Crystal structure predictions for molecules with soft degrees of freedom using intermonomer force fields derived from first principles

Rahul Nikhar,[a] Rafał Podeszwa,[b] Atta U. Rehman,[a] Ommair Ishaque,[a] Aling Jing,[a] John W. Melkumov,[a] Richard Hong[c,d] and Krzysztof Szalewicz [a]*

[a]*Department of Physics and Astronomy, University of Delaware, Newark, Delaware 19716 USA,* [b]*Institute of Chemistry, University of Silesia in Katowice, Szkolna 9, 40-006 Katowice, Poland,* [c]*AbbVie Inc., Research & Development, 1 N Waukegan Road, North Chicago, IL 60064, USA, and* [d]*Department of Chemistry, New York University, New York, New York 10003, USA. E-mail: szalewic@udel.edu*

Abstract

A molecular crystal structure prediction (CSP) protocol used in the seventh blind test (7BT) of CSPs is presented. The 7BT was divided into two stages and included seven targets, with crystals containing from 1 to 3 molecules in asymmetric units, monomers built of up to 100 atoms, and all targets containing monomers with flexible degrees of freedom. Some targets were cocrystals and one target was a salt These diverse targets were treated using a CSP protocol starting from finding the global and local minima conformations of the target molecule. Subsequently, an *ab initio* two-body rigid-monomer 6-dimensional force field (aiFF) was developed for the global-minimum conformer. These aiFFs were then used in CSPs consisting of packing and lattice-energy minimization stages. Flexible-monomer CSPs were used for some targets. To describe the intramonomer FF, either generic empirical FFs or reparametrized FFs of this type were used, with some parameters fitted to *ab initio* energies of monomers in the latter case. A novel packing procedure was applied for two targets in stage 1. The success rate in the structure generation stage was 15% in submission phase and 54% in post-submission phase, while the corresponding values in the structure rating stage were 33% and 89%. We conclude that the inexpensive conformer-based approach with rigid-monomer CSPs can be recommended for investigations of crystals with flexible monomers. An advantage of this protocol is that it is fully based on first-principles quantum mechanics and generates tailor-made FFs suitable for use in subsequent molecular dynamics simulations investigating temperature-dependent effects. However, empirical intramonomer FFs reparametrized using *ab initio* data are not yet adequate for CSPs.

Contents

1 Introduction **3**

2 Methods **7**
2.1 Search and optimization of monomer conformations 7
2.2 aiFF generation methodology 10
2.3 Intramolecular FFs 13
2.4 Crystal structure packing and minimizations 13
2.5 Reranking using periodic DFT calculations 14

3 Stages 1 and 2 **14**

4 Post-submission analysis **17**
4.1 Target XXVII 17
4.1.1 Stage 1 17
4.1.2 Stage 2 17
4.2 Target XXVIII 18
4.2.1 Stage 1 18
4.2.2 Stage 2 19
4.3 Target XXIX 19
4.3.1 Stage 1 19
4.4 Target XXX 20
4.4.1 Stage 1 20
4.5 Target XXXI 21
4.5.1 Stage 1 21
4.5.2 Stage 2 23
4.6 Target XXXII 25
4.6.1 Stage 1 25
4.6.2 Stage 2 26
4.7 Target XXXIII 26
4.7.1 Stage 1 26
4.7.2 Stage 2 27

5 Summary and discussion **28**
5.1 Conformer analysis 28
5.2 Accuracy of intermonomer aiFFs 30
5.3 Accuracy of intramonomer FFs 32
5.4 CCDC polymorphs of stage 2 34
5.5 Overall performance of our CSP protocols in stage 1 including post-submission investigations 39
5.6 Overall performance of our CSP protocols in stage 2 including post-submission investigations 43

6 Final conclusions **46**

1. Introduction

A crystal structure prediction (CSP) protocol is presented and applied to targets of the 7$^{\text{th}}$ blind test (7BT) of CSPs organized by the Cambridge Crystallographic Data Centre (CCDC) (The Cambridge Crystallographic Data Centre, 2024). The 7BT consisted of two stages, hereafter referred to as 'stage 1' and 'stage 2'. The targets under investigation are shown in Fig. 1. The systems cover a wide range of applications such as: an optoelectronic target (XXVII), an organometallic target (XXVIII), a flavouring agent target (XXIX), a co-crystal target (XXX), and pharmaceutical-agrochemical targets (XXXI, XXXII, XXXIII). Most of the targets in 7BT are highly flexible and would require knowledge of both intermonomer and intramonomer force field (FFs). Stage 1 focused on structure generation techniques. The 7BT participants' task was to compile a list of up to 1,500 crystal structures utilizing solely the structural formulas of the target monomers. These structures were then compared with experimental structures that had been previously forwarded to the organizers. The process was conducted blindly, meaning that participants were unaware of the details of these crystal structures. However, for system XXVIII, the crystal structure had already been independently identified, rendering the predictions not entirely blind for this system. Despite this, the predictions for this system were collected, and several groups including ours, did not leverage this pre-existing knowledge in their analysis. For target XXIX, participants were additionally supported with powder X-ray diffraction (PXRD) results to aid their predictions. For target XXX, a co-crystal, they were given the further challenge of predicting stoichiometries and generating a list of the 100 most likely structures. The assessments of the submitted structures against the experimental one were multi-staged. Initially, the simulated PXRD outcomes were aligned with the experimental ones, subsequently, the structures were superimposed using COMPACK (Chisholm & Motherwell, 2005) with a molecular shell of 30 molecules with tolerances of 35% and 35° for distances and angles, respectively. A match was deemed successful if the root mean square deviation (RMSD) of atomic positions (excluding hydrogen) was below 1 Å. Consult the 7BT stage 1 paper (Hunnisett *et al.*, 2024*a*) for further details. The specific ranking of the matches within the submitted list of 1,500 structures did not influence the evaluation process, with energy ranking considerations reserved for stage 2. The second stage excluded targets XXIX and XXX, focusing on ranking the structures based on energy or enthalpy for the remaining systems. Participants were provided with either 100 or 500 structures to rank, which included both modified experimental polymorphs and structures submitted during the first stage. The nature of these modifications was initially undisclosed and was only clarified later, as described in the stage 2 paper (Hunnisett *et al.*, 2024*b*). The modification involved constrained-optimization with the Cambridge Structural Database (CSD) knowledge-based force field (Cole *et al.*, 2016). Note that the collaborative work of the University of Delaware and the New York University team was divided in the 7BT papers (Hunnisett *et al.*, 2024*a*; Hunnisett *et al.*, 2024*b*) into groups 26 and 27. We disregard this division here.

We present a methodical CSP approach applied to the 7BT targets, which consist of (*i*) searching for low-energy monomer conformation(s) (Sec. 2.1); (*ii*) selecting an appropriate *ab initio* method and basis set for calculating interaction energies of

dimers; (*iii*) developing accurate rigid-monomer intermolecular two-body *ab initio* FF (aiFFs) (Metz *et al.*, 2016; Metz & Szalewicz, 2020; Garcia *et al.*, 2020), fitted to interaction energies for one or more monomer conformation(s) (Sec. 2.2); (*iv*) for selected target(s), reparametrization of empirical intramonomer FFs by fitting them to *ab initio* computed monomer energies; (*v*) packing and lattice energy minimization using such rigid- or flexible-monomer FFs (Sec. 2.4); (*vi*) utilization of the global flexible-monomer FFs in molecular dynamics (MD) simulations at finite temperatures. As an independent test, calculations for some crystals were performed using periodic density functional theory augmented with dispersion corrections (pDFT+D). One of the main advantages of our protocol is the straightforward inclusion of MD simulations, as the aiFFs generated during CSPs can be directly employed in MD codes. This approach has been shown to effectively address the problem of polymorph overprediction (Francia *et al.*, 2020), reducing the number of potential polymorphs compared to 0 K methods. However, due to time constraints, this advantage was not fully utilized, as we applied MD only to two targets (XXXI and XXXIII) and only in stage 2. In the case of rigid-monomer CSPs, our approach was fully from first-principles, while flexible-monomer CSPs included an empirical component resulting from the use of empirical intramonomer FFs (even if these FFs were refitted to *ab initio* monomer energies, some parameters retained empirical values).

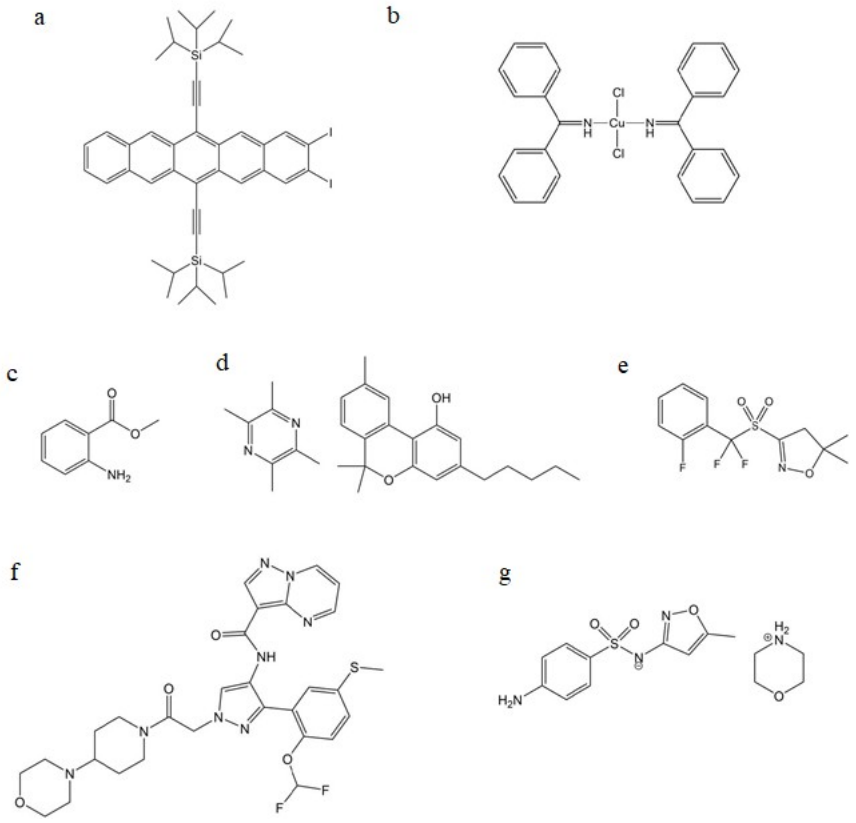

Fig. 1. Molecular diagrams of 7BT targets: (a) XXVII, (b) XXVIII, (c) XXIX, (d) XXX, (e) XXXI, (f) XXXII, and (g) XXXIII.

All the targets of 7BT include non-rigid monomers, i.e., monomers containing degrees of freedom whose values can easily change in the crystal environment (so-called soft degrees of freedom, mainly dihedral angles). Such monomers will deform in crystals relative to their gas-phase equilibrium structures. This occurs to a larger extent when the energy differences between deformed monomers and the gas-phase equilibrium ones are on the order of 10 kJ/mol, a relatively small energy compared to the lattice energies (per monomer), which are of the order of 100 kJ/mol.

While the methods based on intermonomer aiFFs work very well for rigid monomers (Podeszwa *et al.*, 2008; Podeszwa *et al.*, 2009; Reilly *et al.*, 2016; Nikhar & Szalewicz, 2022), extensions to flexible monomers were in an early stage of development when the 7BT started. Although aiFFs were applied in all-dimensional CSPs (Reilly *et al.*, 2016; Metz *et al.*, 2022), in both cases the monomers were nearly rigid. The first-principles-based CSP methodology described in this paper was essentially developed dynamically during our work on 7BT. This work was briefly described in the supporting information B of stage 1 (SI1-B) (Hunnisett *et al.*, 2024*a*) and stage 2 (SI2-B) (Hunnisett *et al.*, 2024*b*) of the 7BT publications. This research explored various previously unknown aspects of first-principles-based all-dimensional CSPs for large monomers with several soft degrees of freedom. For such cases, approximate methods for addressing the monomer flexibility problem are required. In our approach to perform 7BT CSPs, we utilized rigid-monomer intermolecular aiFFs and introduced two distinct strategies to account for monomer flexibility. For monomers exhibiting limited flexibility, we assumed that their geometric configuration within a crystal closely resembles one of the local minima observed on the electronic potential energy surface (PES) in its gas-phase state; essentially, these monomers are in proximity to a specific conformational isomer (conformer) of this molecule. For monomers demonstrating greater flexibility, we applied empirical FFs (including reparameterized ones) for modeling interactions within the monomer, while intermolecular interactions were modeled using aiFF, based on the assumption that there is no coupling between these two sets of forces. This approach raises three fundamental questions. The first revolves around the application of aiFFs, originally designed for rigid monomers, to their flexible counterparts. This is formally possible since such surfaces depend only on distances between atoms in different monomers, allowing for calculations with monomers in various states of deformation. We refer to this process as "flexibilization" of the surfaces. An important consideration is the effectiveness of such a flexibilization approach. The second question relates to the accuracy of generic empirical FFs in reproducing the intramolecular forces. The third question addresses the potential uncertainties introduced by our assumption of no coupling between intermolecular and intramolecular interactions, a premise commonly adopted in empirical FFs. No couplings between inter and intramonomer degrees of freedom that the total energy of a dimer is the sum of intermonomer energy depending only on intermonomer degrees of freedom and of intramonomer energies depending only on intramonomer degrees of freedom. As a result, for a fixed configurations of each monomer in a dimer, the monomer energies remain the same no matter what dimer configuration is. The other way around, the interaction energy for some fixed intermononomer distance and a fixed set of Euler angles determining mutual orientation of monomers does change when monomers deform, but these changes are completely independent of intramonomer

potentials. If such couplings were non-negligible, it would require the intermolecular parameters to depend on the intramonomer coordinates, i.e., be functions of such coordinates. Alternatively, FFs would have to have terms containing producs of inter and intramonomer coordinates. Either option would substantially increase the complexity of the FF. One should mention that autoPES can produce fully coupled FFs (Metz & Szalewicz, 2020). Although we recognized the limitations of empirical FFs for monomers involved in 7BT, time constraints within stage 1 of 7BT did not permit alternative methodologies. In stage 2, we partly reparameterized some empirical FFs to address their limitations. The errors of these three assumptions have been tested and are discussed in the following sections.

The 7BT was much more difficult than the 6BT since the monomers are approximately twice larger, translating into approximately $2^5 = 32$ times longer *ab initio* calculations. However, an even harder aspect of 7BT was the fact that each target's monomers have many soft degrees of freedom. This made the consideration of monomer-flexibility effects a critical aspect of the work. This issue was further amplified due to the crystal disorder present in systems XXVII, XXX, XXXI, and XXXII, which made comparisons with experimental structures difficult. Even more challenging is dynamic disorder, where several structures can coexist in thermal equilibrium, which was identified for system XXVII (Hunnisett *et al.*, 2024*a*). Finally, polytypism, where polymorphs emerge due to variations in the sequence of stacked layers, was observed in system XXIX (Hunnisett *et al.*, 2024*a*). Nevertheless, we worked on all targets and completed all stages for each system. In fact, we were one of only two purely academic group who responded to all challenges. For comparison, in 6BT we were only able to perform CSPs for two targets. These systems were smaller than the smallest one in 7BT, and for most of the 6BT work, we assumed the monomers to be rigid.

Within the human and computational resources available for our team, in stage 1 we were able to develop only a single intermonomer aiFF for each pair of monomers in a target with monomers in their equilibrium conformations for all targets except target XXIX, where we developed three aiFFs, see section S1.3 in Supporting Information. With such aiFFs, we used the following strategies to account for monomer-flexibilities:
(a) Conducting rigid-monomer CSPs (packing and lattice energy minimization) using monomers at their gas-phase global minima, i.e., using the exact protocol from(Nikhar & Szalewicz, 2022). This amounts to ignoring flexibility of monomers.
(b) Implementing the flexibilized version of the aiFF in several rigid-monomer CSPs with monomers geometries corresponding to lowest-energy conformers. Combining such CSP lists into one and adding the deformation penalties for the monomers calculated using a DFT+D method to obtain the final ranking. This protocol is essentially a straightforward extension of the CSP protocol presented in (Nikhar & Szalewicz, 2022).
(c) Utilizing flexibilized rigid-monomer aiFF in conjunction with generic empirical intramonomer FFs in flexible-monomer CSPs (three options: UPACK (van Eijck & Kroon, 1999; Nikhar & Szalewicz, 2022) only, UPACK + MD, MD only). The final rankings in these cases also depend on DFT+D monomer energies providing the deformation penalties.

In stage 2, we improved aiFFs in two ways. One was by using the clusters cut

from crystals (CCfC) method (Nikhar & Szalewicz, 2024), which adds training points representing dimers from crystal structures predicted from the previous version of aiFF. In the second approach, we added training points with monomers in geometries other than from the global minimum (either local minima or geometries from crystals provided by CCDC). Such aiFFs perform better upon flexibilization. Since the crystal structures to be ranked were provided by CCDC, the packing stages of CSP were not performed in stage 2. One important improvement of our protocol from point (c) above was the use of reparameterized intramonomer FFs, where some parameters of empirical FFs were replaced by *ab initio* values calculated for the target monomers.

The paper is organized as follows: first we provide an overview of the methods used in our CSP, followed by the post-submission analysis of each individual system. The final section included a summary and discussion, highlighting the main conclusion and recommendation of this work. Detailed description of the stage 1 and stage 2 work for each system are available in the Supporting Information. Figures, tables, and section numbers from the Supporting Information are prefixed with 'S' throughout this paper; for example, Figure S4 refers to the corresponding figure in the Supporting Information.

2. Methods

2.1. Search and optimization of monomer conformations

In stage 1, the search for conformers began by building an initial three-dimensional (3D) structure using the monomer 2D molecular diagrams, shown in Fig. 1. To refine the 3D structures, an empirical FF was employed to estimate the bond lengths and angles. Subsequently, several potential conformations for each monomer were created using a range of computational tools, including Avogadro (Hanwell *et al.*, 2012), Baloon (Vainio & Johnson, 2007), conformer generator (Cole *et al.*, 2018) (within Mercury (Macrae *et al.*, 2006) program), Conformator (Friedrich *et al.*, 2019), and PLUMED (Plimpton, 1995; Bonomi *et al.*, 2009). A subset of these conformations, were then subject to further optimization through *ab initio* methods, with the exception of one case, to determine their energy rankings. This step was crucial for identifying and discarding conformers that were energetically or geometrically too similar to others. The optimization of conformers using *ab initio* techniques was undertaken utilizing various DFT+D methods. The selection process for the conformers to be used in the CSPs was guided by the energy landscape; the focus was, depending on the system, either on only the global minimum conformer or both the global minimum and several higher-energy conformers that were local minima, provided they were within a few kcal/mol of the global minimum.

