



Cite this: *CrystEngComm*, 2026, 28, 697
Accepted 22nd November 2025

Prediction of the crystal structure of avadomide using machine learning potentials and crystal engineering

Faezeh Bahrami, ^a Alireza Salimi, ^{*a}
Zahrasadat Momenzadeh Abardeh ^b and Artem R Oganov ^{*b}

Today, machine learning (ML) and crystal structure prediction (CSP) are principal tools in computational materials discovery. In this study, CSP of the pharmaceutical compound avadomide is presented, utilizing a workflow that combines the evolutionary algorithm USPEX, machine learning potentials in moment tensor potential (MTP) formulation, and crystal engineering concepts. Our study begins with the optimization of stable molecular conformations of avadomide using DFT, followed by the generation of likely crystal structures with the evolutionary algorithm USPEX. The optimization and ranking of structures based on ML potentials—trained on the subset of DFT data—were further refined through iterative improvements via active learning. The ML potentials employed in this study are currently constrained to local atomic environments and primarily captured short-range interatomic interactions. To address this limitation, the top-ranked structures were analyzed using crystal engineering concepts and found to contain synthons similar to those in experimental structures of related compounds in the CSD database. Using the synthon approach, two potential crystal structures were proposed for avadomide.

Received 1st September 2025,
Accepted 22nd November 2025

DOI: 10.1039/d5ce00849b

rsc.li/crystengcomm

Introduction

Machine learning (ML), a branch of artificial intelligence, is a powerful approach for creating various types of models through intelligent data analysis.¹ In recent years, ML has rapidly evolved into an effective tool across various scientific fields,² including medical diagnostics,³ particle physics,⁴ bioinformatics,⁵ nano science,⁶ and materials discovery.⁷ Today, some computational approaches in materials science require enormous resources and time.⁸ Unlike traditional approaches such as density functional theory (DFT), which rely on fundamental quantum-mechanical principles to model the behavior of electrons within a system, ML works on a different idea.⁹ ML learns from previous data to improve algorithms, manage unknown situations, and extract meaningful insights from vast datasets.¹⁰

As computational methods in materials science have developed, the integration of ML with crystal structure prediction (CSP) has significantly improved the efficiency of CSP.^{11,12} CSP aims to predict stable crystal structures of a substance based solely on its composition or chemical structure diagram prior to any experimental synthesis.^{13,14}

CSP can be divided into two main steps: first, a comprehensive search to find all promising crystal structures; second, selecting and ranking likely structures based on thermodynamic stability and probability of observation.^{15,16} In the early stages of CSP development, it was assumed that the most stable structure corresponded to the experimental one.^{17,18} While a predicted structure may be thermodynamically favorable, its synthesis in the laboratory may be challenging or even impossible.^{19,20} During crystallization, various kinetic factors intervene, leading to diverse often metastable molecular packing arrangements.^{21,22} Accessing direct information about kinetic factors poses a significant challenge. However, crystal engineering offers techniques to indirectly access and utilize this crucial information.^{23–26} Crystal engineering involves designing synthesizable polymorphs of molecular materials with desired properties by using frequently observed and energetically favorable intermolecular interactions known as synthons.^{27–29} Synthons are structural units in a crystal that reflect both thermodynamic and kinetic properties. Synthons can be extracted from the experimental structures in the Cambridge Structural Database (CSD).^{30,31}

Integration of crystal engineering and CSP provides valuable insights into crystal structures, but predicting and identifying the structures of pharmaceutical compounds remains a significant challenge.³² As drug synthesis and design techniques have advanced in the pharmaceutical

^a Department of Chemistry, Faculty of Science, Ferdowsi University of Mashhad, Mashhad, Iran. E-mail: salimi-a@um.ac.ir

^b Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Moscow 143026, Russian Federation. E-mail: A.Oganov@skoltech.ru

industry, the risks associated with designing and producing new solid forms persist. A notable example is ritonavir, for which the unexpected appearance of a new polymorph during manufacturing and distribution compromised its safety and efficacy.³³ To deal with these challenges, a combination of experimental techniques and CSP-based screening for polymorphs is crucial for determining stable forms under standard conditions, preventing the late appearance of polymorphs, and distinguishing stable and metastable pharmaceutical compounds at an early stage.^{34,35}