Table 1. *Summary of methodologies used in stage 1. See text for acronyms. All CSPs used the UPACK package. All intermonomer FFs included SAPT(DFT) asymptotics. All RMSEs are in kcal/mol.*

Target	Monomer geometry	Intermonomer FF	Intramonomer FF	CSP type	Results
XXVII	Avogadro/MMFF94, Avogadro/GA, PBE0+D3(BJ) def2-SVP/def2-TZVPP	PBE0+D3 def2-TZVPP, RMSE: 0.69	none	rigid monomer with equilibrium conformer and aiFF	no polymorph within CCDC criteria
XXVIII	Balloon/MMFF94, Balloon/GA, PBE+D3(BJ) aug-cc-pVTZ	PBE0+D3(BJ) cc-pVTZ, RMSE: 0.42	none	rigid monomer with equilibrium conformer and aiFF	no polymorph within CCDC criteria
XXIX	Avogadro/UFF, Conformer generator/Mercury, PBE0+D3 aug-cc-pVTZ	SAPT(DFT)/PBE aug-cc-pVDZ+mb $MC^{+}BS$, CCfC improvement RMSEs: 0.14, 0.15, 0.15	none	rigid monomer, two conformers and their mixture, refinement with fits to PXRD spectra, additional EVCCPMCE search	no polymorph within CCDC criteria, however our closest experimental match has correctly predicted $Z' = 3$ with $RMSD_{27} = 0.197$ Å, see Sec. S1.3.5.
XXX	Avogadro/UFF, Conformer generator/Mercury, PBE0+D3 aug-cc-pVTZ	SAPT(DFT)/PBE aug-cc-pVDZ+mb $MC^{+}BS$, CCfC improvement RMSEs: 0.17, 0.39, 0.47	OPLS-AA/q(aiFF)	rigid and flexible monomer UPACK searches followed by EVCCPMRE searches, monomer deformation penalties from OPLS-AA/q(aiFF)	no polymorph within CCDC criteria, stoichiometry predicted correctly
XXXI	Avogadro/UFF, Conformator, PBE+D3(BJ) aug-cc-pVTZ	SAPT(DFT)/PBE aug-cc-pVDZ+mb $MC^{+}BS$, RMSE: 0.16	none	rigid-monomer and 8 conformers plus 2 optical isomers, PBE+D3(BJ) energy penalties	polymorph #1192 matched experimental polymorph A_{min} with $RMSD_{30} = 0.857$ Å
XXXII	Avogadro/GAFF	PBE0+D3(BJ) def2-TZVP, RMSE: 0.487	GAFF/q(aiFF)	all flexible monomer with either GAFF for all interaction or aiFF for inter- and GAFF for intramonomer interactions UPACK followed by EVCCPMRE	no polymorph within CCDC criteria
XXXIII	Avogadro/UFF, Conformer generator/Mercury, PBE0+D3 aug-cc-pVTZ, CSD search MP2/6-31G*	SAPT(DFT)/PBE aug-cc-pVDZ+mb $MC^{+}BS$, RMSEs: 1.52, 2.27, >100	GAFF/q(RESP)	flexible monomer UPACK all GAFF followed by aiFF/GAFF UPACK, deformation penalties from PBE0+D3/6-31G**	no polymorph within CCDC criteria

Table 2. *Ranges of energies (in kJ/mol) for the monomers from the CCDC stage 2 polymorphs lists relative to the given monomer equilibrium energy. The latter energies are also listed (in hartree).*

	XXVII	XXVIII	XXXI	XXXII	XXXIII
lowest	10.41	4.95	13.1	6.29	1.10
highest	1310.47	991.00	554.9	1232.00	137.03
method	PBE0+D3BJ	PBE+D3BJ	PBE+D3BJ	PBE0+D3BJ	PBE0+D3BJ
basis set	def2-TZVPP	aug-cc-pVTZ	aug-cc-pVTZ	def2-TZVP	aug-cc-pVTZ
E_{equil}	−2880.886276	−3673.074491	−1441.717218	−2459.456556	−1174.512790

In stage 2, the goal was ranking the crystal structures provided by CCDC. One straightforward strategy could be to perform lattice energy calculations using the flexibilized aiFF and then add penalties calculated with respect to the global minimum conformer. Table 2 shows the ranges of monomer energies from the CCDC lists.

Table 3. *Summary of methodologies used in stage 2. See text for acronyms. All CSPs used the UPACK package. All intermonomer FFs included SAPT(DFT) asymptotics.*

Target	Monomer geometry	Intermonomer FF	Intramonomer FF	CSP type	Results
XXVII	Constrained minimization of monomers from CCDC crystals, PBE0+D3(BJ) def2-TZVPP	CCfC based on stage 1, dimers from CCDC structures, RMSE: 0.66	none	rigid monomer UPACK minimization with aiFF, PBE0+D3(BJ) def2-TZVPP monomer deformation penalties, pDFT+D used in ranking	experimental crystal ranked as #3
XXVIII	Constrained minimization of CCDC monomers, PBE+D3(BJ) aug-cc-pVTZ	CCfC based on CCDC polymorphs with original monomer replaced by constrained-optimized monomers, RMSE: 0.48	none	rigid monomer UPACK minimization with aiFF, PBE+D3(BJ) aug-cc-pVTZ monomer deformation penalties	experimental crystal ranked as #1
XXXI	Constrained minimization of CCDC monomers, PBE+D3(BJ) aug-cc-pVTZ	CCfC based on CCDC polymorphs with original monomer replaced by constrained-optimized monomers, RMSE: 0.25	revised GAFF with aiFF charges	rigid monomer UPACK minimization followed by MD GAFF monomer deformation penalties	A_{maj} - rank 21, A_{min} - rank 35, B - beyond CCDC criteria
XXXII	Constrained minimization of CCDC monomers, PBE+D3(BJ) aug-cc-pVTZ	CCfC based on CCDC polymorphs with original monomer replaced by constrained-optimized monomers, RMSE: 0.63	none	rigid monomer UPACK minimization with aiFF, PBE0+D3(BJ) def2-TZVP monomer deformation penalties	Form A - rank 23
XXXIII	Constrained minimization of CCDC monomer using PBE+D3(BJ) aug-cc-pVTZ	CCfC using dimers from CCDC polymorph, SAPT(DFT)/PBE aug-cc-pVDZ+mb $MC^{+}BS$, RMSEs: 0.18, 0.70, 1.46	refitted GAFF with hard parameters: Joyce fit to equilibrium *ab initio* monomer (including Hessian), and soft parameters: Paramfit fit to 2000 *ab initio* geometries, RMSEs: S: 1.57 kcal/mol, M: 0.50 kcal/mol	flexible monomer UPACK minimizations with generic GAFF for inter and refitted GAFF for intramonomer interactions, rigid monomer UPACK minimization with aiFF, MD minimization at 0K followed by NVT, followed by NPTF all with aiFF for intermonomer and refitted GAFF for intramonomer interactions, PBE0+D3 aug-cc-pVDZ monomer deformation penalties	Form A ranked 345, form B ranked 132, correct energetic order

For some systems, these ranges are as large as 1000 kJ/mol. From stage 1 calculations, we knew that intermonomer contributions to lattice energies for these targets are in the hundreds of kJ/mol. This meant that the monomer-deformation penalties of the order of 1000 kJ/mol would make the crystals with such monomers unbound. Moreover, the range of monomer energies coming from CSPs with empirical FFs is typically within dozens of kJ/mol more. Specifically, for the flexible CSP of target XXX, the monomer C energy range is less than 40 kJ/mol, in stark contrast with the ranges in Table 2; see Fig. 1 in SI2-B of the stage 2 7BT paper (Hunnisett *et al.*, 2024*b*). The presence of the highly-deformed monomers in the structures provided by CCDC could be therefore only due to purposeful deformations introduced by the CCDC team. Indeed, as we know now, the CCDC team performed limited minimizations of lattice energies of some or all crystals sent to the participants using an in-house empirical FF (Cole *et al.*, 2016). Apparently, this FF performs poorly in representing monomer energies. Based on these data, we assumed that stage 2 CSPs had to involve significant optimization of monomer geometries, because otherwise most of the provided crystal structures would have to be immediately rejected due to monomer-deformation penalties. To bring monomer energies to within reasonable ranges, we constrained-optimized all monomer geometries extracted from the CCDC crystal lists for all stage 2 targets. The degrees of freedom constrained in the optimizations were soft dihedral angles. The rationale for this was that the CCDC team probably did not distort such dihedral angles since these distortions would have changed the shapes of monomers and led

to unreasonable crystal structures. For the largest systems, we have constrained all of the dihedral angles for computational efficiency reasons. Tests on selected monomers have shown that the energies obtained in these two ways differed in most cases by less than 1 kJ/mol.

2.2. aiFF generation methodology

The aiFFs were developed using the autoPES software package (Metz *et al.*, 2016; Metz & Szalewicz, 2020). The autoPES aiFF development procedure is divided into six parts: asymptotic calculations, grid points generation for close-range dimer configurations, interaction energy calculation at each grid point, fitting an analytic functional form to the data, removal of 'holes' (see below), and evaluation of the quality of the fit. If the quality is insufficient, the process is repeated, except for the asymptotic component. Such iterations continue until convergence. Each step is described in detail in autoPES paper (Metz & Szalewicz, 2020).

For the *ab initio* calculations of interaction energies, depending on the system, we used perturbation theory or the supermolecular method, as detailed in Table 1. Specifically, for four systems, symmetry-adapted perturbation theory (SAPT) (Jeziorski *et al.*, 1994) based on a DFT description of the monomers, SAPT(DFT) (Misquitta *et al.*, 2005), was our selection. This method was used within the SAPT2020 (Garcia *et al.*, 2020; Bukowski *et al.*, 2020) package, with ORCA-4.2.1 (Neese, 2012; Neese, 2018) being used for the DFT calculations of the monomers. The SAPT(DFT) calculations typically include the $\delta E_{\rm int}^{\rm HF}$ term, which accounts for the difference between the sum of SAPT terms based on the Hartree-Fock (HF) description of the monomers [SAPT(HF)] and the supermolecular HF energy, and we added this component to our SAPT(DFT) energies. Unless otherwise stated, our calculations were performed using the aug-cc-pVDZ basis set (Kendall *et al.*, 1992; Woon & Dunning, Jr., 1993). These were employed in the so-called monomer-centered plus basis set approach (MC^+BS) (Williams *et al.*, 1995) and augmented with additional midbond functions (3s3p2d2f) positioned at the midpoint of the line linking the centers of mass of the monomers. This setup is the default basis set choice in the autoPES software (Metz *et al.*, 2016; Metz & Szalewicz, 2020).

For the remaining three systems, we adopted the supermolecular approach. We calculated the intermolecular energies using the PBE0 (Perdew *et al.*, 1996; Adamo & Barone, 1999) DFT functional, combined with Grimme's D3 dispersion correction (Grimme *et al.*, 2010), and we applied the counterpoise corrections. In essence, the latter involves computing monomer energies using the dimer-centered basis set (DCBS). We did not include midbond functions in our supermolecular calculations. We used the aug-cc-pVDZ basis set, unless noted differently for specific cases.

The functional form of the aiFF used for fitting the intermolecular energies (Metz *et al.*, 2016; Metz & Szalewicz, 2020) is given by:

$$V = V_{\rm elst} + V_{\rm exp} + V_{\rm asymp}^{(2)} = \sum_{a\in A, b\in B} \Big[u_{{\rm elst},ab}(r_{ab}) + u_{{\rm exp},ab}(r_{ab}) + u_{{\rm asymp},ab}^{(2)}(r_{ab}), \Big] \quad (1)$$

where a and b iterate over atoms in monomers A and B, respectively, with r_{ab} denoting

the distance between atoms a and b. The expressions for the pairwise atom-atom interaction functions are:

$$u_{\mathrm{elst},ab}(r_{ab}) = f_1(\delta_1^{ab}, r_{ab})\frac{q_a q_b}{r_{ab}}$$

$$u_{\mathrm{exp},ab}(r_{ab}) = \left[1 + \sum_{i=1}^{k} a_i^{ab}(r_{ab})^i\right] e^{\alpha^{ab}-\beta^{ab}r_{ab}} + \frac{A_{12}^{ab}}{(r_{ab})^{12}}$$

$$u_{\mathrm{asymp},ab}^{(2)}(r_{ab}) = -\sum_{n=6,8} f_n(\delta_n^{ab}, r_{ab})\frac{C_n^{ab}}{(r_{ab})^n}, \tag{2}$$

where q_a, q_b, a_i^{ab}, α^{ab}, β^{ab}, A_{12}^{ab}, C_n^{ab}, and δ_n^{ab} are parameters that are modified during the fitting process. The function f_n is the Tang-Toennies damping function (Tang & Toennies, 1984) defined as:

$$f_n(\delta, r) = 1 - e^{-\delta r}\sum_{m=0}^{n}\frac{(\delta r)^m}{m!}. \tag{3}$$

All atom-atom functions included the following terms: Coulomb, repulsion $A_{12}^{ab}/(r_{ab})^{12}$, induction plus dispersion C_6^{ab}/r_{ab}^6, and exponential with the second order polynomial ($k = 2$). The Coulomb and induction plus dispersion terms were always damped. This gives the total of 10 parameters: 3 asymptotics and 7 close-range. For target XXVII, there were 2 more parameters: δ_8^{ab} and C_8^{ab}. The set of distributed induction plus dispersion coefficients C_n^{ab}—all of which were required to be positive—along with partial charges q_a and q_b on atoms were fitted on a comprehensive grid of long-range interaction energies. These energies were calculated from an *ab initio* distributed asymptotic (aiDA) expansion (Rob & Szalewicz, 2013*a*; Rob & Szalewicz, 2013*b*) that is consistent with SAPT(DFT) and was used in SAPT2020 (Garcia *et al.*, 2020; Bukowski *et al.*, 2020) framework. Ensuring that aiFFs behaved correctly at long ranges, these parameters were held constant during later stages of fitting. The subsequent step in the process involved optimizing parameters a_i^{ab}, α^{ab}, β^{ab}, A_{12}^{ab}, and δ_n^{ab} against the *ab initio* interaction energies of the dimer, E_{int}. These calculations spanned intermediate- and short-range intermolecular distances, referred to as close-range grid points. In case of SAPT(DFT) calculations, we fitted the damping parameters δ_1^{ab} and δ_6^{ab} separately: the former to the electrostatic energies and the latter to the aggregated induction and dispersion energies, including their exchange counterparts. For supermolecular calculations, we fitted all close-range parameters to the total interaction energies. The initial selection of grid points for computing the intermolecular energies was based in the first iteration on the values of interaction energies calculated using OPLS-AA (Jorgensen & Tirado-Rives, 1988) and was refined by a tentative aiFF in subsequent iterations. In the process of fitting, we prioritized the grid-points with negative interaction energies by assigning them larger weights compared to those with positive interaction energies (see AutoPES publications (Metz *et al.*, 2016; Metz & Szalewicz, 2020) for details of the weighting function). In the final iteration, the fit was checked for the presence of "holes", namely, regions with nonphysical negative or low energy at short intermonomer separations, where the interactions should be strongly repulsive. The

holes that were detected were removed by adding grid-points in such regions and refitting the aiFF. To test the aiFF convergence, we divided the entire set of grid points into a test set (comprising 30% of the total) and a training set (making up the remaining 70%). The convergence was deemed achieved when the root mean square error (RMSE) observed in the test data was less than 1.2 times the RMSE of the training set. For target XXIII, with ionic monomers, we should have used also the term $f_4(\delta_4^{ab}, r_{ab})C_4^{ab}/(r_{ab})^4$, the leading asymptotic induction term for charged systems. However, this term was incompatible with software used in CSPs for this target. Therefore, we could not include this term although autoPES can produce such fits. This omission is not consequential for ion-ion interactions since the asymptotics is strongly dominated by the $1/R$ Coulomb interactions of permanent charges.

As already mentioned in Sec. 2.1, in stage 1 the aiFFs were developed using only rigid, equilibrium monomers in the training set. In stage 2, dimers extracted from the CCDC crystals were added to the training set. Such dimers contain monomers which are all different from the equilibrium monomer optimized by us. This can be done for the same reason that one can use flexibilized forms of aiFFs: the fit form, provided by Eqs. (1)–(3), depends only on intermolecular atom-atom distances and excludes any explicit reference to the monomer geometry. This exclusion inherently allows, in general, for the use of grid points generated for monomers possessing a range of arbitrary geometries during the fitting process. In turn, any aiFF in the form of Eqs. (1)–(3) can be employed to compute interaction energies across a spectrum of arbitrarily deformed monomers. We call an aiFF used in such a way a "flexibilized" form of this aiFF. Deploying such flexibilized aiFFs was important in CSP scenarios involving monomer conformations that did not conform to the gas-phase global minimum (even if only the equilibrium monomer was used in the aiFF development) and was particularly valuable in CSPs that used fully flexible monomers.

It is also worth noting that autoPES (Metz & Szalewicz, 2020) is capable of creating flexible-monomer PESs by including only specific degrees of freedom for each monomer. For monomers that have up to three soft degrees, the resulting PESs would have been up to twelve-dimensional. Such aiFFs could enable performing flexible-monomer CSPs entirely grounded in first principles. However, the computational resources required for this approach are too large to be feasible within 7BT's time constraints.

In stage 1, for targets XXIX and XXXI, the first generation aiFFs underwent refinement by adding into the training set a select number of nearest-neighbor dimer configurations. We extracted these dimers from 5×5×5 supercells of the best-ranking polymorphs from CSPs that were performed with the original version of this aiFF and used this enlarged training set in a new fit. This is called the clusters cut from crystals (CCfC) approach (Nikhar & Szalewicz, 2024). Despite the improvements in RMSE of the refitted aiFFs being relatively minor, the addition of dimer configurations from the top-ranking polymorphs enhanced the performance of aiFFs in the CSPs. This is due to a notable improvement in regions of the aiFF that are relevant to the investigated crystals.

2.3. Intramolecular FFs

In stage 1, CSPs for three targets (XXX, XXXII, and XXXIII) were performed in all dimensions, i.e., including monomer energies in the minimization of lattice energies. To represent intramonomer interactions, we used literature generic empirical FFs. Specifically, for target XXX, OPLS (Jorgensen & Tirado-Rives, 1988) was used, while for the remaining two targets, we used GAFF (Wang *et al.*, 2004).

In stage 2, targets XXXI and XXXIII included intramonomer contributions in the optimization of lattice energies and in MD simulations. For target XXXI, the intramolecular part was reparameterized by replacing the original equilibrium bond lengths and angles by the corresponding values from the *ab initio* equilibrium structure. Also the partial charges were taken from the aiFF. For target XXXIII, partly *ab initio* intramonomer FF was developed using the AMBER Paramfit software (Betz & Walker, 2015) and Joyce (Barone *et al.*, 2013) codes. In the first step, the Antechamber software, as implemented in AmberTools20 (Case *et al.*, 2020), was used for determining the generic intramolecular parameters, i.e., bond, angle, improper and proper dihedrals, Lennard-Jones parameters, and partial charges, which were used as a starting point for refitting the intramolecular FF. Point charges were then replaced by those from the aiFF intermolecular potential. Briefly, Joyce was used to fit bond-stretch, angle-bend, improper dihedrals, and rigid proper dihedrals parameters to the Hessian of the global minimum geometry. In the subsequent phase of refinement process, while keeping most of the parameters derived from Joyce constant, the fitting involved the flexible dihedrals and improper torsions. This was achieved using the AMBER Paramfit software (Betz & Walker, 2015) and, as a result, the improper torsions from Joyce were modified. Thus for XXXIII, the only empirical component remaining in the reparameterized FF were the Lennard-Jones parameters.

2.4. Crystal structure packing and minimizations

A locally tailored version of UPACK was used for generating random crystal structure candidates, filtering them to obtain a collection of plausible candidates, and minimizing lattice energies of the remaining set of polymorphs using aiFF. This adaptation allowed us to apply a more complex version of the interatomic functional form specifically developed for aiFFs, as described in Eqs. (1)–(3). This approach surpasses the simpler models found in standard UPACK, which are limited to the basic Buckingham plus charges (exp-6+q) and Lennard-Jones with charges (LJ+q or 12-6-1) potentials. In the first step of the rigid-monomer UPACK CSP protocol, we generated 20 potential polymorphs for each space group, imposing no constraints. During the next phase, known as the packing step, we chose the polymorph exhibiting the highest density from this initial pool to establish limits for further exploration in coordinate space. This strategy led to the generation of a significant number of hypothetical crystal polymorphs using a quasi-random approach, creating, by default, 5,000 structures per space group. Our analysis consistently covered (unless noted otherwise) at least 13 space groups, following the UPACK default settings: $P1$, $P\bar{1}$, $P2_1$, $C2$, Pc, Cc, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Pbcn$, and $Pbca$. In most cases, we assumed a single molecule in the asymmetric unit ($Z' = 1$) using UPACK's 'pack12' program. To increase diversity of the potential structures in the $P2_1/c$ space group specifically, we

also conducted searches in the non-standard equivalent variants of this group ($P2_1/n$ and $P2_1/a$) using the identical random seed as for $P2_1/c$, rather than starting each search with a fresh seed. The first step of the energy minimization process of the hypothetical polymorphs involved optimization using the OPLS-AA FF (Jorgensen & Tirado-Rives, 1988) for the dispersion and repulsion components of the LJ potential and incorporating either SAPT or the restrained electrostatic potential (RESP) charges (Bayly *et al.*, 1993) for the Coulombic interactions within the 12-6-1 potential model. Following this, polymorphs that were duplicates, identified by falling within a predetermined similarity threshold, were removed. This was achieved by applying a clustering algorithm (van Eijck & Kroon, 1997), using the 'dist' program from the UPACK suite of tools. The remaining polymorphs then underwent further optimizations with tight thresholds using aiFFs, employing the 'pack3' program of UPACK. This process also included an additional round of clustering to refine the ultimate CSP lists. This standard CSP procedure was followed in stage 1. In stage 2, since a set of polymorphs for each target was provided by CCDC, only the fine optimization step using aiFFs was performed.

In addition to the UPACK structure generation codes, we used the Extended Variables Coupled to Crystal Polymorph Modified Replica Exchange (EVCCPMRE) method (Chan & Tuckerman, 2024). For details of the procedure of using EVCCPMRE in the application to 7BT targets, see the SI1-B of the stage 1 7BT paper (Hunnisett *et al.*, 2024*a*). The EVCCPMRE method provides generally better sampling of the structures than the quasi-random algorithm in UPACK. However, due to time constraints, we used this method for systems XXIX, XXX, and XXXII only.

2.5. Reranking using periodic DFT calculations

In stage 2, we have also used periodic dispersion-corrected DFT calculations (pDFT+D), exclusively employing single-point evaluations (fixed crystal geometry, no optimizations). For target XXVII, the pDFT+D lattice energy calculations were performed using the SCAN (Sun *et al.*, 2015) functional with the projector augmented-wave pseudopotentials (Blöchl, 1994; Kresse & Joubert, 1999) plus the D3 dispersion correction (Grimme *et al.*, 2010) in VASP 5.4.4 (Kresse & Hafner, 1993; Kresse & Hafner, 1994; Kresse & Furthmüller, 1996*a*; Kresse & Furthmüller, 1996*b*; Kresse & *et al.*, 2021). For targets XXVIII, XXXI, and XXXIII, we utilized the PBE (Perdew *et al.*, 1996) functional with pseudopotentials (Rappe *et al.*, 1990) plus the D3 dispersion correction (Grimme *et al.*, 2010) within Quantum ESPRESSO (QE) (Giannozzi *et al.*, 2009; Giannozzi *et al.*, 2017) software. We employed these pDFT+D calculations mostly as an additional means to validate our findings. However, in one instance (target XXVII), we also elevated certain polymorphs ranked highly by pDFT+D in our submission list.

3. Stages 1 and 2

A detailed report of the computational methodology used in the predictions and our performance in both stages of 7BT are given in the SI, and the additional references used in performing these calculations are Ref. (Halgren, 1996; Halgren, 1999;

Holland, 1975; Grimme, 2011; Weigend & Ahlrichs, 2005; Peterson *et al.*, 2003; de Oliveira *et al.*, 2010; Dodda *et al.*, 2017; Spellmeyer *et al.*, 1997; Żuchowski *et al.*, 2008; Riplinger *et al.*, 2016; Weigend & Häser, 1997; Weigend *et al.*, 1998; Feyereisen *et al.*, 1993; Bernholdt & Harrison, 1996; Mardirossian & Head-Gordon, 2016; Podeszwa *et al.*, 2006; Misquitta & Szalewicz, 2005; Riplinger *et al.*, 2016; Podeszwa *et al.*, 2006; Rappé *et al.*, 1992; Becke, 1988; Lee *et al.*, 1988; Misquitta & Szalewicz, 2005; Riplinger & Neese, 2013; de Gelder *et al.*, 2001; Kullback & Leibler, 1951; Chan *et al.*, 2021; Groom *et al.*, 2016; Clark *et al.*, 1989; Heinzerling *et al.*, 2012; Dunning, Jr., 1989; Frisch *et al.*, 2016; Chen *et al.*, 2012; Barducci *et al.*, 2008; Abrams & Tuckerman, 2008; Jakalian *et al.*, 2002; Ditchfield *et al.*, 1971; Binkley *et al.*, 1980; Frisch *et al.*, 1990; Petersson & Al-Laham, 1991; Weigend & Ahlrichs, 2005; Tuckerman *et al.*, 2000; Spek, 2009; Barone *et al.*, 2013). Briefly, the structure generation stage, i.e., stage 1, consisted of providing a list of 1500/100 possible polymorphs. The standard protocol used for stage 1 began with assembling a set of potential conformers through a conformational search, as discussed in Sec. 2.1. Subsequently, the global minimum from conformational search was selected for aiFF development, as outlined in Sec. 2.2, with the exception of target XXIX, for which both conformers contributed to the aiFF creation. The following step involved selecting an appropriate *ab initio* method to calculate two-body interaction energies. Given the sizes of targets XXVII, XXVIII, and XXXII, supermolecular DFT+D method was deemed an appropriate choice for performing interaction energy calculations, which were consistently executed in counterpoise-corrected basis sets. Conversely, the remaining targets used SAPT(DFT) for interaction energy calculations. Following these steps, the developed aiFFs were applied in CSPs as described in Sec. 2.4. The polymorph list for targets XXVII, XXVIII, XXIX, and XXXI was generated using rigid-monomer CSP, whereas the remaining targets used flexible-monomer CSP. In flexible CSPs, the aiFF was used in a flexibilized manner, as explained in Sec 2.2. *Ab initio* intramonomer penalties, calculated with respect to the target's global minimum geometry, were added to the polymorphs' lattice energy to arrive at the final CSP lists.

Stage 2 of the 7BT focused on ranking of crystal structures by stability. This stage involved ordering lists of either 100 (targets XXVII and XXXI) or 500 polymorphs (targets XXVIII, XXXII, and XXXIII) for each target, as provided by CCDC. The methods behind the preparation of these lists remained undisclosed to the participants. Our team inferred that the preparation process might mirror the informal practices observed in the 6th blind test (6BT) (Reilly *et al.*, 2016), where lists generated by empirical FFs were shared among groups specializing in polymorph ranking. We also anticipated the absence of any polymorph sufficiently resembling the experimental polymorph within the 7BT criteria of similarity ($RMSD_{30} < 1$ Å and within 25%/25° tolerances), based on the common limitations of empirical FFs of providing sufficiently accurate geometries. Furthermore, we expected the list to include a variant of the experimental polymorph, but deformed enough to push its $RMSD_{30}$ beyond the test's acceptance thresholds. In this case, accurately ranking of such a crystal would be impossible without a significant optimization of both monomer and crystal geometries. This perspective was reinforced by our analysis of the energy range of the monomers for the targets, as shown in Table 2, see Sec. 2.1 for detailed explanation.

Thus, we reoptimized the geometry of each monomer with constraints. This process

was applied to every target, including XXXI and XXXIII, where we employed flexible-monomer CSPs strategies. These reoptimized geometries were needed for improvements of aiFFs using the CCfC method. We selectively constrained the degrees of freedom pertaining to soft dihedral angles under the assumption that the CCDC probably avoided distorting these angles to prevent altering the shapes of monomers and leading to implausible crystal structures. For largest systems, we constrained all dihedral angles, aiming to improve computational efficiency. Our tests on a subset of monomers revealed that the energy differences resulting from these two constrained approaches were minor, in most cases not exceeding 1 kJ/mol.

In the development of the intermonomer aiFFs in stage 2, we further refined the potentials by using the CCfC approach. This involved generating new grid points, representing dimer configurations of the provided CCDC polymorphs, and integrating these into the existing training sets for refining of the fits. Given the goal to rank the crystal structures as provided by CCDC, we bypassed the structure generation steps for CSPs at stage 2. Instead, we modified the provided structures by replacing the original monomers with those that were optimized with constraints. Following this modification, we either minimized their lattice energies or conducted MD simulations, starting with these adjusted crystals.

For specific targets, namely XXVII, XXVIII, and XXXII, we optimized lattice energies using a rigid-monomer approach with the newly constrained-optimized monomers. To prepare for these optimizations, we embedded the optimized monomers into crystal lattices from the CCDC lists, aligning their centers of mass (COM) and the principal axes of inertia with those of the original monomer. Target XXVII was an exception, which is discussed in more detail in Sec. S2.1. For targets XXXI and XXXIII, our approach combined aiFFs for intermonomer interactions with enhanced empirical FFs for intramonomer interactions. This was applied in both flexible-monomer lattice-energy (0 K) minimizations and in MD simulations of crystals at various temperatures, starting with the constrained-optimized monomers. Enhancements to empirical FFs included updating hard geometric parameters (equilibrium bond lengths and angles) to those derived from the *ab initio* calculations of the equilibrium monomer and, in some cases, also adjusting the force constants to align with *ab initio* values for different monomer geometries. This marked a progression from stage 1, where we limited our approach to rigid-monomer CSPs for XXXI. On the other hand, for XXXII we opted for rigid-monomer CSPs in stage 2, despite using flexible-monomer ones in stage 1. The shift was due to the large size of monomer XXXII, which made enhancements impractical within the time limits of 7BT, and the unreliability of generic empirical FFs for accurate structure evaluation.

Additionally, we performed periodic dispersion-corrected DFT calculations (pDFT+D) on some systems, focusing only on single-point calculations (fixed crystal geometry, no optimizations). In these calculation, the lattice energies were obtained using the following formula:

$$E_{\mathrm{latt}} = \frac{E_{\mathrm{c}}}{Z} - E_{\mathrm{eq}}, \tag{4}$$

where E_{c} is the lattice energy of the unit cell, Z is the number of monomers per unit cell, and E_{eq} is the energy of the equilibrium monomer in the gas phase. These calculations were primarily used for verification purposes, although in certain cases,

they influenced our decision to prioritize some polymorphs higher on our submission list (see, individual sections).

Next, we discuss the improvements observed in our results after refining the methodology applied to each of the targets listed below.

4. Post-submission analysis

4.1. Target XXVII

4.1.1. Stage 1

We were unable to generate a polymorph close to the experimental crystal since rigid-monomer CSPs were only conducted for the global minimum conformer. The RMSD between this conformer and the monomer from Form A is 0.71 Å accounting for a large fraction of the allowed discrepancy. After submitting the prediction, we conducted rigid-monomer CSPs using conformer 7 (see Sec. S1.1.1), and the RMSD compared to the experimental monomer, equal to 0.60 Å, was the smallest among conformers, optimized by us, with an energy is 2 kcal/mol above the global minimum monomer. Although this RMSD is still large, our CSP gave a polymorph with $RMSD_{30}$ = 0.98 Å (tolerances 30%/30°, including only pentacene core nonhydrogen atoms), ranked as #45, which should be considered a good prediction for a system of that size. Thus, the approach of conducting rigid-monomer CSPs for each low-energy conformer would be successful for target XXVII. However, due to time constraints, we were unable to perform CSP utilizing conformers other than the global minimum one.

The large values of RMSDs do not fully explain the main reason for the poor match. The primary issue was the locations of the TIPS groups. The global minimum conformer has nearly linear C6-C12-C13-Si bonds, with a C12-C13-Si bond angle of 177° (the same value applies to the other TIPS group). The silicon atoms are nearly coplanar with the rings, and the isopropyl groups exhibit approximate symmetry relative to the rings, resulting in steric hindrance within the crystal. To reduce this hindrance, one of the TIPS groups should bend off the plane. In fact, the experimental monomer had a C12-C13-Si bond angle of 170° (176° for the other TIPS group) and a C6-C12-C13-Si dihedral angle of 45° (22°). The experimental monomer structure was also impacted by the dynamic disorder due to rotations of the TIPS groups (Hunnisett *et al.*, 2024*a*). Additionally, our conformers in stage 1 were optimized with a double-zeta basis set. When optimized using def2-TZVPP basis set, a conformer with an RMSD of 0.35 Å relative to the experimental monomer was obtained . This improvement can be attributed to the use of a larger basis set, which provides a more accurate description of the electronic correlation effect.

4.1.2. Stage 2

The composition of the CCDC-provided list is partly responsible for the good predictions for target XXVII (see Sec. S2.1.4). One reason is that the monomers in structures #28, #59, and #61 were very similar to experimental monomer and their geometries were minimally affected by our constrained optimizations. Additionally, the pentacene core structures remained close to their original positions within the CCDC polymorph during our lattice-energy minimizations. The original CCDC polymorphs (no opti-

mizations) #28, #59, and #61 were ranked 1, 5, and 3, respectively by our aiFF. The ranking of these polymorphs with the monomers replaced by constrained-optimized ones (and still no crystal structure optimizations) was even slightly better: 1, 5, and 2, respectively. Thus, our rigid-monomer optimizations actually made rankings worse (cf., Table 12).

4.2. Target XXVIII

4.2.1. Stage 1

We would have carried out CSPs with all eight of the low-energy conformers (see Sec. S1.2.1) that we found if not for the time limits. However, it would be challenging to locate the experimental structure to within the 7BT accuracy standards, even if we had done so. This is because the lowest RMSD for our conformers with respect to the experimental monomer (0.8 Å for conformer 2, the second-lowest energy conformer) was already very close to the limit of 1.0 Å for crystal's RMSD. In our post-submission CSPs, using this conformer, only 8 matching molecules were found within the standard tolerances, and the $RMSD_8$ was 1.19 Å. With eight soft degrees of freedom, the system appeared at this phase of our investigations to be too flexible for the conformer-based protocol. Clearly, the forces exerted by the other molecules in the crystal cause the monomers to undergo significant deformations compared to their gas-phase conformer geometries. In order to shed further light on this matter, we conducted CSPs utilizing the experimental monomer geometry, except that the experimental X–H bonds (X denoting any atom bound to hydrogen) were optimized using the aug-cc-pVTZ basis set and the PBE+D3(BJ) method. This procedure was necessary due to the substantial uncertainties in the positions of hydrogen atoms in experimental structures. We then executed the CSP procedure with this monomer. The experimental polymorph was identified at rank 1 with an $RMSD_{30}$ value of 0.184 Å, indicating that the flexibilized intermolecular aiFF is sufficiently accurate. This is further supported by our group's performance in stage 2, where we identified the experimental polymorph at rank 1 on the list of structures supplied by CCDC.

Next, we addressed the question why all our conformers are so different from the experimental monomer. We fully optimized this monomer to determine its closest conformer, which turned out to be conformer 2. The main difference between experimental monomer and conformer 2 is that only the former is centrosymmetric. The latter is not centrosymmetric due to Jahn-Teller's effect: lowering of symmetry removes orbital degeneracy and therefore lowers the energy of an unpaired electron. The experimental monomer can be centrosymmetric since interactions with other monomers in the crystal can remove degeneracy without need for the Jahn-Teller deformation. Thus, using conformer 2, we constructed a monomer by taking half of conformer 2 and obtaining the other half via the inversion symmetry operation. This monomer is just 1.4 kcal/mol above conformer 2 and has an RMSD of only 0.137 Å relative to the experimental monomer.

CSPs were then conducted using the centrosymmetric *ab initio* monomer, resulting in an experimental-like crystal at rank 1 with a $RMSD_{30}$ value of 0.292 Å. Therefore, target XXVIII actually does fall within the category of crystals suitable for examination via the conformer-based protocol, provided that symmetrized structures are

taken into account. Another possible explanation for our findings is dynamic disorder, where the monomers within the crystal oscillate between conformer 2 and its mirror image, causing the experiment to observe the average spatial arrangement (Hunnisett *et al.*, 2024*a*).

4.2.2. Stage 2

Our successful prediction of CCDC structure #144 at rank 1 (see Sec. S2.2.5) is partly due to remarkable similarity between structure #144 and the experimental crystal. Its monomer was extremely similar to the experimental one, with an RMSD of 0.21 Å, compared to crystal's $RMSD_{30}$ of 0.234 Å. The lattice-energy calculation, without optimization, yielded rank 1. Constrained optimization had minimal impact on this monomer, reducing the RMSD to 0.025 Åfor non-hydrogen atoms. Additionally, the high quality of our aiFF contributed significantly to our stage 2 prediction performance. We utilized a triple-zeta basis set (without diffuse functions), which is quite large for a monomer of this size. We also carefully eliminated the "holes" in the PES during the fitting process. Finally, a variety of monomer geometries were fitted with the stage 2 aiFF.

In addition, we conducted independent single-point pDFT+D calculations on the structures resulting from the lattice-energy minimizations with rigid monomers (using monomers that were constrained-optimized). The PBE+D3(BJ) functional (Perdew *et al.*, 1996; Grimme *et al.*, 2010), implemented in the QE package (Giannozzi *et al.*, 2009; Giannozzi *et al.*, 2017), was utilized. The respective cutoff energies for charge density and plane waves were 685.86 eV and 4435.46 eV. Although these calculations contributed to the validation of the stage 2 aiFF's convergence and accuracy, they were omitted from the submission. The rankings obtained from the pDFT+D calculations exhibited consistency with the aiFF-based calculations for the top two structures.

4.3. Target XXIX

4.3.1. Stage 1

With an RMSD of less than 0.001 Å, conformer A's overlap with the experimental monomer was perfect while conformer B's RMSD is 1.1 Å (see Sec. S1.3.5). Thus, monomer flexibility was not an issue for target XXIX. Therefore, rigid-monomer CSPs with $Z' = 1$–4 and monomers in the conformer A geometry should yield predictions as accurate as any flexible-monomer CSPs. However, we used both conformers A, B, and their mixture in our CSPs. The resulting best match with experiment among submitted polymorphs had $RMSD_{27} = 0.197$ Å (structure #927), so it was beyond the CCDC criteria for the match. However, it was close to the threshold and it was a pure A-monomer polymorph with $Z' = 3$, in agreement with experiment. The space group of #937 was $P2_1/c$, different from *Pbcn* of the experimental crystal. Although the match of structure #937 with the experimental data was reasonably satisfactory, its energy ranking, which was not considered by CCDC during the evaluation of the 1500 structures in submission 1, was not. A similar protocol based on the use of aiFFs in rigid-monomer CSPs was employed in ranking experimental-like crystals in a recent publication from our group (Nikhar & Szalewicz, 2022), yielding rankings of 16 or better, with the majority at rank 1.

To find out the reasons for high rankings of our best match polymorph, we repeated CSPs for target XXIX using UPACK and considering only homogeneous crystals of conformer A in $Z' = 1$, 2, and 3. We found the experimental polymorph at rank 68 in $Z' = 3$ with $\mathrm{RMSD}_{30} = 0.142$ Å in the *Pbca* space group. Such predictions would be among the best in 7BT. This result suggests some mistakes made in the 7BT work on this target. We have not analyzed our 7BT calculations, however, one reason for the worse predictions in the test could be the inclusion of monomer B, and in particular of mixed AB polymorphs. For the latter polymorphs, the AB aiFF was potentially insufficiently accurate, in particular at close-range configurations. CSPs using such aiFF can result in exploring regions that are too deep, consequently placing the resulting polymorphs high on the CSPs list. Such polymorphs might have pushed pure-A polymorphs matching well the experimental polymorph beyond the submitted list of 1500 polymorphs. The CCfC protocol can effectively address this issue, but due to time constraints, we were unable to apply CCfC for the AB aiFF.

We also considered the reason behind the reduced number of matching molecules observed as a result of the PXRD/EVCCPMRE refinement of the structures predicted by UPACK in the post-submission investigation. Although we recognize the difficulties associated with repurposing information from near-miss structures to generate an accurate experimental structure, we currently do not have a definitive answer to this question. These challenges might be overcome in the future by integrating EVCCPMC (Chan & Tuckerman, 2024) with crystal adiabatic free energy dynamics (CAFED) (Yu & Tuckerman, 2011).

As part of our post-analysis, we further explored the recycling of structural data from near-miss structures within the EVCCPMC framework. We categorized false positive structures—those that are not clear negatives but also not precisely the correct structures required to be identified—as either "near-miss" or "near-hit." Specifically, the experimental structure was constructed with a single EVCCPMC step by taking the EV directly from the experimental structure and using it to seed the EVCCPMC algorithm. Remarkably, structures with similarity scores as high as 29 out of 30 molecules ($\mathrm{RMSD}_{29} = 0.167$ Å) could be generated with only a few EVCCPMC steps when a particular near-miss EV was employed as a seed. These findings underscore the potential usefulness of incorporating PXRD similarity measures into the screening process.

4.4. Target XXX

4.4.1. Stage 1

In stage 1, the only stage for target XXX, we have correctly predicted the stiochiometry of the polymorphs. However, our best matches with experiment on the list of 1500 submitted crystals included only 11 molecules for form A and 10 for form B (see Sec. S1.4.5). The main reason for such performance was clearly the use of empirical intramonomer FFs.