In addition, pharmaceutical compounds encounter two challenges in CSP due to their large molecular size and structural flexibility.^{36,37} First, due to the large number of degrees of freedom for flexible molecules, there is a risk that the actual crystal structures may be overlooked or mistakenly excluded. Second, a given compound may have many crystalline modifications (polymorphs).^{29,38,39} The primary difficulty is a large number of polymorphs and their usually small energy differences, often below 10 kJ mol⁻¹.⁴⁰ To overcome this, computational methods must be able to efficiently screen a wide range of structures while accurately accounting for intermolecular interactions at different energy levels.^{41,42} There is a pressing need for high-level methods capable of precisely ranking polymorphs.^{43,44} Due to the high costs associated with traditional methods like DFT, machine learning potentials were employed to efficiently evaluate a broad range of candidates and accurately rank them.⁴⁵

For materials modeling, moment tensor potentials (MTPs) in machine learning interatomic potential (MLIP) code are a promising tool due to their affordability and high accuracy.⁴⁶ As a short-range potential, MTP has its own limitations.^{47,48} To address this, the ML-derived structure ranking was combined with insights from crystal engineering and the synthon approach, as established in our previous research.^{49–51} The Results and discussion section examines MTP in more detail.

Our research group focuses on integrating CSP and crystal engineering. Based on our previous research on the importance of synthons in identifying both kinetic and thermodynamic structures, the purpose of this article is to explore the outcomes of employing CSP with ML in conjunction with this approach. In this research, the 3-(5-amino-2-methyl-4-oxoquinazolin-3(4H)-yl)piperidine-2,6-dione (avadomide⁵²) molecule was selected for this purpose (Fig. 1). Avadomide is a new small pharmaceutical molecule with

potent immunomodulatory and antitumor properties.^{52,53} The powder X-ray diffraction (PXRD) pattern of avadomide was revealed in the international patent number AU 2017225168 B2, but its crystal structure remains unsolved. Avadomide and similar compounds are described in the international patent number EP4327878A2, including their methods of treatment and the prescribed unit doses. Also, the effectiveness of avadomide was examined in treating patients who were unable to tolerate standard cancer treatments based on the prescribed dosage.⁵⁴ The application of CSP combined with crystal engineering provides an opportunity to explore the structural landscape of avadomide. This approach can address challenges such as the drug's low solubility while enhancing its medicinal properties, all without disturbing the chemical form of the drug.

Methods

In this study, hierarchical design was employed to perform CSP calculations to solve complex problems systematically by breaking them down into manageable steps.⁵⁵ The energy profile with respect to torsion angles was analyzed to investigate the conformational states of the avadomide molecule, employing the Dreiding force field with the equilibration atomic charge (Q_{eq}). This gave us the initial molecular geometries which were then relaxed using the Gaussian09 program.⁵⁶ Stable conformational states were optimized using DFT at the B3LYP/6-311++G** (p,d) level of theory. In order to generate avadomide structures in the crystallographic asymmetric unit with one independent molecule ($Z' = 1$), two conformational forms and the most commonly occurring space groups ($P\bar{1}$, Pc , $P2_1/c$, $P2_1/c$, $P2_1/m$, Cc , Cm , $C2/c$, $Pca2_1$, $Pna2_1$, $Pbca$, $Pbcn$, $Aba2$, $Pnna$ and $Pccn$) were used. The evolutionary algorithm, as implemented in the USPEX 10.5 code,^{57–60} was used to search for low-energy crystal structures. Structure relaxation was done using the Dreiding force field⁶¹ within the GULP code,⁶² with atomic charges determined by atomic charge, Q_{eq} .⁶³ The rigidity of the molecular structure was maintained during crystal structure relaxation. Then, structures exhibiting similarity based on an RMSD15 < 0.15 Å criterion⁶⁴ were eliminated using Mercury software,⁶⁵ and unique structures underwent re-optimization using the density functional tight binding (DFTB) and ML method.^{43,66–68}

This study employed the most widely used DFTB variant for organic molecules. The DFTB calculations were performed using a third-order density functional Hamiltonian with self-consistent charge redistribution, as implemented in the DFTB⁺ code.^{69–72} Slater-Koster 3OB parameter files were used, and the missing London dispersion interactions were accounted for using the Tkatchenko-Scheffler (TS) correction.⁶⁶ For some structures the DFTB energies did not fully converge.