In the post-submission analysis, we performed conformer-based rigid-monomer CSPs to find out how well such approach works for target XXX. We first evaluated the similarity between the conformers used in the development of the aiFFs and the monomers present in the experimental polymorphs. The lowest-energy conformer of

tetramethylpyrazine displayed excellent agreement with the monomers in both variations of the target XXX cocrystal, with RMSD values of 0.017/0.051 Å for the A/B forms. In contrast, the lowest local minimum conformer of cannabinol used in the aiFF developments, i.e., conformer 2, exhibited noticeable differences from the cannabinol monomer found in experimental polymorph A, with an RMSD of 1.219 Å (see Figure 7a of SI1-B of the stage 1 7BT paper (Hunnisett *et al.*, 2024*a*)). Most of these discrepancies stemmed from the pentane appendage, particularly evident when aligning the rings. On the other hand, conformer 2 bore a closer resemblance to one of the monomers in experimental polymorph B, as shown in Figure 7b of SI of the stage 1 paper(Hunnisett *et al.*, 2024*a*), with an RMSD of 0.421 Å. Again, the primary deviations were observed in the two terminal carbons of the pentane appendage, positioned at distances of 0.806 and 1.591 Å from the equivalent carbon atoms in the experimental monomer. For the other monomer of form B, the RMSD is 0.960 Å. Thus, the conformers employed in aiFF developments were reasonably close to experimental monomers in form B but not in form A. Our set of cannabinol conformers includes some closer matches than those of conformer 2: conformer 1 with RMSDs of 0.845 with respect to the monomer of form A and 0.372/0.826 Å for the monomers of form B.

We then performed conformer-based rigid-monomer CSPs with conformer 1 of T and 2 of C. The best predictions had ranks 23 for form A and 1 for form B, but the matches were poor. We also performed CSPs using the experimental monomer C geometry, with rationalized X-H bonds, taken from polymorph A. We found the experimental polymorph A at rank 17 with $RMSD_{30}$ = 0.231 Å (such analysis was not performed for polymorph B). Thus, the reason for the unsatisfactory outcome of the rigid-monomer CSPs was monomer C geometry and not the limitations of the CC aiFF discussed in Sec. S1.4.3. We have not investigated target XXX further, but the next step would be to perform rigid-monomer CSPs with the conformers of C. The results from aiFF-based CSPs may still be outside CCDC criteria, but if this step is followed by pDFT+D calculations on, say, top 100 polymorphs, the final predictions should be successful.

4.5. Target XXXI

4.5.1. Stage 1

We analyzed the generated conformers by comparing them with the experimental monomers. The global-minimum conformer C01 is significantly different from any experimental monomer, with the smallest RMSD of 1.81 Å compared to the monomer from the A_{min} polymorph. This is due to its relatively compact configuration, shown in Fig. S16. Conformers 2 and 4 (see Sec. S1.5.1), which have 'open' structures, closely matched experimental monomers. Conformer 4 aligns with the monomers from experimental structures B, A_{maj}, and C with RMSDs of 0.16, 0.31, and 0.37 Å, respectively (Fig. 2(b-d)). Conformer 2 aligns with the monomer from A_{min} with an RMSD of 0.277 Å (Fig. 2(a)). The relationships between conformers 2 and 4 and the monomers from A_{min} and A_{maj}, respectively, is anticipated due to both groups differing by the same rotation around the S1-C5 axis (see Fig. S16 for atom numbering) among the two conformers and the disorder forms. Conformers 2 and 4 exhibit small RMSDs with

respect to the experimental monomer compared to the CCDC threshold for crystals, indicating that the approach of rigid-monomer searches using low-energy conformers is likely to be effective for target XXXI. To understand why our method did not yield more matches and why the only match obtained was ranked so high (the ranking was not mandated by CCDC, but our list was ranked), we repeated rigid-monomer CSPs; the only difference compared to the CSPs used in the submission was that we replaced GAFF with OPLS-AA, both with aiFF charges, in the crude optimization steps of UPACK. The fine optimizations were conducted in the post-submission phase using the same aiFF as in the original CSPs. We first performed CSPs utilizing the rationalized monomer from the A_{min} polymorph. The rationalization involved optimizing the X-H bonds (with X representing any atom connected to a hydrogen) using the PBE+D3(BJ) method and the aug-cc-pVTZ basis set to improve positions of hydrogen atoms from experimental structures. These rigid-monomer CSPs led to the experimental structure being ranked second with an $RMSD_{30}$ of 0.204 Å. We repeated the procedures using conformer 2, which closely resembles the monomer from A_{min}. This resulted in a match to the A_{min} crystal with an $RMSD_{30}$ of 0.963 Å at rank 5, indicating a satisfactory prediction. The same steps were followed using the rationalized version of the monomer from structure B, resulting in the prediction of the B crystal form ranked 4 with an $RMSD_{30}$ of 0.241 Å. The crystal A_{maj} was also found with rank 2 and with an $RMSD_{30}$ value of 0.666 Å. Similar CSPs using conformer 4 resulted in A_{maj} rank of 3 with $RMSD_{30}$ of 0.630 Å and B at rank 24 with $RMSD_{30}$ of 0.361 Å. It appears that the change in the empirical FF and charges as well as small changes in the conformer geometry can impact the initial step of the CSP search such that these structures were not present in our original CSP submitted to the 7BT.

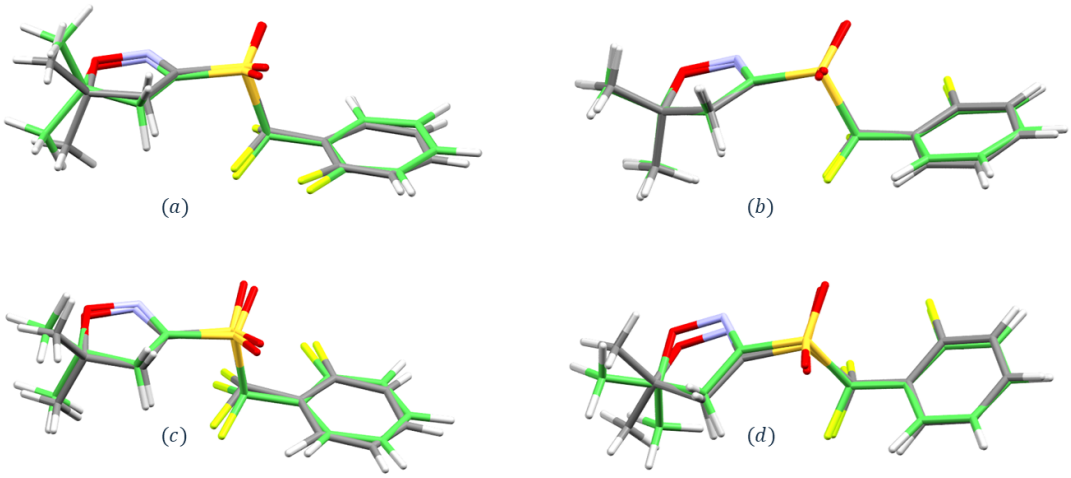

Fig. 2. Target XXXI: Monomer overlap for (a) Conformer C02 vs. monomer from Form A_{min}, RMSD = 0.277 Å. (b) Conformer C04 vs. monomer from Form B, RMSD = 0.16 Å. (c) Conformer C04 vs. monomer from Form A_{maj}, RMSD = 0.313Å. (d) Conformer C04 vs. monomer from Form C, RMSD = 0.365Å.

4.5.2. Stage 2

To understand why our predictions were unsatisfactory, we present in Table 4 the results of different methods used to rank the structures. Some calculations were done before submission, but they were considered less reliable than the submitted results due to not accounting for finite temperature effects (see Sec. S2.3.6). The first set of columns display results obtained by predicting lattice energies of the actual polymorphs provided by CCDC(no geometry optimization) using aiFF and DFT+D (PBE+D3(BJ) with aug-cc-pVTZ) to calculate monomer-deformation penalties. The rankings for A_{maj}/A_{min}/B improved from ranks 21/35/− to 2/22/14, where "−" means that the match was way beyond 7BT limits. The predictions were still better when the original monomers were replaced with constrained-optimized ones, resulting in ranks 1/5/3, which would have been one of the best prediction in the test for this target. This validates our constrained-minimization approach, and demonstrates the accuracy of our aiFF. Rigid-monomer lattice-energy minimizations slightly worsen the rankings, to 1/7/5, as shown in the third set of columns, but also slightly enhances all RMSDs. The poor predictions in the last columns, showing the submitted results from flexible-monomer approaches are most likely due to using the modified GAFF for intramolecular interactions and possibly not substituting the GAFF monomer-deformation energy penalties in the final rankings with *ab initio* values. This point is illustrated in Table 5, where the second set of columns displays relative lattice enthalpies under specific temperature and pressure conditions, whereas the first set of columns displays the lattice energies at $T = 0$ K for the structures obtained from the MD simulations, incorporating monomer-deformation penalties calculated using an *ab initio* approach. The polymorph resembling form A_{min} is ranked 1 for this set of conditions, the one resembling A_{maj} is ranked 15, and the one ranked first in the submission, which does not approximate form B due to its high $RMSD_{14}$, is unbound. Note that the energy comparisons in Table 5 are not rigorous because the first set of columns shows internal energies at $T = 0$ K, while the second column shows enthalpies averaged at $T = 150$ K and $P = 1$ atm in the MD simulations.

Table 4. *$RMSD_{30}$ and rankings for target XXXI. Tolerances of 25%/25° are used for matching criteria. The RMSD in the parentheses in the last column indicates that only 14 molecules were matching out of 30. The first two columns are single-point calculations, i.e., no crystal-geometry optimizations were performed. The crystal structures in the first column are exactly as provided by CCDC, whereas in the second column, the original monomers were replaced by constrained-optimized ones. The third column shows results from rigid-monomer UPACK optimizations. The fourth column shows the ranking according to the enthalpies averaged over fully flexible MD simulations at $T = 150$ K and $P = 1$ atm. In all calculations, aiFF was used to represent intermonomer interactions.*

	single-point unmodified		single-point constrained-optimized		rigid-mono opt constrained-optimized		MD fully flexible	
polymorph	rank	$RMSD_{30}$ [Å]	rank	$RMSD_{30}$ [Å]	rank	$RMSD_{30}$ [Å]	rank	$RMSD_{30}$ [Å]
A_{maj} (#98)	2	0.180	1	0.164	1	0.161	21	0.420
A_{min} (#1)	22	0.332	5	0.336	7	0.310	35	0.274
B (#25)	14	0.247	3	0.252	5	0.235	1	(1.953)
C (#89)	100	0.199	91	0.239	92	0.336	98	0.397

Table 5. *Crystal lattice energies and enthalpies of thermally averaged structures from MD simulations at $T = 150$ K and $P = 1$ atm. $E_{\rm latt}$ represents the lattice energy at 0 K, while $\Delta E_{\rm latt}/\Delta H_{\rm latt}$ denote the lattice energy/enthalpy relative to the global minimum polymorph, respectively. The lattice energies were computed during the post-submission work using the same aiFF as in the submitted list and with the same* ab initio *monomer penalties.*

Struct	$E_{\rm latt}$ (kJ/mol)	$\Delta E_{\rm latt}$ (kJ/mol)	Rank	Submitted $\Delta H_{\rm latt}$ (kJ/mol)	Rank
$A_{\rm maj}$ (#98)	-71.871	36.691	15	7.366	21
$A_{\rm min}$ (#1)	-108.562	0	1	9.933	35
B (#25)	34.329	142.892	43	0	1
C (#89)	1171.691	1280.253	87	21.489	98

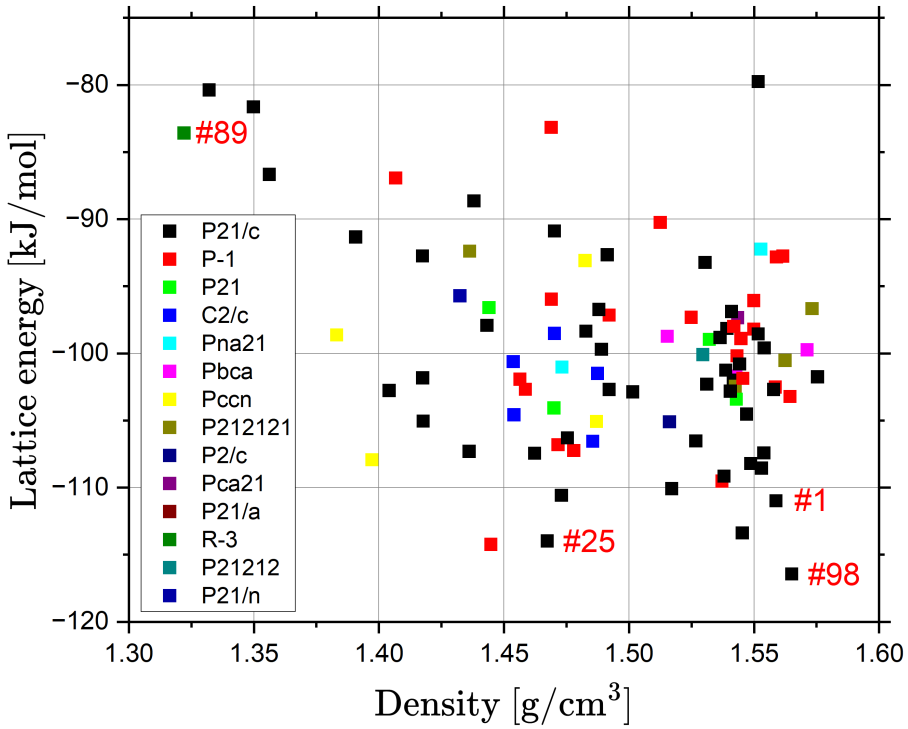

Fig. 3. Target XXXI: Lattice energy $E_{\rm latt}$ versus density from single point pDFT+D calculations for CCDC crystal structures with monomers replaced by constrained-optimized ones using UPACK. Structures corresponding to the experimental representatives: $A_{\rm maj}$ (#98), $A_{\rm min}$ (#1), B (#25), C (#89) are marked with their respective CCDC numbers given in parentheses.

Table 6 presents the top 20 rankings for both unmodified and constrained-optimized CCDC structures using stage 2 aiFF. The ranking in this table is based on energies calculated using UPACK (single point energy and rigid monomer optimization) and QE (Giannozzi *et al.*, 2009; Giannozzi *et al.*, 2017) for pDFT+D (PBE with plane

wave forms) lattice energies. The plane-wave energy cutoff was set at 1224.513 eV for wave functions and 8163.422 eV for charge densities. The Monkhorst-Pack method (Monkhorst & Pack, 1976; Pack & Monkhorst, 1977) was used for the k-point mesh with a sampling rate of 0.2 Å^{-1}. The crystals were ranked using pDFT+D calculations based on their single-point lattice energies. Figure 3 shows that ranks of the A_{maj}, A_{min}, and B crystals calculated with single point pDFT+D calculations of constrained-optimized CCDC crystal using UPACK were among top five (cf., Table. 6, which are also similar to the results from the second set of column of Table 4). With the exception of structure #89, all other three experimental representatives (#1, #25 and #98) are in the top 5 ranks with constrained-optimized structures. The pDFT+D lattice energies were calculated using the formula of Eq. (4).

Table 6. *System XXXI lattice energies of the top 20 ranked constrained-optimized and original CCDC structures using stage 2 aiFF and pDFT+D. pDFT+D lattice energies were computed using the same geometries as preceding UPACK calculations. All energies are in kJ/mol.*

	Original CCDC Structures				Constrained-Optimized CCDC structures							
					Original CCDC crystals structures				UPACK rigid-monomer optimized structures			
	UPACK		pDFT+D		UPACK		PDFT+D		UPACK		PDFT+D	
Rank	Struct. No	Energy	Struct. No	Energy	Struct. No	Energy	Struct. No	Energy	Struct. No	Energy	Struct. No	Energy
1	70	-127.368	59	-137.139	**98**	-116.428	59	-132.944	**98**	-119.687	**98**	-134.961
2	**98**	-126.438	**98**	-135.925	53	-114.234	**98**	-132.314	57	-117.257	59	-134.294
3	73	-126.018	34	-134.952	**25**	-113.971	**25**	-131.566	53	-117.093	**1**	-133.875
4	53	-125.317	**25**	-134.794	70	-113.373	**1**	-131.056	70	-116.981	**25**	-132.88
5	32	-125.290	57	-134.704	**1**	-110.982	52	-130.312	**25**	-115.069	57	-132.602
6	11	-123.951	11	-134.443	24	-110.569	24	-129.912	59	-114.670	70	-130.951
7	12	-123.807	22	-133.545	38	-110.072	58	-129.220	**1**	-114.482	24	-130.832
8	34	-123.627	32	-133.222	28	-109.522	20	-128.569	38	-114.197	71	-130.248
9	57	-123.546	70	-132.906	57	-109.144	53	-128.506	20	-113.454	53	-130.247
10	38	-122.738	39	-132.659	59	-108.543	75	-128.323	63	-113.081	52	-130.124
11	28	-122.553	52	-132.625	20	-108.210	85	-128.277	28	-112.421	36	-129.993
12	48	-122.077	24	-131.656	15	-107.929	70	-128.091	30	-112.207	38	-129.866
13	63	-121.436	38	-131.375	91	-107.443	36	-128.009	24	-111.780	99	-129.821
14	**25**	-121.396	20	-131.253	36	-107.416	28	-127.687	73	-111.344	28	-129.798
15	59	-120.698	63	-130.899	52	-107.303	71	-127.562	36	-111.260	75	-129.779
16	17	-120.240	99	-130.821	75	-107.242	15	-127.562	26	-111.095	20	-129.778
17	20	-120.149	85	-130.796	47	-106.799	57	-127.394	75	-110.971	58	-129.759
18	4	-119.845	10	-130.777	58	-106.548	5	-127.290	86	-110.873	18	-129.744
19	8	-119.727	48	-130.762	18	-106.525	37	-127.125	47	-110.737	34	-129.693
20	45	-119.036	4	-130.594	86	-106.300	29	-126.998	23	-110.553	85	-129.655

4.6. Target XXXII

4.6.1. Stage 1

The configuration of target XXXII in the room temperature experimental form B was 'flat', identified as a high energy transition state from the vapor phase UFED simulations. When the initial ranking was performed using GAFF, the rank #1 structure was also flat (see Sec. S1.6.1). A comparison of the monomers from our submitted lists with those from the experimental structures found the lowest RMSDs between the monomers equal to 0.65 (for #87), 0.87 (for #528), and 0.81 (for #225) Å for forms B, A_{maj}, and A_{min}, respectively. To evaluate the quality of our FFs, we performed a lattice-energy optimization with UPACK starting from the experimental form B (space group $P2_1/c$, room temperature). Initially, GAFF was used for both the inter- and intramonomer interactions. This CSP resulted in a packed structure and unit cell that was similar to the experimental one; however, the molecular geometry differed significantly: the similarity between the experimental and GAFF-optimized structures

was only 13/30 molecules with an $RMSD_{13} = 0.581$ Å. (The GAFF lattice energy was −547.170 kJ/mol, giving it a relative ranking #163 within the intermediate structure pool of EVCCPMRE/GAFF generated structures). The overlap of these structures is shown in Fig. 11 SI1-B of the stage 1 7BT paper (Hunnisett *et al.*, 2024*a*). Although this match is below CCDC criteria, it is significantly better than any match on the submitted list. Thus, possible inadequacies of GAFF alone cannot explain the poor results. One explanation could be that our search protocol was unable to find a crystal structure similar to form B. Another possibility is that the final optimizations with aiFF worsened the structures. To check this, we performed a rigid-monomer reoptimization of the GAFF-optimized form B using the aiFF, which changed the structure insignificantly: $RMSD_{30} = 0.087$ Å with respect to the GAFF-optimized structure. Interestingly, the rank improved: the lattice energy was −289.855 kJ/mol from aiFF combined with −220.883 kJ/mol GAFF-based intramolecular component to give −510.738 kJ/mol, a relative rank of #69 within the submitted structures. Note that these two ranks discussed above cannot be directly compared. Thus, the aiFF quality is not the reason for the poor predictions.

4.6.2. Stage 2

The structure #232 which was the representative of form B_{LT} was not ranked due to a error in running UPACK for crystals with $Z' > 1$ (see Sec. S2.4.5). After submission, we reran the process for such crystals, obtaining rank 30 with $RMSD_{30} = 0.334$ Å (tolerances 25% and 25°). The UPACK rigid-monomer optimization significantly improved the rankings compared to the original CCDC structures, where monomers were replaced by constrained-optimized ones. The rankings improved from 116 to 23 for form A_{maj} and from 241 to 30 for form B_{LT}, indicating the high quality of our aiFF. During the test period, we also calculated the lattice energies of the crystals using pDFT+D. We performed these calculations using the PBE+D3(BJ) functional as implemented in the VASP software package (Kresse & *et al.*, 2021). These were single-point calculations on the CCDC polymorphs with monomers replaced by the constrained-optimized ones. The plane-wave energy and charge density cutoffs used were 600 eV and 4,800 eV respectively. The k-spacing was 0.15 $Å^{-1}$. This process aimed to assess how the aiFF-based UPACK minimization method's ranking correlates with the pDFT+D ranking. The correlation was low, leading us to conclude that the aiFF-based results were more reliable. Thus, pDFT+D calculations had no impact on our submission. However, the pDFT+D rankings turned out to be better. The experimental structures A and B were ranked by pDFT+D as numbers 12 and 13, respectively. When similar pDFT+D calculations were performed on the structures after UPACK minimizations (conducted for #232 after the submission), the rankings were 2 and 9, respectively. In comparison. the aiFF-based rankings for the same structures were 23 and 30. Such an improvement by pDFT+D calculations aligns with the aiFF@CSPs protocol (Nikhar & Szalewicz, 2022), which includes pDFT+D single-point calculations on structures from rigid-monomer UPACK minimizations.

4.7. Target XXXIII

4.7.1. Stage 1

The comparison between the experimental monomers and the *ab initio* optimized conformers (see Sec. S1.7.1) revealed that the cations were virtually identical, with RMSDs of 0.016 (0.015) Å for polymorphs A (B). However, for the anions, the RMSDs for the four consecutive energy-ordered conformers were 0.092 (1.026), 1.355 (0.854), 1.079 (0.273), and 0.875 (1.052) Å for polymorphs A (B). Unfortunately, our predictions were inaccurate due to the substandard quality of the SS aiFF, which resulted from a flaw in the autoPES procedure. Our post-submission evaluation revealed that the SS aiFF contained 'holes' due to an oversight in the design of the grid generation and hole-search algorithms of autoPES, which did not account for the possibility of two identically charged monomers. Specifically, for purely repulsive same-charge monomer surfaces, the grid generation algorithm's decreasing grid point density with increasing positive interaction energy led to insufficient sampling at critical separations crucial for crystal structures. Moreover, the hole-search algorithm, which assumes the existence of radial minima on the scans of a PES, failed to detect the presence of holes due to the absence of minima on surfaces with the same charge. The very large RMSE for all grid points actually indicated the presence of these holes, but accurate PESs occasionally produce a significant number of grid points at extremely close intermonomer separations (with closest-contact atoms separated by distances well below the sum of their van der Waals radii), so this warning was ignored. During stage 2, the issues were fixed by taking into account the same-charge characteristic of the PES.

4.7.2. Stage 2

For target XXXIII, the experimental representative polymorphs closely resembled the experimental crystals, with $RMSD_{30}$ values of 0.387 and 0.254 Å for forms A and B, respectively. Our optimizations of these CCDC-provided polymorphs, including separate constrained minimization of monomers, marginally improved the value for form A to 0.322 Å, but worsened it for form B to 0.506 Å. Nonetheless, both final structures demonstrated reasonable $RMSD_{30}$ values. However, our energy rankings were notably poor (see Sec. S2.5.7).

In our post-submission analysis, we obtained better rankings: form A at 201 and form B at 24, based on lattice energies computed in the same manner as in the submission but for the exact original CCDC-provided structures. Consequently, our significant efforts in generating tailor-fit GAFF, performing UPACK minimizations, and MD simulations were counterproductive. Nonetheless, even these rankings were not accurate enough, warranting further investigation on how to improve our approach.

Additionally, the experimental polymorphs A and B were ranked at 123 and 1, respectively, in our pDFT+D calculations. These calculations were performed on the CCDC-provided structures with the monomers replaced by constrained-optimized monomers. Remarkably, despite using the lowest level of pDFT+D: the non-hybrid PBE functional, a coarse k-point mesh, and loose plane-wave cutoffs, we achieved excellent results for form B. However, the ranking of form A was poor, rendering these results partly fortuitous. This suggests that for target XXXIII, the many-body polarization effects may be significant given the ionic nature of the monomers. Such effects were small in previous research by our group (Nikhar & Szalewicz, 2022); however none of the monomers considered in that study were ionic. Consequently, autoPES can generate polarizable aiFFs which can be used in UPACK.

5. Summary and discussion

The 7BT stages 1 and 2 were designed to test the cutting-edge methodologies in the fields of structure generation and energy rankings of CSPs, respectively. Our team dedicated significant time and effort in both of these stages, attempting all test targets and responding to all challenges for each target. Although our submissions' agreements with experimental crystals in stage 1 were not satisfactory, we argue that the methodology developed, over the course of and after the end of 7BT, works reasonably well. Specifically, we show that in stage 1, the use of rigid-monomer CSPs with flexibilized aiFFs for a number of low-energy conformers would identify experimental polymorphs near the top of the ranking list. In stage 2, our flexibilized aiFFs ranked experimental polymorphs of two of the three largest targets in the top 5, and the post-submission analysis found 5 (8) out of 9 experimental polymorphs within the top 5 (30). In the previous sections, we provided a detailed target-by-target description of these CSPs. In this summary, we will examine them from a broader perspective. We can categorize our CSPs into two groups: those involving rigid monomers and those involving flexible monomers. The factors influencing the quality of predictions varied significantly between these two groups, therefore we will organize our summary accordingly.

5.1. Conformer analysis

As discussed in the Introduction, rigid-monomer predictions can be successful only if the geometries of monomers that were used in the CSPs are reasonably close to those of the monomers in experimental polymorphs. Table 7 presents comparisons of these geometries. It lists the energies and monomer RMSDs (in parentheses) for the monomer geometry used in the aiFF development and for the conformer that has the lowest RMSD with respect to the experimental monomer. Let us assume that an RMSD below 0.6 Å is sufficiently small for CSPs that can potentially produce polymorphs with $RMSD_{30}$ below 1 Å. Then the equilibrium conformer is in this category for 7 cases out of 17 (counting two different monomers in XXX B (C)). If one considers low-energy conformers (with energies up to 10 kJ/mol above the equilibrium conformer), 13 out of 17 cases are in this category. Thus, the conformer-based rigid-monomer CSPs could potentially be successful for all targets but XXX and XXXII. For the former case, monomer C has an aliphatic appendage that can form a large number of close-energy configurations. In fact, the three forms of this appendage found in experimental polymorphs are all quite different geometrically. We found conformers with RMSDs below 0.85 Å relative to experimental monomers, so the conformer-based method can still work if the top few dozens of polymorphs are reoptimized with pDFT+D. Hence, only target XXXII does not appear to be feasible with conformer-based rigid-monomer CSPs. For this large molecule with 11 soft degrees of freedom, a couple dozens of low-energy conformers have folded structures, whereas the experimental monomers have open structures and conformers similar to these monomers are almost 200 kJ/mol above the equilibrium monomer. One could account for this effect by eliminating the folded monomers, but this would result in a fairly complicated protocol.

Despite the fact that our conformer-based protocol should be able to provide good predictions for 8 out of 12 polymorphs, only one of our predicted polymorphs satis-

fied CCDC criteria. There were three reasons we did not do better: (*i*) For 3 targets: XXX, XXXII, and XXXIII, we performed flexible-monomer CSPs with empirical intramonomer FFs which we now know are not accurate enough, see Sec. 5.3 below; (*ii*) for 2 targets: XXVII and XXVIII, we used only the equilibrium conformer which is too different from the experimental monomer to provide predictions within CCDC limits; (*iii*) thus, bona fide conformer-based CSPs were performed only for targets XXIX and XXXI. For the former target, the use of the second conformer actually backfired since the experimental polymorph includes only one monomer geometry, which is very close to the equilibrium conformer. Our best predictions was actually quite close to the experimental polymorph, with $Z' = 3$ and $\mathrm{RMSD}_{27} = 0.197$ Å. It was, however, a structure with a mixture of conformers in the asymmetric unit with conformer ratio A:B = 2:1. One can suspect that the reason for this miss is the fact that we have not performed CCfC refinement for the AB aiFF and it contains some too deep wells at short intermonomer separations. The second target for which we performed conformer based CSPs, with 8 conformers included, was target XXXI. This was the only target for which one of our predicted polymorphs matched an experimental polymorph to within CCDC criteria. However, even here, we were able to match only one of the disordered structures.

The potential of the conformer-based method was demonstrated in our post-submission work where we predicted 6 polymorphs out of 12 within CCDC criteria (see Sec. 5.5 below). The ranks of these polymorphs were also very good, ranging from 1 to 45. This group includes targets XXVII, XXVIII, XXIX, and XXXI. As stated above, the conformer-based method in the form applied here is not expected to work for targets XXX and XXXII. The method did not work for target XXXIII for reasons unrelated to monomer flexibility: our intermonomer aiFFs miss many-body effects that are significant for this target. Our conformer-based approach should be able to predict this target if polarizable aiFFs are used. A special note is required for target XXVIII. The monomer listed in Table 7, that differs from the experimental monomer by 0.137 Å, is not a conformer but was obtained by constructing a centrosymmetric monomer from conformer 2. It is not a conformer since it does not represent a local minimum (it is possible that there is a centrosymmetric monomer beyond the 10 kJ/mol range considered by us, but within this range the conformer closest to the experimental monomer has RMSD of 0.8 Å and is not centrosymmetric).

Table 7. *Comparison of* ab initio*-derived conformers to experimental monomers. For each polymorph, we list the conformer or monomer energy (in kJ/mol) with respect to the equilibrium conformer and RMSD (in Å) with respect to experimental monomer(s), nonhydrogen atoms only. These quantities are provided for the conformers/monomers used in the aiFF developments and for the conformers/monomers that show the lowest RMSD with respect to the experimental monomer(s).*

Target	Polymorph	Conformer/monomer used in aiFF fitting	Closest conformer or monomer
XXVII	A	0.00 (0.710)	8.37 (0.600)
XXVIII	A	0.00 (2.100)	5.85 (0.137)[†]
XXIX	A	0.00 (0.001)	0.00 (0.001)
XXX	A (T)	0.00 (0.017)	0.00 (0.017)
	A (C)	1.14 (1.219)	0.00 (0.845)
	B (T)	0.00 (0.051)	0.00 (0.051)
	B (C)[‡]	1.14 (0.421, 0.960)	3.44 (0.372, 0.827)
XXXI	A_{maj}	0.00 (2.121)	5.19 (0.313)
	A_{min}	0.00 (1.788)	3.86 (0.277)
	B	0.00 (2.195)	5.19 (0.160)
XXXII[§]	A	453.80 (2.560)	187.35 (0.873)
	B	453.80 (2.570)	183.89 (0.646)
XXXIII	A (M)	0.00 (0.016)	0.00 (0.016)
	A (S)	0.00 (0.092)	0.00 (0.092)
	B (M)	0.00 (0.015)	0.00 (0.015)
	B (S)	0.00 (1.026)	7.00 (0.273)

[†] Monomer obtained by symmetrization of conformer 2 to remove the Jahn-Teller deformation.
[‡] This polymorph with $Z' = 2$ has two distinct C conformers.
[§] As discussed in the text, the conformer used in aiFF development later turned out to be substantially above the equilibrium conformer.

5.2. Accuracy of intermonomer aiFFs

The performance of the aiFFs on subsets of close-range grid points is shown in Table 8. This table includes data for both stages, except for targets XXIX and XXX, for which only stage 1 was conducted. In stage 2, the aiFFs were refitted by adding dimers extracted from the CCDC-provided polymorphs to the training set used in

stage 1.

Table 8. *RMSEs of aiFFs evaluated on subsets of the close-range grid points are given in kcal/mol. Numbers of grid points in the subsets are given in parentheses. The number of free parameters in the close-range fitting stage is denoted* N_{FP}*, the total number of grid points by* N_{grid}*, and the number of detected local minima of the aiFF by* N_{min}*. A dash '-' indicates no data points were available in the energy range.*

							RMSE	
Target	Stage	aiFF	N_{grid}	N_{FP}	$N_{\mathrm{grid}}/N_{\mathrm{FP}}$	N_{min}	$E_{\mathrm{int}} < 0$	$E_{\mathrm{int}} < 10$ kcal/mol
XXVII	1	$\overline{\mathrm{AA}}$	12266	929	13.20	599	0.69 (7570)	0.91 (9637)
	2	$\overline{\mathrm{AA}}$	15390	929	16.57	728	0.66 (10691)	0.83 (12760)
XXVIII	1	AA	6282	504	12.46	156	0.43 (2502)	0.56 (3144)
	2	$\overline{\mathrm{AA}}$	10249	504	20.34	251	0.47 (5077)	0.65 (6146)
XXIX	1	$\overline{\mathrm{AA}}$	6465	460	14.05	90	0.14 (5290)	0.22 (6058)
		$\overline{\mathrm{BB}}$	6957	460	15.12	85	0.15 (5043)	0.23 (6073)
		AB	14293	880	16.24	58	0.15 (10153)	0.24 (12046)
XXX	1	$\overline{\mathrm{TT}}$	896	28	32.00	7	0.17 (769)	0.23 (883)
		CC	5471	340	16.09	292	0.47 (2885)	0.58 (3748)
		TC	2463	178	13.84	70	0.39 (1701)	0.53 (2039)
XXXI	1	AA	10672	378	28.23	92	0.22 (6605)	0.28 (8391)
	2	$\overline{\mathrm{AA}}$	14958	378	39.57	213	0.25 (10213)	0.33 (12986)
XXXII	1	AA	11148	460	24.23	230	0.48 (7558)	0.73 (10257)
	2	AA	13852	460	30.11	467	0.63 (9793)	0.89 (12992)
XXXIII	1	MM	4500	180	25.00	0	-	-
		SS	10000	418	23.92	0	-	-
		MS	13879	518	26.79	19	1.52 (8715)	1.65 (9058)
	2	$\overline{\mathrm{MM}}$	8496	180	47.20	0	-	0.02 (41)
		$\overline{\mathrm{SS}}$	12996	418	31.09	0	-	0.62 (37)
		$\overline{\mathrm{MS}}$	19899	518	38.22	19	1.46 (18507)	1.57 (18925)

The ratio $N_{\mathrm{grid}}/N_{\mathrm{FP}}$ was always larger than 10 for all fits, ensuring enough grid points per free parameters were used in the developments to avoid overfitting. The RMSEs for negative interaction energies was less than 0.7 kcal/mol for all targets except target XXXIII. As discussed in Sec. S1.7.3, for target XXXIII, the RMSEs for MS aiFF appear large but correspond to less than a 2% error in the global minimum region. The aiFFs MM and SS are repulsive everywhere since they consist of identically charged monomers; thus, in the table these entries are marked with dashes. Also, in stage 1 there were no grid points with energies below 10 kcal/mol, while in stage 2 there were just a couple dozens points. As shown in Table 8, RMSEs of stage 2 aiFFs are slightly larger than those from stage 1, except for targets XXVII and XXXIII where they decreased a bit. However, the performance on dimers extracted from CCDC crystals improved dramatically. This improvement is illustrated in Table 9 for target XXXIII. The RMSEs in this table are calculated using stage 1 and stage 2 aiFFs. For MM, there is an 18-fold decrease of the RMSE. For SS, the improvement was less dramatic, but still substantial, reducing the RMSE to a third of its value from stage 1. Finally, for MS, even though the stage 1 aiFF was already relatively accurate, the

RMSE became 2 times smaller. Similar trends of improvement were observed across other targets in stage 2.

Table 9. *Target XXXIII: RMSEs (in kcal/mol) of stage 1 and stage 2 aiFFs for a set of dimers extracted from the CCDC-provided polymorphs.*

aiFF	# of dimers	Stage 1 RMSE	Stage 2 RMSE
MM	3996	3.78	0.21
SS	3996	4.00	1.38
MS	6020	1.79	0.92

5.3. Accuracy of intramonomer FFs

When 7BT began, our team believed that older, mostly unsuccessful predictions with empirical FFs were primarily due to the poor quality of their intermonomer component. We thought that replacing this component with aiFFs would result in reliable CSPs, a method referred to as ai-inter/emp-intra. This belief was largely based on our groups' previous successful use of such a mixed approach (Reilly *et al.*, 2016; Metz *et al.*, 2022). However, in those cases, the monomers were significantly more rigid than most of the monomers in 7BT. During stage 1, we realized that the ai-inter/emp-intra approach does not work well for the 7BT targets. Therefore, in stage 2, we attempted to improve the empirical component by refining some parameters of these FFs using *ab initio* values for the equilibrium parameters for target XXXI. For XXXIII, we also reoptimized force constants and torsional parameters by fitting them to a set of *ab initio* monomer energies. A comparison of quality of empirical and refined empirical intramonomer FFs is presented in Table 10.

Table 10. *RMSEs (in kcal/mol) of stage 1 and stage 2 intramolecular FFs used in flexible-monomer CSPs for target XXXIII. The RMSEs are computed with respect to the* ab initio *monomer energies used in the refinement of stage 2: 3,001 and 2,000 points for monomers M and S, respectively.*

Stage	Monomer	Intramolecular FF	RMSE
1	M	GAFF	1.07
	S	GAFF	4.19
2	M	tailor-fit GAFF	0.46
	S	tailor-fit GAFF	2.11

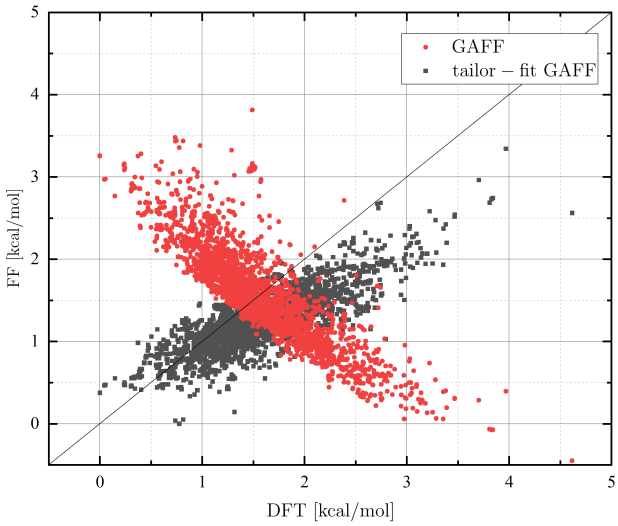

Fig. 4. Target XXXIII: Performance of GAFF and of stage 2 tailor-fit GAFF for monomer M.

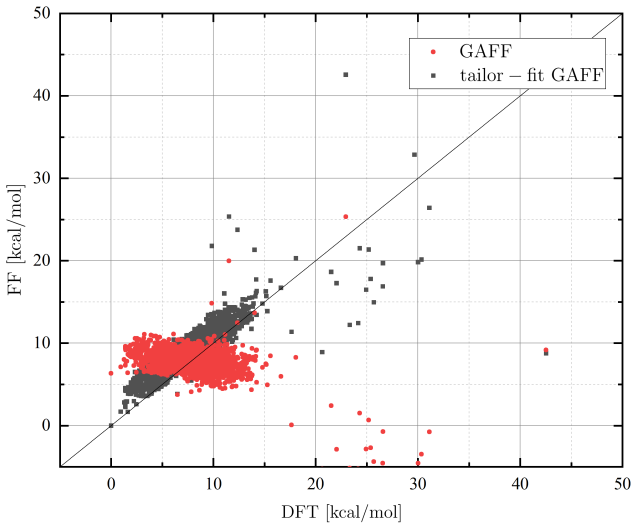

Fig. 5. Target XXXIII: Performance of GAFF and of stage 2 tailor-fit GAFF for monomer S.

Table 10 shows that tailor-fitting of GAFF lead to reductions of RMSEs by 2.32 and 2 times for M and S monomers, respectively. Figures 4 and 5 present scatter plots of DFT and FF energies calculated for monomer M and S, respectively. Note that the original GAFF energies were shifted to minimize the RMSE and such shifted energies are shown on the figures. Although the reductions of the RMSEs due to fitting are not particularly large, the figures show that the overall correlation between the FF and *ab initio* energies improved significantly.