For full structure optimization, the next optimization step involved using the MTP potentials within the MLIP code,⁷³ coupled with an active learning methodology. For creation of

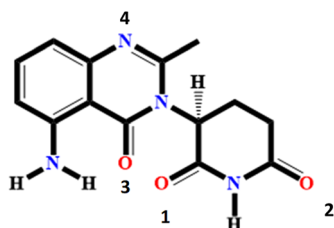


Fig. 1 Chemical structure of avadomide, with numbering of some atoms.

training sets, the Vienna *Ab initio* Simulation Package (VASP) code^{74–76} was utilized. In these calculations, we used the PBE-D3 approach and the projector augmented wave (PAW) method.

For the final optimization step, one hundred lowest-energy structures were re-optimized using the PBE-D3 method,⁷⁷ implemented in VASP. In these calculations, a kinetic energy cutoff of 700 eV was utilized. The Brillouin zone was sampled with a *k*-point grid resolution of $2\pi \times 0.06 \text{ \AA}^{-1}$. The convergence criteria were set to 1×10^{-5} eV per cell for total energies within the self-consistent loop and 5×10^{-3} eV $\text{\AA}^{-1}$ for forces in structure relaxation.

A model illustrating how molecules similar to avadomide may be oriented and packed in crystal structures can facilitate the selection of the most likely configurations. Structural information from the CSD⁷⁸ was used to determine and analyze the structural search space. ConQuest⁷⁹ (CSD version 5.38, updated November 2021) was employed to identify structures similar to avadomide. Considering that the crystal structure of avadomide is not known, the molecule was divided into several fragments for the search. Based on our knowledge of compounds related to avadomide, the intermolecular interactions of each fragment were analyzed. Detailed information for each query is available in the SI. In the CSD search, structures were selected based on criteria including an *R* factor of ≤ 0.075 , no disorder, no structural errors, no molecular polymerization, and no ions. The Mercury software⁸⁰ in the CSD package was used to study crystal packing, intermolecular interactions, and crystallographic analysis of experimental structures. The structures similar to avadomide reported in the CSD were analyzed for interactions and synthon geometries (see the SI for more details).

Results and discussion

Conformational analysis

Avadomide, a racemic^{81–83} pharmaceutical compound, has a flexible torsion angle and can adopt different conformational states in its solid forms, which affect crystal structure generation. To explore these conformational states, the torsion angles were calculated using the Dreiding force field combined with atomic charge (Q_{eq}). The energy profile of the molecule in the gas phase, plotted as a function of its torsion angle (see Fig. S2), revealed two possible conformational states. The energies of these two conformations were

calculated using the B3LYP/6-311++G**(p,d) method. Form I and form II (Fig. 2) have torsion angles of 123.99° and -50.97° , respectively, with form I having lower energy by $6.588 \text{ kcal mol}^{-1}$. A notable feature of avadomide and its related structures is that the hydrogen atom attached to the chiral carbon lies in the same plane as the carbonyl group containing oxygen atom 3. This feature is present in both form I and form II (details provided in the SI). Additionally, analysis of the piperidinedione ring and OH group orientations revealed specific positions in two unstable states where steric hindrances resulted in high-energy conformations. The energy differences between the conformational states, as detailed in Table S1, demonstrate a significant variation among the conformational states.