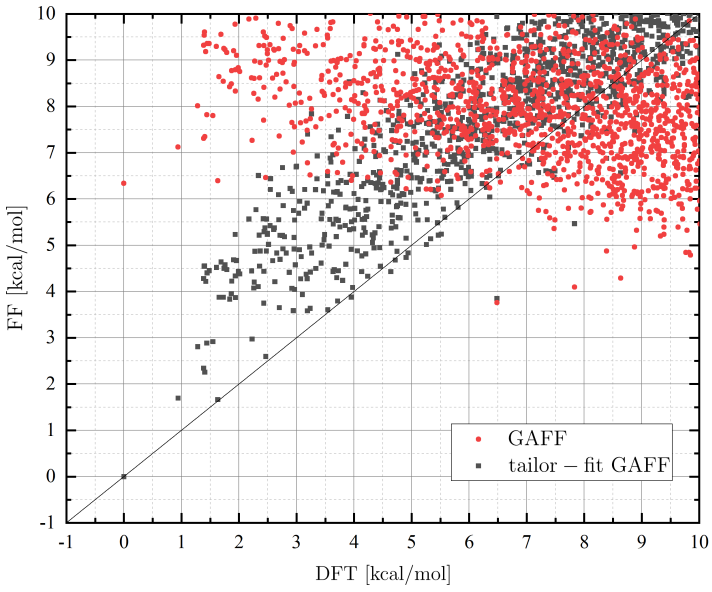

Fig. 6. Target XXXIII: Performance of GAFF and of stage 2 tailor-fit GAFF for monomer S in low-energy region. RMSEs in this region are are 3.08 and 1.84 kcal/mol, respectively.

Despite the improvements due to fitting, the remaining RMSEs are about 2 (M) and 8 (S) kJ/mol, too large for CSPs since the energy differences between polymorphs are of the order of 1 kJ/mol. One might hope that the performance is better in the low-energy region. Figure 6 shows such performance for monomer S. As seen in this figure, the tailor-fit GAFF does not reproduce low-energy *ab initio* points for the monomer accurately. In fact, there is a systematic upward shift of the FF values by a couple kcal/mol. Since monomer S has flexible torsional degrees of freedom, such low-energy regions are sampled during CSPs. These torsions generally change energies within 20 kJ/mol (5 kcal/mol) of the global minimum configuration. Hence, accurately reproducing low-energy regions is important for CSPs and even the tailor-fitted GAFF does not seem to be sufficiently accurate for CSPs.

Since the performance of the original empirical intramonomer FFs for other targets for which we used flexible-monomer CSPs is similar to the performance for target XXXIII discussed above, the overall poor predictions for these targets can be attributed to this problem. Even with the tailor-fit GAFF used for target XXXIII in stage 2, the accuracy of the intramonomer part was probably insufficient for accurate enough predictions. We say "probably" since the missing many-body effects were another factor reducing the quality of predictions. The improved GAFF for target XXXI in stage 2 was still less accurate than the tailor-fit GAFF for XXXIII, and indeed the predictions with flexible monomers were poor.

5.4. CCDC polymorphs of stage 2

Most CSP protocols are divided into two stages: the set of polymorphs resulting from the structure generation stage becomes the pool for the ranking stage. Accordingly, in

stage 2 , CCDC provided participants with pools of 100 or 500 polymorphs, depending on the target. At the time of the test, no information was available on how they were constructed. There were two features of these pools that made them significantly different from real-life pools: while some polymorphs had monomers with energies relative to the *ab initio* equilibrium conformer significantly larger than expected from typical structure generation steps, some other polymorphs, including their monomers, were so close to experimental ones that their structures were well within CCDC criteria for successful predictions. The latter polymorphs are called by CCDC "experimental-representative structures". The first feature was found already in our initial analysis, as discussed in the Introduction. This led us to perform flexible-monomer CSPs for some targets. The reason for the excessive monomer energies was that the polymorphs selected from stage 1 submissions were optimized with constraints (Hunnisett *et al.*, 2024*b*) with CCDC's own empirical FF (Cole *et al.*, 2016), which apparently is not accurate in the intramonomer sector. This optimization was performed to "anonymize" the structures, so that the original group that submitted the structure could not immediately identify it as their entry. Note that monomers with energies of several kJ/mol above the equilibrium monomer energy cannot form crystals since monomer energy penalties would be larger than binding effects of intermonomer interactions. We were unaware of the second feature until experimental structures were revealed. Apparently, the experimental representatives were not optimized with the CCDC field (or perhaps optimized with very tight constraints), otherwise their energies would not remain so close to the energies of the equilibrium monomers, see below. Had we known this, we would have just performed single-point calculations for the set of CCDC polymorphs. This procedure, as we will show, leads to much better predictions than our flexible-monomer optimizations. Such a pool of polymorphs is not what one would expect in real-world CSPs, where the structure generation stage would have been performed with a high-quality empirical FF, a very inexpensive approach, and the energies of all monomers would be in a fairly narrow range of a few dozen of kJ/mol. There definitely would not be any polymorphs as close to experimental ones as those in the CCDC sets. The reason such polymorphs were present is that in stage 1 several groups submitted structures which could not be considered as resulting from a structure generation protocol only, but rather were a result of a comprehensive, expensive CSPs, representing the best possible attempts by a group. Nobody expects real-world structure generation methods to produce polymorphs close enough to the experimental polymorph to satisfy the 7BT criteria of similarity, as it was the case for experimental representatives in the pools. One can make here a point that the ranking pools of 7BT benefited one type of methods over other types. In particular methods based on the total lattice energy, such as pDFT+D methods, which can only do local minimizations, work well only if the starting configuration is reasonably close to the minimum configuration. Thus, the application of pDFT+D to the experimental representatives in 7BT was a guaranteed success. The polymorphs with high lattice energies were just ending up at far places of the submitted lists. In this section, we will present evidence for these statements. We will first discuss the monomers of the polymorphs on CCDC lists and then compare the experimental representatives to

experimental polymorphs.

Table 11. *Comparisons of RMSDs (in Å) of monomers from CCDC representative crystals relative to the corresponding experimental monomers (only non-hydrogen atoms) and of energies (in kJ/mol) of these monomers relative to the energies of the corresponding* ab initio *equilibrium monomers. In the column labeled "original", the results are shown for monomers from the representative structures, with structure number in parenthesis. In the column "constrained-optimized", similar results are provided for the corresponding monomers after constrained minimizations. The methods and basis sets are consistent with those specified in Table 2.*

Target	Polymorph	original RMSD/ΔE	constrained-optimized RMSD/ΔE
XXVII	A (#28)	0.260/15.302	0.262/1.097
	A (#38)	0.229/72.192	0.510/27.348
	A (#59)	0.261/14.717	0.260/0.592
	A (#61)	0.285/13.977	0.287/-0.394*
XXVIII	A (#144)	0.021/30.198	0.025/21.475
XXXI	A_{maj} (#98)	0.083/17.539	0.070/8.700
	A_{min} (#01)	0.166/12.899	0.176/5.481
	B (#25)	0.061/10.011	0.062/6.348
	C (#89)	0.001/554.93	0.132/6.208
XXXII	A (#317)	0.070/42.367	0.089/36.260
	B (#232)	0.095/48.000	0.106/43.300
XXXIII	A (#233) M	0.024/9.200	0.016/0.200
	A (#233) S	0.110/21.922	0.107/1.895
	B (#452) M	0.018/13.60	0.014/0.080
	B (#452) S	0.034/40.984	0.067/13.044

*The reason for the negative energy difference is that the reference structure is the global-minimum monomer of stage 1, optimized with the def2-SVP basis set. However, all the energies for monomers of target XXVII and the reference monomer energy used here were evaluated in the def2-TZVPP basis set. In the latter basis, the reference monomer does not represent the minimum structure.

We will first discuss the monomers of the experimental-representative polymorphs. Table 11 lists RMSDs of these monomers relative to monomers found in experimental polymorphs. Additionally, it shows the energies of these monomers relative to those of the corresponding *ab initio* equilibrium monomers. Similar comparisons are also made with the constrained-optimized monomers. The RMSDs range between 0.0 and 0.1 Å for 9 out of the 15 monomers. For the remaining 6 cases, the RMSDs ranged between 0.1 and 0.29 Å. Moreover, in the majority of cases, 9 out of 15, the energies of the monomers from CCDC experimental representatives fall within the 9–22 kJ/mol range, while for the additional 4 monomers the range is 30–48 kJ/mol. This contrasts with the values for the complete CCDC set which extended into hundreds of kJ/mol, as

shown in Table 2. We will now examine the 5 cases with energies above 40 kJ/mol. The larger values for target XXXII could be expected, given that the equilibrium monomer exhibits a folded configurations while the experimental representative monomers are nearly planar. For monomer S of the target XXXIII polymorph #452, the factors contributing to the relatively large energy are inaccurate X–H bond lengths and the essentially flat NH_2 configuration. After the constrained optimization of this monomer, the RMSD underwent only a minor change, while the energy was reduced to 13 kJ/mol. So perhaps this monomer was taken from an experimental crystal. The largest energy differences are seen for monomers #38 of XXVII (72 kJ/mol) and #89 of XXXI (555 kJ/mol). The former case was an unintended experimental representative and can be disregarded. The latter monomer has only 0.001 Å RMSD for non-hydrogen atoms. Clearly, this is just the experimental monomer. The energy of this monomer decreased to 29 kJ/mol after the optimization of the X-H bond lengths alone. Given that the target XXXI form C polymorph was eventually excluded from the 7BT, we will omit it from further considerations.

Now let us examine the crystal structures of the representative polymorphs. Column (a) of Table 12 illustrates how closely these structures align with those of the associated experimental polymorphs. The $RMSD_{30}$ values (see captions of Table 12 for details) range from 0.18 to 0.53 Å (considering only the closest representation for target XXVII here and all subsequent discussions). Thus, the RMSDs for all the original experimental representative crystals were within the matching range of 7BT. This, in our view, should not be the case, as optimizations should have been necessary for the representative polymorphs to be included within the range.

We will now discuss what strategy could our team select had we known how the stage 2 pools were created. Table 12 shows our rankings based on the lattice energies with intermonomer contributions from aiFF and monomer-deformation penalties calculated *ab initio*. Target XXXI is an exception, as the penalties were calculated from the modified GAFF. Column (d) of Table 12 shows that our enormous effort invested in CSPs with flexible monomers was counterproductive, in particular for target XXXI, as generally the rankings in this column are the worst, despite the fact that this protocol required the largest effort. The remaining columns of Table 12 show rankings and RMSDs for (a) the original CCDC structures; (b) the same structures with monomer substituted by constrained-optimized ones; (c) the structures (b) after rigid-monomer optimizations. The best ranks are found in column (a) in 3 instances, column (b) in 6 instances, column (c) in 4 instances, and column (d) in 0 instances. The average rankings for these columns are 86, 73, 66, and 108, respectively. The $RMSD_{30}$ averages are 0.30, 0.25, 0.32, and 0.70 Å, respectively. Clearly, the outcome is not proportional to the computational effort, as this effort increases from (a) to (d). When constrained-optimized monomers were substituted for the original ones, this led to an overall improvement and increased stability in the rankings, with the exception of XXXIII #233. Thus, the constrained optimization of monomers, an inexpensive step, should be a recommended part of the ranking protocol. The UPACK rigid-monomer optimizations in column (c) led to a modest overall worsening for targets XXVII, XXVIII, and XXXI, but polymorphs of target XXXII experienced a very significant boost in rankings. The only negative change was for form B of target XXXIII where the ranking dropped from 19 to 283. Overall, such optimization proved to be worth doing for

the very specific ranking pool in 7BT. They are definitely worth doing for typical ranking pools. The reason for the underperformance for XXXIII is that many-body induction effects—which our aiFFs did not include—are crucial for this target (see the discussion in Sec. 4.7.2). Apparently, these effects lead to worsening of the geometry of form B: indeed, RMSD changed from 0.264 to 0.665 Å. The "best-of" rankings for the consecutive polymorphs in Table 12 are 1, 1, 1, 5, 3, 23, 30, 201, and 19, reasonably accurate predictions given the complexity of targets, Single-point pDFT+D calculations on these structures could still result in an improvement in these rankings, following the protocol of the aiFF@CSPs technique (Nikhar & Szalewicz, 2022). For systems where such calculations were performed, these improvements were indeed observed for systems XXXI, XXXII, and XXXII, while for system XXVIII the ranking

remained 1.

Table 12. *Comparisons of ranks and $RMSD_{30}$'s (in Å) of the original experimental-representative polymorphs (the structure numbers in parentheses) and of polymorphs obtained after various types of optimizations. RMSDs are relative to the experimental polymorphs (considering only non-hydrogen atoms, except for XXVII, where core-only non-hydrogen atoms are considered). At least 30 monomers matched the experimental structures, except for XXXI/B (indicated by parentheses), where only 14 monomers matched. The tolerances were 25%/25°. The column "original" shows rankings from single-point calculations (without any optimizations) on the exact experimental representatives. The "constr-opt" column shows ranks and RMSD after single-point calculations for the same crystals, but with the monomers replaced by the constrained-optimized ones. The next two columns present results from lattice energy minimizations with rigid constrained-optimized monomers ("rigid-mono opt") and (when performed) MD simulations with flexible monomers ("flex-mono opt"). Italics indicate the submitted findings, and boldface font highlights the best predictions.*

Target	Polymorph	(a) original rank/RMSD	(b) constr-opt rank/RMSD	(c) rigid-mono opt rank/RMSD	(d) flex-mono opt rank/RMSD
XXVII	A (#28)	**1**/0.530	**1**/0.531	*3*/***0.437***	
	A (#38)	87/0.804	85/0.804	*70/0.554*	
	A (#59)	5/0.828	5/0.828	*4/0.841*	
	A (#61)	3/0.569	2/0.569	*11/0.468*	
XXVIII	A (#144)	**1**/0.234	**1**/0.234	***1/0.200***	
XXXI	A_{maj} (#98)	2/0.180	**1**/0.164	**1/0.161**	*21/0.420*
	A_{min} (#01)	22/0.332	**5**/0.336	7/0.310	*35*/***0.274***
	B (#25)	14/0.247	**3**/0.252	5/**0.235**	*1/(1.953)*
XXXII	A (#317)	201/0.264	116/0.266	***23***/***0.261***	
	B (#232)	307/**0.298**	241/0.303	**30**/0.334*	
XXXIII	A (#233)	**201**/0.387	272/0.383	244/0.326	*349*/***0.322***
	B (#452)	24/**0.254**	**19**/0.264	283/0.665	*132/0.506*

*Due to a mistake in UPACK calculations, structure #232 was submitted at unranked position 487 (the geometry was the original CCDC geometry, but with monomers substituted by the constrained-optimized ones), see Sec. S2.4.5. The values in the table correspond to the correctly optimized structure in the post-submission.

5.5. Overall performance of our CSP protocols in stage 1 including post-submission investigations

The overall performance of our team in stage 1 of 7BT CSPs is shown in Table 13. We have successfully predicted only one of the polymorphs, A_{min} of target XXXI, and the stoichiometry of target XXX. While this is a low success rate, it places our team in the middle of the 21 teams participating in stage 1. There were 13 challenges in

this stage including the stoichiometry challenge. We count different polymorphs of a given target as separate challenges and in the case of target XXXI we also count A_{min} and A_{maj} as separate challenges, but not count form C as it was impossible to predict. Thus, our 2 correct predictions mean 15% success rate. There were 10 groups with correct predictions for more than 2 challenges and 9 groups with less than 2 correct predictions. For target XXIX, our prediction was close to the CCDC thresholds, but for the remaining targets our predictions were quite far from the thresholds, mostly due to the number of matching molecules below 30. On the other hand, the post-submission work shows that if we were able to include more conformers in our predictions, as initially planned, and if some mistakes were avoided, we would have obtained 7 successful predictions, 54% success rate. Only 2 groups achieved more than 7 successful predictions in 7BT. Of course, the success rates for other groups could have increased if there were some mistakes in their procedures and these mistakes were fixed.

The reasons for such underperformance in the stage 1 of 7BT were, depending on a system, a combination of the following three factors: (*a*) some rigid-monomer CSPs used monomer geometries far from the experimental one, (*b*) some aiFFs were insufficiently accurate, and (*c*) flexible-monomer CSPs used intramonomer FFs that were insufficiently accurate, i.e., the generic empirical intramonomer FFs in flexible CSPs failed to accurately represent the true monomer energy landscape, compromising the reliability of predictions. We will now discuss these issues target by target, explaining also how fixing of them in the post submission work led to significantly improved success rate.

In rigid-monomer CSPs performed for target XXVII, the failure in stage 1 was due to the geometry of the conformer used, both in the development of aiFF and in CSPs, being significantly different from experimental monomer geometry, see Table 7. Therefore, even with a perfect aiFF, we would not have met the CCDC match criteria. In the post submission work, we performed CSPs using conformer number 7, much closer to the experimental monomer, although the RMSD of 0.6 Å is still fairly large. These CSPs resulted in an $RMSD_{30} = 0.98$ Å for our best-match polymorph, where the majority of the discrepancy is at the monomer level. Still, this prediction, at rank 45 after merging conformers 1 and 7 predictions, is within CCDC criteria. The rank would certainly become worse if we have included other conformers, but is likely to stay within the top 1500. Note that for most targets the geometries of one of the local minima identified in our conformer searches match the experimental monomer geometries quite reasonably, except for targets XXVII and XXXII, see Table 7.

For target XXVIII, similarly as for target XXVII, the failure in stage 1 was due to the conformer used in the aiFF development and in CSPs being significantly different from experimental monomer, see Table 7. Since the closest conformer has an RMSD of 0.8 Å relative to the experimental monomer, using this monomer in rigid-monomer CSPs did not produce a polymorph within CCDC limit. The reason for no good conformer match is that the experimental monomer is planar with inversion symmetry, while none of our conformers has this symmetry. It actually could have been anticipated within the blind test that theoretically derived conformers will not be symmetric due to the Jahn-Teller effect. Since experimental monomers tend to have more planar than folded structures, one should add to the list of conformers a monomer obtained

by symmetrization of one of the conformers. We symmetrized conformer 2 and CSPs with this conformer led to the rank 1 polymorph with a very low $RMSD_{30} = 0.292$ Å.

For target XXIX, rigid-monomer CSPs should have have identified the experimental polymorphs since conformer A is nearly identical to the experimental monomer. Not knowing this, we performed CSPs with conformers A, B, and their mixtures. Our best match to the experimental polymorph was reasonable, but beyond CCDC criteria. In post-submission work, we performed rigid-monomer CSPs in $Z' = 1$, 2, and 3 using only conformer A and found a polymorph close to the experimental polymorph at rank 68 with $RMSD_{30} = 0.142$ Å. This would had been an excellent prediction in the test as only two groups found this polymorph and the RMSDs of their predictions were similar to our. Our post-submission result suggests that a mistake was made in the stage 1 predictions. Another possibility is that the failure was due to the inaccuracy of the heterogeneous dimer aiFF used in the CSPs of the mixed-conformer crystals, leading to these polymorphs being erroneously ranked high and pushing the polymorph now at rank 68 beyond the range of 1500 on the submission list.

For target XXX, we correctly predicted the stoichiometry of crystals, but our flexible-monomer CSPs gave within 1500 submitted structures only polymorphs with small number of molecules matching the experimental crystal. Our post-submission rigid-monomer CSPs with conformers used in the development of aiFFs (conformer 1 for T, conformer 2 for C) were not much more successful in terms of RMSDs (although we did find polymorphs with ranks 23 and 1 for forms A and B, respectively). This could be expected since while the conformer 1 of T is very close to the experimental monomer, conformer 2 of C has RMSDs of 1.219 Å relative to the monomer of form A and 0.421/0.960 Å relative to the two monomers in the asymmetric unit of form B (cf. Table 7). We also performed CSPs using the experimental monomer C geometry, with rationalized X-H bonds, taken from polymorph A. We found the experimental polymorph A at rank 17 with $RMSD_{30} = 0.231$ Å (such analysis was not performed for polymorph B). This indicates that the monomer geometry was the main problem of our stage 1 post-submission predictions and that the intermonomer aiFF is reasonably accurate despite some inadequacies of the CC aiFF discussed in Sec. 5.3 (this aiFF was not fully converged and had some holes remaining). The empirical intramonomer FF used in our flexible-monomer CSPs was apparently not able to optimize monomer C geometry to resemble the experimental monomer C. While CSPs with empirical intramonomer FFs would not succeed for target XXX, conformer-based rigid-monomer CSPs might. As seen in Table 7, there are conformers of C with RMSDs relative to experimental monomers below 1 Å. While aiFF-based CSPs with such conformers will most likely produce polymorphs with RMSDs above 1 Å, pDFT+D calculations, a part of aiFF@CSP protocol (Nikhar & Szalewicz, 2022), on the set of, say, 100 top-ranked polymorphs from such CSPs are likely to reduce these RMSDs to below 1 Å.

For target XXXI, we performed multiple rigid-monomer CSPs using all 8 low-energy conformers. For this target we successfully predicted the structure of form A_{min}, but not the other two forms. This is surprising since some of the conformers used in CSPs are very close to experimental monomers, cf. Table 7. Also the aiFF was fairly accurate aiFF, as seen in Table 8. In the post-submission work, we repeated these CSPs using at this time only conformers 2 and 4. These CSPs resulted in excellent predictions, with ranks 3, 5, and 24 for forms A_{maj}, A_{min}, and B, respectively, with $RMSD_{30}$

values within CCDC limits. With such good ranks, it is unlikely that inclusion of all 8 conformers would have resulted in these ranks becoming larger than 1500. The reasons for for the submitted polymorphs being so different are unclear.

Target XXXII was a difficult task for CSPs due to its size and significant flexibility. Indeed, it was predicted by only 3 groups. Our flexible-monomer CSPs found only polymorphs with small number of matching molecules. To identify reasons for such predictions, one should bear in mind that the intermonomer aiFF was quite successful in rigid-monomer rankings of stage 2 (taking into account good rankings of polymorph B after a human mistake was fixed). We also refitted the aiFFs in stage 2, but the improvements were not large, see Figs. S35a and S35b. Thus, we may conclude that the insufficient accuracy of the GAFF intramonomer contributions led to poor predictions in stage 1. This conclusion is consistent with our observations from other targets: generic empirical FFs are inadequate for CSPs of highly flexible molecules. We have not performed any extensive work on target XXXII in the post submission phase since it would either require developing an intramonomer FF fitted to *ab initio* data or performing rigid-monomer CSPs with a large number of conformers. However, we carried out a flexible-monomer pure GAFF optimization starting from the experimental polymorph B. The resulting polymorph has $\mathrm{RMSD}_{13} = 0.581$ Å and rank 163 among the pool from complete pure GAFF CSPs. A further optimization of this structure with intermonomer part of GAFF replaced by aiFF led to very small changes. This result, while markedly outside CCDC criteria, is significantly better than the best submitted result, indicating some issue in the structure generation procedures. One may add that while conformer-based CSPs would be hard for target XXXII, they are not impossible. Lowest-energy flat conformers are energetically above a set of a couple dozens of folded conformers. The latter can be eliminated by geometric criteria and then the pool of flat conformers may be small enough to manage.

For target XXXIII, we performed flexible-monomer CSPs with intermonomer aiFFs and intramonomer empirical FFs. As for some other targets, our best predictions have small numbers of matching monomers. The main reason for such predictions is again that empirical intramonomer FFs are not accurate enough for the monomers of target XXXIII, as demonstrated in Sec. 5.3. The issues with quality of the aiFF for dimer CC, related to this dimer containing two equally charged monomers, see Sec. 4.7.1, could lead to an additional worsening of predictions. To shed more light on the latter factor, we performed rigid-monomer CSPs for target XXXIII. As seen in Table 7, the monomer geometries used for M and S aiFF development had RMSDs of 0.016 and 0.092 Å, respectively, relative to the experimental monomers of polymorph A. For polymorph B, M matched the experimental monomer equally well, but RMSD for S was 1.026 Å. Thus, one would anticipate to get good ranking for polymorph A at least. Such CSPs with stage 1 aiFFs, however, gave results not much better than in submission. We could find polymorphs with better matches than in submitted results, $\mathrm{RMSD}_{11} = 1.575$ and $\mathrm{RMSD}_{19} = 0.845$ Å for forms A and B, respectively, but rankings were very poor: 2632 and 6561. This may indicate that the substandard quality of the SS aiFF contributed negatively to the stage 1 results. To further explore this issue, we also performed rigid-monomer CSPs with stage 2 aiFFs and using for monomer S the 4 lowest-energy conformers (conformers 1 and 2 are optical isomers, conformer 3 is the conformer at 7.00 kJ/mol that is closest to the experimental monomer in form

B, cf. Table 7). The results, listed in Table 13, did improve, but would not constitute successful predictions in 7BT. We interpret these results in the following way. Even the aiFFs of stage 2 are not accurate enough since they do not not include many-body polarization effects which are large for charged monomers. These effects apparently are not making huge differences in ranking of a given set of crystals, as shown by the results of stage 1 where rankings of form B were good and of form A reasonable, but significantly impact optimizations of crystal structures. In particular, in stage 2 the results were worse after optimizations, see Table 12. The autoPES program can develop polarizable FFs and such FFs can be used in future to perform CSPs for target XXXIII, hopefully leading to much improved results.

Table 13. *Performance of the CSP protocol used in stage 1, where N in $RMSD_N$ (in Å) signifies the number of monomers matched with the experimental crystal structure.* ***Bold*** *font represents the submitted result that was considered a successful prediction.*

Target	Polymorph	Stage 1 submission		Post submission		Post-submission comment
		Rank	$RMSD_N$	Rank	$RMSD_N$	
XXVII	A	749	3.700_6	45	0.98_{30}	Rigid-monomer CSPs using conformers 1 and 7
XXVIII	A	1482	2.100_4	1	0.292_{30}	Rigid-monomer CSPs using symmetrized conformer 2
XXIX	A	937	0.197_{27}	68	0.142_{30}	Rigid-monomer CSPs with equilibrium conformer in $Z' = 1$-3
XXX	A	1038	4.075_{11}	23	2.700_6[†]	Rigid-monomer CSPs using
	B	1359	0.577_{10}	1	3.600_8	conformer 2 of C
XXXI[‡]	A_{maj}	-	-	3	0.630_{30}	Rigid-monomer CSPs using
	A_{min}	**1192**	$\mathbf{0.857}_{30}$	5	0.963_{30}	conformers 2 and 4
	B	-	-	24	0.361_{30}	
XXXII	A	418	2.846_6			UPACK pure-GAFF optimization
	B	543	1.620_6	163	0.581_{13}	starting from experimental form B
XXXIII	A	1220	1.687_8	1543	2.181_{19}	Rigid-monomer CSP using confor-
	B	792	1.551_{12}	1699	0.821_{19}	mers M: 1, S: 1-4 and stage 2 aiFF

[†] Found at rank = 17 with $RMSD_{30}$ = 0.231 Å using experimental monomer C geometry with rationalized X-H bonds, see discussion.
[‡] We have not included a row for Form C since this form was essentially impossible to predict due to voids in crystal structures and indeed it was not predicted by any group (Hunnisett *et al.*, 2024*a*).

5.6. Overall performance of our CSP protocols in stage 2 including post-submission investigations

In stage 2, the participants were asked to rank sets of polymorphs supplied by the organizes. Thus, this stage did not involve any structure generation but only ranking or optimizations followed by rankings of the provided structures. Our protocols used in stage 2 were partly analogous to those of stage 1, but included several new elements: (*a*) aiFFs were refitted using dimers extracted from the CCDC-provided polymorphs and the CCfC method. Our flexibilized aiFFs demonstrated reasonable accuracy for deformed monomers, as detailed in Table 9 and Sec. S2.5.2, and a refit to a training set containing a more diverse pool of monomers improves their performance. (*b*) For

targets XXXI and XXXIII, empirical intramonomer FFs were refitted using *ab initio* data (only the original empirical FFs were used in stage 1). (*c*) MD simulations were performed for targets XXXI and XXXIII. For the remaining targets, XXVII, XXVII, and XXXII, we performed only rigid-monomer CSPs using UPACK. The monomer deformation contributions to lattice energies were computed using DFT+D except for target XXXI. In post-submission, we performed only rigid-monomer polymorph optimizations or single-point energy calculations using aiFFs or pDFT+D. For details, see Sec. 5.4 and Table 12.

Table 14 presents the result submitted and our best predictions obtained in post-submission work. These best post-submission results are those highlighted in bold in Table 12, except for target XXXIII where we list pDFT+D predictions. The data in Table 14 show that we achieved a success rate of 33% (3 out of 9) in stage 2 submission and 89% (8 out of 9) in the post-submission CSPs. Here 'success' is defined as identifying the experimental polymorph in the top 10% of the structures, i,e., within the top 10 for targets with 100 structures or the top 50 for targets with 500 structures. There were 21 groups participating in stage 2 and 3 (12) groups successfully predicted more than 8 (3) polymorphs, 3 groups predicted 3 polymorphs, and 9 groups predicted less than three polymorphs. Most groups with high success rate used pDFT+D, the method favored by the construction of the CCDC lists of polymorphs, as discussed in Sec. 5.4. Since experimental representative polymorphs on CCDC lists were very close to actual experimental polymorphs, see Sec. 5.4, our successful predictions for targets XXVII and XXVIII were expected. The same would have been true for target XXXII if not a human mistake made in the CSPs (however, for polymorph A our predictions was successful). For the remaining 2 targets, XXXI and XXXIII, our predictions were disadvantaged by the use of refined empirical intramonomer FFs. Below we discuss

the results presented in Table 14 target by target.

Table 14. *Performance of our CSP protocols used in stage 2. All $RMSD_{30}$ values (in Å) are calculated with respect to the experimental crystal structures. 'Stage 2 submissions' column refers to the stage 2 submissions. 'Post submissions' column lists the best aiFF-based predictions obtained in the post-submission phase except for target XXXIII where pDFT+D results are given (aiFF-based results are given in Table 12). 'Post submission details' column specifies the type of monomer used and the types of CSP performed. Submitted results (post-submission results) considered successful predictions are shown in boldface (italic boldface).*

Target	Poly	Stage 2 submission		Post submission		Post-submission details	
		Rank	$RMSD_{30}$	Rank	$RMSD_{30}$	Monomer	Method
XXVII	A	**3**	**0.437**	***1***	***0.530***	original	Single-point energy
XXVIII	A	**1**	**0.200**	***1***	***0.200***	constr-opt	Rigid-monomer optimization
XXXI	A_{maj}	21	0.420	***1***	***0.161***	original	Rigid-monomer optimization
	A_{min}	35	0.274	***5***	***0.336***	constr-opt	Single-point energy
	B	1[†]	1.953[‡]	***3***	***0.252***	constr-opt	Single-point energy
XXXII	A	**23**	**0.261**	***23***	***0.261***	constr-opt	Rigid-monomer optimization
	B	487	0.282	***30***	***0.334***	constr-opt	Rigid-monomer optimization
XXXIII	A	349	0.322	123	0.383	constr-opt	pDFT+D
	B	132	0.506	***1***	***0.264***	constr-opt	pDFT+D

[†] Cannot be counted as a successful match due to large RMSD.
[‡] RMSD is for overlap of 14 molecules.

For target XXVII, our submission based on rigid-monomer CSPs with the constrained-optimized monomer was successful with rank 3 and had the lowest $RMSD_{30}$ of all our predictions. However, the single-point lattice-energy calculation with the original monomer is at rank 1, so it is even better. In fact, all predictions listed in Table 12) are successful.

For target XXVIII, our submission based on rigid-monomer CSPs with the constrained-optimized monomer was successful with rank 1 and had the lowest $RMSD_{30}$ of all our predictions. The same result is used in the post-submission column. All our predictions for target XXXVII are at rank 1.

In the submission phase for target XXXI, we used a modified GAFF to describe intramonomer interactions. This modification consisted in replacing the original equilibrium bond lengths and bond angles with those from the conformer used in the aiFF fitting. The partial charges were taken from aiFF and all the remaining parameters were from the original GAFF. As discussed in the context of target XXXIII in Sec. S2.5.3, such simple modification are unlikely to result in the modified intramolecular FF giving energies reasonably similar to those from *ab initio* calculations, in particular in the low-energy region most relevant for CSPs. As a consequence, our

submission in stage 2 was not successful for any of the polymorphs of target XXXI. As we now know, due to the construction of CCDC lists flexible-monomer predictions were not needed in stage 2, although constrained optimization of monomers improved the predictions significantly. The single-point calculations and rigid-monomer optimizations with such monomers both gave excellent predictions, with best ranks 1, 3, 5 for forms A_{maj}, B, and A_{min}, respectively.

For target XXXII, the experimental polymorph A was ranked reasonably well at 23, in the submission. Polymorph B was placed by us at the end of the submitted list (unranked position 487) due to an error in lattice energy calculations for crystals with $Z' = 2$. When these calculations were repeated in the post-submission phase, the rank became 30.

For target XXXIII, we developed an improved intramonomer FF. The improvements were more substantial than for target XXXI: all GAFF parameters except for LJ ones were refitted to a set of *ab initio* monomer energies (charges were taken from aiFF), see Sec. S2.5.3. However, the low-energy regions of monomer S still had fairly large errors as seen in Fig. 6. The submitted results are reasonable in terms of $RMSD_{30}$, but rankings are poor. These results do not allow use to evaluate the quality of the refitted intramonomer FF since another effect that impacted rankings negatively was the lack of many-body effects in our aiFFs. In post-submission investigations, single-point calculations on the set 500 polymorphs provided by CCDC gave good ranking for polymorph B, 19 and 24 for the constrained-optimized and the original monomers, respectively, but rigid-monomer optimization made these rankings much worse. For polymorph A, the best ranking was 201. Therefore, we list in Table 14 the results from pDFT+D single-point calculations. We believe the latter predictions are better since many-body polarization effects are included in pDFT+D.

6. Final conclusions

This paper describes the CSP protocols used by our team in 7BT of CSPs and the performance of these protocols in this test. There are two main version of our protocols: rigid-monomer CSPs involving several low-energy conformers and flexible-monomer CSPs. Monomer deformation penalties are computed using DFT+D methods. All CSPs are performed using FFs which are simple analytic potential functions, therefore our approach is inexpensive, with main costs due to the developments, fully automatic, of tailor-made FFs fitted to results of *ab initio* calculations for monomers and dimers. This is a significant advantage of our approach since some other protocols used in 7BT are extremely expensive in terms of computational resources.

In stage 1, the structure generation stage of 7BT, our success rate in submission was 15%, while after fixing some mistakes, we reached success rate of 54% in postsubmission work. The success rate was generally low in stage 1, with only 10 (2) groups out of 21 having success rate above 15% (54%). In stage 2, the structure ranking stage, our success rate in stage 2 was 33% (89%) in submission (post-submission). There were 12 (3) groups out of 21 that achieved success rates above 33% (89%). We were one of only two academic group that answered all challenges. The main reasons that we did not perform better in the submission phases were the use of inaccurate intramonomer FFs, the use of only a single conformer in rigid-monomer CSPs, and human errors.

We believe that our protocols, if properly executed, offer one of the most reliable and inexpensive methods of predicting crystal structures. 7BT provided both the guidance how our protocols should be used and a general information about crystal structure predictions. Below we discuss our findings.

Our approach uses intermonomer aiFFs fitted using autoPES codes to *ab initio* calculations for dimers using either SAPT(DFT) or DFT+D methods. In the standard approach, only a single monomer configuration is used in *ab initio* calculations, but the form of aiFFs allows one to use them for arbitrary geometries of monomers, which is called flexibilization of aiFFs. The fits can also be performed to training data containing monomers with varying geometries and aiFFs fitted in this way are more robust. aiFFs used in CSPs should be refined by the CCfC procedure which uses a given aiFF in CSPs, cuts a number of dimers from best-ranked polymorphs, and uses these dimers to refit aiFF. This process is iterated until the fit achieves desired accuracy on dimers cut from last iteration crystals. 7BT was a critical test for aiFFs developed in this way and our good results in post-submission phase show that this approach performs very well. CCfC step is critical and some aiFFs used in stage 1 without this step performed unsatisfactorily. There was one exception to good performance of aiFFs: target XXXIII. This target contains two charged monomers which leads to large three-body polarization effects, not included in our current protocol. Since both autoPES (UPACK) have option to generate (use) polarizable FFs, our protocol can applied to ionic monomers in future CSPs. A part of the aiFF@CSPs protocol (Nikhar & Szalewicz, 2022) are pDFT+D calculations on a set of best ranked polymorphs from aiFF-base calculations. We have not utilized this option in submitted results.

In contrast to intermonomer aiFFs, there is no clear picture concerning intramonomer FFs. Our work in stage 1 of 7BT has shown that standard empirical FFs are completely unreliable in CSPs involving monomers with soft degrees of freedom. In stage 2, we improved such FFs by reparametrizing them to results of *ab initio* calculations on monomers. The simplest form of such reparametrizations, replacement of equilibrium constants by *ab initio* values, was shown to be insufficient (cf. CSPs for target XXXI). More extensive reparametrization for target XXXIII, involving fits to a set of training monomer energies computed *ab initio*, produced an FF significantly better than the empirical one. Still, in low-energy region critical for CSPs, the deviations from *ab initio* values were still above 1 kJ/mol, possibly too large for accurate predictions. Unfortunately, this applications was not conclusive due to the fact that target XXXIII has significant three-body interactions neglected by us.

Despite the issues with intramonomer FFs, we were able to perform successful predictions for monomers with soft degrees of freedom using rigid-monomer CSPs with several low-energy conformers. We found that for most targets of 7BT the monomers from experimental polymorphs have geometries close to the geometry of one of the conformers, which is the reasons that this method works. Our findings are similar to those of (Thompson & Day, 2014) who investigated crystals of 15 molecules with soft degrees of freedom and found that in 75% of cases studied the monomers were deformed by the crystal field relative to their gas-phase equilibrium geometry in such a way that their energies were less than 10 kJ/mol above their gas-phase minimum energy value. There were rare instances where this deformation penalty reached up to 20 kJ/mol. Since we include all conformers within the 20 kJ/mol range, the crystal-

monomer geometry is more likely close to the the geometry of one of the conformers in this energy range rather than to be at a point far from any conformer since the former geometry has lower deformation penalty. If this is the case, pDFT+D crystal geometry optimization which is a part of our CSP conformer-based protocol (although we have not implemented it in 7BT) should converge to experimental crystal geometry. Clearly, there will be cases where such protocol will fail, but the work of Thompson and Day indicates that it should work in majority of cases. In 7BT, there were two cases where the rigid-monomer CSP did not perform as well. The conformers of the monomer of target XXVIII are nonsymmetric due to Jahn-Teller's effect, while the experimental monomer is centrosymmetric. Since the presence of Jahn-Teller's effect is known a priori, one can symmetrize low-energy conformers and use them in CSPs. After we did this, our predictions for target XXVIII were excellent. The other exception is target XXXII, where the experimental monomers are planar, while couple dozens lowest-energy conformers have folded structures. It might be possible to use conformer-based CSPs for target XXXII after eliminating folded conformers based on their geometry, but we have not done this.

In stage 2, we performed MD simulations for targets XXXI and XXXIII, which might allow us to include temperature effects in CSPs. However, due to poor quality of intramonomer FFs, not conclusions could be made on how well MD simulations perform relative to UPACK CSPs.

Based on the performance of our protocols in 7BT, we believe we can recommend our conformer-based aiFF@CSPs to be used in predictions for crystals built of monomers with soft degrees of freedom. The important aspect of this approach is that it is based on first principles. The 7BT results indicate that it should be sufficient to include conformers with energies up to 20 kJ/mol above the global minimum (very similar to 25 kJ/mol suggested in (Thompson & Day, 2014)). The aiFFs should be refined using the CCfC procedure. The programs performing all such calculations are available on web pages of UD group. After the proposed protocol is completed, the results can be fine-tuned by computing monomer-deformation penalties on a higher level of theory than DFT+D, if feasible. Such an approach was shown by (Greenwell & Beran, 2020) to improve ranking. For ionic monomers, the FFs should contain a polarizations term, but this option has not been tested yet. The final step should be pDFT+D calculations on about 50 top-ranked polymorphs, possibly with lattice energy minimizations. The alternative approach, with intramonomer FFs fitted to monomer *ab initio* energies, cannot be yet recommended. However, the refinement procedures used by us in 7BT can certainly be improved, hopefully resulting in first-principles predictions for all types of molecular crystals.

Acknowledgements: The work at the University of Delaware was supported by the U.S. Army Research Laboratory and Army Research Office under grant W911NF-19-0117 and by National Science Foundation under grants CHE-1900551, CHE-2154908, and CHE-2313826. The use of the DARWIN computing system funded by NSF grant 1919839 is also acknowledged. The authors thank Rajni M. Bhardwaj, Eric J. Chan, Jutta Rogal, Hongxing Song, Leslie Vogt-Maranto, and Mark E. Tuckerman who participated in the 7BT work that formed the onset of the present paper.

References

Abrams, J. B. & Tuckerman, M. E. (2008). *J. Phys. Chem. B*, **112**(49), 15742–15757.

Adamo, C. & Barone, V. (1999). *J. Chem. Phys.* **110**(13), 6158–6170. https://doi.org/10.1063/1.478522

Barducci, A., Bussi, G. & Parrinello, M. (2008). *Phys. Rev. Lett.* **100**(2), 020603. https://link.aps.org/doi/10.1103/PhysRevLett.100.020603

Barone, V., Cacelli, I., De Mitri, N., Licari, D., Monti, S. & Prampolini, G. (2013). *Phys. Chem. Chem. Phys.* **15**, 3736–3751. http://dx.doi.org/10.1039/C3CP44179B

Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. (1993). *J. Chem. Phys.* **97**(40), 10269–10280. https://doi.org/10.1021/j100142a004

Becke, A. D. (1988). *Phys. Rev. A*, **38**, 3098–3100.

Bernholdt, D. E. & Harrison, R. J. (1996). *Chem. Phys. Lett.* **250**, 477–484.

Betz, R. M. & Walker, R. C. (2015). *JCC*, **36**(2), 79–87. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23775

Binkley, J. S., Pople, J. A. & Hehre, W. J. (1980). *J. Am. Chem. Soc.* **102**(3), 939–947. https://doi.org/10.1021/ja00523a008

Blöchl, P. E. (1994). *Phys. Rev. B*, **50**, 17953–17979. https://link.aps.org/doi/10.1103/PhysRevB.50.17953

Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., Broglia, R. A. & Parrinello, M. (2009). *Comput. Phys. Commun.* **180**(10), 1961–1972. https://www.sciencedirect.com/science/article/pii/S001046550900157X

Bukowski, R., Cencek, W., Jankowski, P., Jeziorska, M., Jeziorski, B., Garcia, J., Kucharski, S. A., Lotrich, V. F., Metz, M. P., Misquitta, A. J., Moszyński, R., Patkowski, K., Podeszwa, R., Rob, F., Rybak, S., Szalewicz, K., Williams, H. L., Wheatley, R. J., Wormer, P. E. S. & Żuchowski, P. S., (2020). SAPT2020: An *ab initio* program for many-body symmetry-adapted perturbation theory calculations of intermolecular interaction energies. University of Delaware and University of Warsaw. http://www.physics.udel.edu/~szalewic/SAPT.

Case, D. A., Belfon, K., Ben-Shalom, I. Y., Brozell, S. R., Cerutti, D. S., Cheatham, III, T. E., Cruzeiro, V. W. D., Darden, T. A., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Goetz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C. L., Skrynnikov, N. R., Smith, J., Swails, J., Walker, R. C., Wang, J., Wilson, L., Wolf, R. M., Wu, X., Xiong, Y., Xue, Y., York, D. M. & Kollman, P. A., (2020). AMBER 2020. University of California, San Francisco.

Chan, E. J., Shtukenberg, A. G., Tuckerman, M. E. & Kahr, B. (2021). *Crystal Growth & Design*, **21**(10), 5544–5557. https://pubs.acs.org/doi/10.1021/acs.cgd.1c00261

Chan, E. J. & Tuckerman, M. E. (2024). *Acta Cryst. B*, **80**. https://doi.org/10.1107/S205252062400132X

Chen, M., Cuendet, M. A. & Tuckerman, M. E. (2012). *J. Chem. Phys.* **137**(2), 024102. https://doi.org/10.1063/1.4733389

Chisholm, J. A. & Motherwell, S. (2005). *J. Appl. Crystallogr.* **38**(1), 228–231. https://doi.org/10.1107/S0021889804027074

Clark, M., Cramer III, R. D. & Van Opdenbosch, N. (1989). *J. Comp. Chem.* **10**(8), 982–1012. https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540100804

Cole, J. C., Groom, C. R., Read, M. G., Giangreco, I., McCabe, P., Reilly, A. M. & Shields, G. P. (2016). *Acta Cryst. B*, **72**(4), 530–541.

Cole, J. C., Korb, O., McCabe, P., Read, M. G. & Taylor, R. (2018). *J. Chem. Inf. Model.* **58**(3), 615–629. https://doi.org/10.1021/acs.jcim.7b00697

de Oliveira, P. J. P., Barros, C. L., Jorge, F. E., Canal Neto, A. & Campos, M. (2010). *J. Mol. Struct. (Theochem)*, **948**(1), 43–46. https://www.sciencedirect.com/science/article/pii/S0166128010001302

Ditchfield, R., Hehre, W. J. & Pople, J. A. (1971). *J. Chem. Phys.* **54**(2), 724–728. https://doi.org/10.1063/1.1674902

Dodda, L. S., Cabeza de Vaca, I., Tirado-Rives, J. & Jorgensen, W. L. (2017). *Nucleic Acids Res.* **45**(W1), W331–W336. https://doi.org/10.1093/nar/gkx312

Dunning, Jr., T. H. (1989). *J. Chem. Phys.* **90**, 1007–1023.

van Eijck, B. P. & Kroon, J. (1997). *J. Comp. Chem.* **18**(8), 1036–1042. https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199706%2918%3A8%3C1036%3A%3AAID-JCC7%3E3.0.CO%3B2-U

van Eijck, B. P. & Kroon, J. (1999). *J. Comp. Chem.* **20**(8), 799–812. https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199906%2920%3A8%3C799%3A%3AAID-JCC6%3E3.0.CO%3B2-Z

Feyereisen, M., Fitzgerald, G. & Komornicki, A. (1993). *Chem. Phys. Lett.* **208**, 359–363.

Francia, N. F., Price, L. S., Nyman, J., Price, S. L. & Salvalaglio, M. (2020). *Cryst. Growth Des.* **20**(10), 6847–6862.

Friedrich, N.-O., Flachsenberg, F., Meyder, A., Sommer, K., Kirchmair, J. & Rarey, M. (2019). *J. Chem. Inf. Model.* **59**(2), 731–742. https://doi.org/10.1021/acs.jcim.8b00704

Frisch, M. J., Head-Gordon, M. & Pople, J. A. (1990). *Chem. Phys. Lett.* **166**(3), 275–280. https://www.sciencedirect.com/science/article/pii/000926149080029D

Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Scalmani, G., Barone, V., Petersson, G. A., Nakatsuji, H., Li, X., Caricato, M., Marenich, A. V., Bloino, J., Janesko, B. G., Gomperts, R., Mennucci, B., Hratchian, H. P., Ortiz, J. V., Izmaylov, A. F., Sonnenberg, J. L., Williams-Young, D., Ding, F., Lipparini, F., Egidi, F., Goings, J., Peng, B., Petrone, A., Henderson, T., Ranasinghe, D., Zakrzewski, V. G., Gao, J., Rega, N., Zheng, G., Liang, W., Hada, M., Ehara, M., Toyota, K., Fukuda, R., Hasegawa, J., Ishida, M., Nakajima, T., Honda, Y., Kitao, O., Nakai, H., Vreven, T., Throssell, K., Montgomery, Jr., J. A., Peralta, J. E., Ogliaro, F., Bearpark, M. J., Heyd, J. J., Brothers, E. N., Kudin, K. N., Staroverov, V. N., Keith, T. A., Kobayashi, R., Normand, J., Raghavachari, K., Rendell, A. P., Burant, J. C., Iyengar, S. S., Tomasi, J., Cossi, M., Millam, J. M., Klene, M., Adamo, C., Cammi, R., Ochterski, J. W., Martin, R. L., Morokuma, K., Farkas, O., Foresman, J. B. & Fox, D. J., (2016). Gaussian 16 Revision C.01. Gaussian Inc. Wallingford CT.

Garcia, J., Podeszwa, R. & Szalewicz, K. (2020). *J. Chem. Phys.* **152**, 184109–(1:23). https://doi.org/10.1063/5.0005093

de Gelder, R., Wehrens, R. & Hageman, J. A. (2001). *J. Comput. Chem.* **22**(3), 273–289. https://onlinelibrary.wiley.com/doi/10.1002/1096-987X(200102)22:3¡273::AID-JCC1001¿3.0.CO;2-0

Giannozzi, P., Andreussi, O., Brumme, T., Bunau, O., Nardelli, M. B., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Cococcioni, M., Colonna, N., Carnimeo, I., Corso, A. D., 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., de-la Roza, A. O., 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. (2017). *J. Phys.: Condens. Matter*, **29**(46), 465901.
https://dx.doi.org/10.1088/1361-648X/aa8f79

Giannozzi, P., Baroni, S., Bonini, N., Calandra, M., Car, R., Cavazzoni, C., Ceresoli, D., Chiarotti, G. L., Cococcioni, M., Dabo, I., Corso, A. D., de Gironcoli, S., Fabris, S., Fratesi, G., Gebauer, R., Gerstmann, U., Gougoussis, C., Kokalj, A., Lazzeri, M., Martin-Samos, L., Marzari, N., Mauri, F., Mazzarello, R., Paolini, S., Pasquarello, A., Paulatto, L., Sbraccia, C., Scandolo, S., Sclauzero, G., Seitsonen, A. P., Smogunov, A., Umari, P. & Wentzcovitch, R. M. (2009). *J. Phys.: Condens. Matter*, **21**(39), 395502.
https://dx.doi.org/10.1088/0953-8984/21/39/395502

Greenwell, C. & Beran, G. J. O. (2020). *Cryst. Growth Des.* **20**(8), 4875–4881.

Grimme, S. (2011). *WIREs Comput. Mol. Sci.* **1**(2), 211–228.
https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.30

Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. (2010). *J. Chem. Phys.* **132**(15), 154104.
https://doi.org/10.1063/1.3382344

Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst. B*, **72**(2), 171–179.
https://doi.org/10.1107/S2052520616003954

Halgren, T. A. (1996). *J. Comput. Chem.* **17**(5-6), 490–519.
https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199604%2917%3A5/6%3C490%3A%3AAID-JCC1%3E3.0.CO%3B2-P

Halgren, T. A. (1999). *J. Comp. Chem.* **20**(7), 730–748.
https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%28199905%2920%3A7%3C730%3A%3AAID-JCC8%3E3.0.CO%3B2-T

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E. & Hutchison, G. R. (2012). *J. Cheminf.* **4**, 1–17.
https://doi.org/10.1186/1758-2946-4-17

Heinzerling, L., Klein, R. & Rarey, M. (2012). *J. Comp. Chem.* **33**(32), 2554–2565.
https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.23094

Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI, USA: University of Michigan Press.

Hunnisett, L. M., Abraham, N. S., Aitipamula, S. & *et al.* (2024*a*). *Acta Crystallogr., Sect. B: Struct. Sci*, **80**(6).

Hunnisett, L. M., Abraham, N. S., Aitipamula, S. & *et al.* (2024*b*). *Acta Crystallogr., Sect. B: Struct. Sci*, **80**(6).

Jakalian, A., Jack, D. B. & Bayly, C. I. (2002). *J. Comput. Chem.* **23**(16), 1623–1641.
https://onlinelibrary.wiley.com/doi/10.1002/jcc.10128

Jeziorski, B., Moszyński, R. & Szalewicz, K. (1994). *Chem. Rev.* **94**, 1887–1930.
https://doi.org/10.1021/cr00031a008

Jorgensen, W. L. & Tirado-Rives, J. (1988). *J. Am. Chem. Soc.* **110**(6), 1657–1666.
https://doi.org/10.1021/ja00214a001

Kendall, R. A., Dunning, Jr., T. H. & Harrison, R. J. (1992). *J. Chem. Phys.* **96**(9), 6796–6806.
https://doi.org/10.1063/1.462569

Kresse, G. & *et al.*, (2021). VASP: Vienna ab initio simulation package.
http://www.vasp.at

Kresse, G. & Furthmüller, J. (1996*a*). *Comput. Mater. Sci.* **6**(1), 15–50.
https://www.sciencedirect.com/science/article/pii/0927025696000080

Kresse, G. & Furthmüller, J. (1996*b*). *Phys. Rev. B*, **54**, 11169–11186.
https://link.aps.org/doi/10.1103/PhysRevB.54.11169

Kresse, G. & Hafner, J. (1993). *Phys. Rev. B*, **47**, 558–561.

Kresse, G. & Hafner, J. (1994). *Phys. Rev. B*, **49**, 14251–14269.
https://link.aps.org/doi/10.1103/PhysRevB.49.14251

Kresse, G. & Joubert, D. (1999). *Phys. Rev. B*, **59**, 1758–1775.
https://link.aps.org/doi/10.1103/PhysRevB.59.1758

Kullback, S. & Leibler, R. A. (1951). *Ann. Math. Stat.* **22**, 79–86.

Lee, C., Yang, W. & Parr, R. G. (1988). *Phys. Rev. B*, **37**, 785.

Macrae, C. F., Edgington, P. R., McCabe, P., Pidcock, E., Shields, G. P., Taylor, R., Towler, M. & van de Streek, J. (2006). *J. Appl. Crystallogr.* **39**(3), 453–457.
https://doi.org/10.1107/S002188980600731X

Mardirossian, N. & Head-Gordon, M. (2016). *J. Chem. Phys.* **144**, 214110–(1:23).

Metz, M. P., Piszczatowski, K. & Szalewicz, K. (2016). *J. Chem. Theory Comput.* **12**, 5895–5919.
https://doi.org/10.1021/acs.jctc.6b00913

Metz, M. P., Shahbaz, M., Song, H., Vogt-Maranto, L., Tuckerman, M. E. & Szalewicz, K. (2022). *Cryst. Growth & Des.* **22**(2), 1182–1195.
https://doi.org/10.1021/acs.cgd.1c01117

Metz, M. P. & Szalewicz, K. (2020). *J. Chem. Theory Comput.* **16**, 2317–2339.

Misquitta, A. J., Podeszwa, R., Jeziorski, B. & Szalewicz, K. (2005). *J. Chem. Phys.* **123**, 214103–(1:14).
https://doi.org/10.1063/1.2135288

Misquitta, A. J. & Szalewicz, K. (2005). *J. Chem. Phys.* **122**, 214109–(1:19).
https://doi.org/10.1063/1.1924593

Monkhorst, H. J. & Pack, J. D. (1976). *Phys. Rev. B: Condens. Matter*, **13**(12), 5188.

Neese, F. (2012). *WIREs Comp. Mol. Sci.* **2**(1), 73–78.
https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.81

Neese, F. (2018). *WIREs Comput. Mol. Sci.* **8**(1), e1327.
https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/wcms.1327

Nikhar, R. & Szalewicz, K. (2022). *Nature Comm.* **13**, 3095.
https://doi.org/10.1038/s41467-022-30692-y

Nikhar, R. & Szalewicz, K. (2024). Manuscript in preparation.

Pack, J. D. & Monkhorst, H. J. (1977). *Phys. Rev. B: Condens. Matter*, **16**(4), 1748.

Perdew, J. P., Burke, K. & Ernzerhof, M. (1996). *Phys. Rev. Lett.* **77**, 3865–3868.
https://link.aps.org/doi/10.1103/PhysRevLett.77.3865

Peterson, K. A., Figgen, D., Goll, E., Stoll, H. & Dolg, M. (2003). *J. Chem. Phys.* **119**(21), 11113–11123.
https://doi.org/10.1063/1.1622924

Petersson, G. A. & Al-Laham, M. A. (1991). *J. Chem. Phys.* **94**(9), 6081–6090.
https://doi.org/10.1063/1.460447

Plimpton, S. (1995). *J. Comp. Phys.* **117**(1), 1–19.
https://www.sciencedirect.com/science/article/pii/S002199918571039X

Podeszwa, R., Bukowski, R. & Szalewicz, K. (2006). *J. Phys. Chem. A*, **110**, 10345–10354.

Podeszwa, R., Rice, B. M. & Szalewicz, K. (2008). *Phys. Rev. Lett.* **101**, 115503–(1:4).

Podeszwa, R., Rice, B. M. & Szalewicz, K. (2009). *Phys. Chem. Chem. Phys.* **11**, 5512–5518.

Rappé, A. K., Casewit, C. J., Colwell, K. S., Goddard III, W. A. & Skiff, W. M. (1992). *J. Am. Chem. Soc.* **114**(25), 10024–10035.
https://doi.org/10.1021/ja00051a040

Rappe, A. M., Rabe, K. M., Kaxiras, E. & Joannopoulos, J. D. (1990). *Phys. Rev. B*, **41**, 1227–1230.
https://link.aps.org/doi/10.1103/PhysRevB.41.1227

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. (2016). *Acta Cryst. B*, **72**, 439–459.

Riplinger, C. & Neese, F. (2013). *J. Chem. Phys.* **138**, 034106.

Riplinger, C., Pinski, P., Becker, U., Valeev, E. F. & Neese, F. (2016). *J. Chem. Phys.* **144**(2), 024109.
https://doi.org/10.1063/1.4939030

Rob, F. & Szalewicz, K. (2013*a*). *Chem. Phys. Lett.* **572**, 146–149.

Rob, F. & Szalewicz, K. (2013*b*). *Mol. Phys.* **111**(9-11), 1430–1455.
https://doi.org/10.1080/00268976.2013.808770

Spek, A. L. (2009). *Acta Crystallogr. D*, **65**(2), 148–155.
https://doi.org/10.1107/S090744490804362X

Spellmeyer, D. C., Wong, A. K., Bower, M. J. & Blaney, J. M. (1997). *J. Mol. Graph. Model.* **15**(1), 18–36.
https://www.sciencedirect.com/science/article/pii/S1093326397000144

Sun, J., Ruzsinszky, A. & Perdew, J. P. (2015). *Phys. Rev. Lett.* **115**, 036402.
https://link.aps.org/doi/10.1103/PhysRevLett.115.036402

Tang, K. T. & Toennies, J. P. (1984). *J. Chem. Phys.* **80**(8), 3726–3741.
https://doi.org/10.1063/1.447150

The Cambridge Crystallographic Data Centre, (2024). The Cambridge Crystallographic Data Centre (CCDC). https://www.ccdc.cam.ac.uk/. Accessed: March 21, 2024.

Thompson, H. P. G. & Day, G. M. (2014). *Chem. Sci.* **5**(8), 3173–3182.

Tuckerman, M. E., Yarne, D., Samuelson, S. O., Hughes, A. L. & Martyna, G. J. (2000). *Comput. Phys. Commun.* **128**(1), 333–376.
https://www.sciencedirect.com/science/article/pii/S0010465500000771

Vainio, M. J. & Johnson, M. S. (2007). *J. Chem. Inf. Model.* **47**(6), 2462–2474.
https://doi.org/10.1021/ci6005646

Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. (2004). *J. Comp. Chem.* **25**(9), 1157–1174.
https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20035

Weigend, F. & Ahlrichs, R. (2005). *Phys. Chem. Chem. Phys.* **7**, 3297–3305.
http://dx.doi.org/10.1039/B508541A

Weigend, F. & Häser, M. (1997). *Theor. Chem. Acc.* **97**, 331–340.

Weigend, F., Häser, M., Patzelt, H. & Ahlrichs, R. (1998). *Chem. Phys. Lett.* **294**, 143–152.

Williams, H. L., Mas, E. M., Szalewicz, K. & Jeziorski, B. (1995). *J. Chem. Phys.* **103**, 7374–7391.
https://doi.org/10.1063/1.470309

Woon, D. E. & Dunning, Jr., T. H. (1993). *J. Chem. Phys.* **98**(2), 1358–1371.
https://doi.org/10.1063/1.464303

Yu, T.-Q. & Tuckerman, M. E. (2011). *Phys. Rev. Lett.* **107**, 015701. https://link.aps.org/doi/10.1103/PhysRevLett.107.015701

Żuchowski, P. S., Podeszwa, R., Moszyński, R., Jeziorski, B. & Szalewicz, K. (2008). *J. Chem. Phys.* **129**, 084101–(1:17). https://doi.org/10.1063/1.296855

iucr

Synopsis

Supply a synopsis of the paper for inclusion in the Table of Contents.