Crystal structure generation

Two forms, I and II, were used for generating crystal structures in USPEX, with $Z' = 1$ in common space groups ($P\bar{1}$, Pc , $P2_1/c$, $P2_1/c$, $P2_1/m$, Cc , Cm , $C2/c$, $Pca2_1$, $Pna2_1$, $Pbca$, $Pbcn$, $Aba2$, $Pnna$, and $Pccn$) at $T = 0 \text{ K}$ and ambient pressure. A symmetric random structure generator⁵⁷ was used to generate 120 structures in the initial generation. In the following generations, 100 structures were generated from low-energy configurations: 50% through heredity, 10% through soft mutation, 10% through rotational mutation, and 30% using a random symmetric structure generator. The GULP code and Dreiding force field were used to relax the generated structures. Approximately 13 000 candidate structures were generated during the evolutionary search. The evolutionary searches were stopped when the best structure remained unchanged for 30 generations.

Next, duplicate structures were eliminated and low-energy structures within 60 kJ mol^{-1} of the most stable generated structure were identified. Duplicate structures were found using the packing similarity search feature of Mercury software, which was based on an RMSD threshold of $0.15 \text{ \AA}$. Structures with RMSD greater than $0.15 \text{ \AA}$ were considered different. At this stage, 4316 crystal structures for form I and form II remained, which were then used in the subsequent re-optimization steps.

The generated structures were re-optimized using DFTB calculations. Since some structures could not be relaxed fully, they were processed using ML techniques.

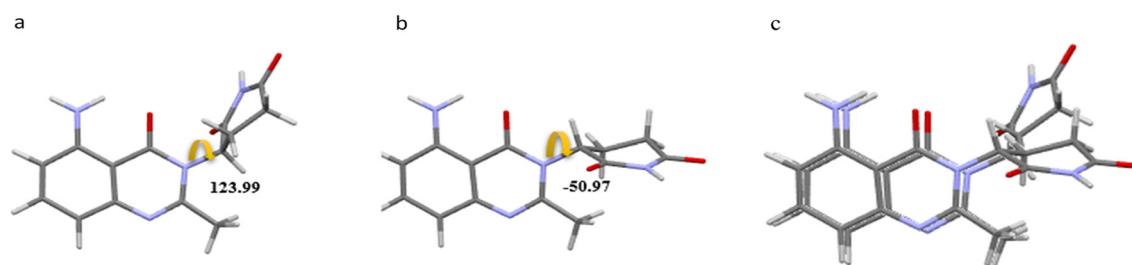


Fig. 2 Molecular structures of (a) form I, (b) form II, and (c) overlap of form I and form II.

Optimization step: machine learning potential

In this study, we employed a method combining machine learning potentials, using MTPs as the ML-based interatomic interaction model. In the ML process, the first step is to generate appropriate reference data for the ML model. Machine learning potentials (MLPs), such as MTPs, utilize training data from quantum-mechanical calculations to learn the relationships between atomic configurations and their energies, focusing on the local atomic environment. Interatomic machine-learning potentials model the total energy of a system as the sum of atomic energies, where the energy assigned to each atom depends on its local chemical environment.

$$E^{\text{mtp}}(\text{cfg}) = \sum_{i=1}^n V(n_i) \quad (1)$$

A key element of constructing any ML model is the training set. Employing a comprehensive and complete training set enhances the accuracy of the machine learning potential and facilitates training over a larger area of the energy landscape. However, it also significantly increases computational costs. Traditionally, the construction of training sets used to be made manually and by trial and error, where in every iteration, the performance of the trained potential is evaluated. Active learning simplifies the process and improves efficiency.

For each step of the machine learning potential model (MTP potential, training, and active learning) a set of hyperparameters must be defined, which will be outlined in the following. Fig. 3 provides an overview of the complete CSP and ML methodology, including the key components and processes involved. In the following, we will discuss each step individually.

Training and active learning

In the initial stage of training on quantum-mechanical data, 500 structures were randomly selected as the first training set to ensure broad coverage of configurational space, including different molecular packings and conformations of

avodomide. The cutoff parameters were set to $R_{\text{cut}} = 5 \text{ \AA}$ and $R_{\text{min}} = 0.8 \text{ \AA}$. Thus, configurations that have atoms distanced by less than $0.8 \text{ \AA}$ are filtered out. The training parameters were configured with energy weight = 1, force weight = 0.1 and stress weight = 0.1.

After training the potential, active learning was performed. In this process, an active learning algorithm identifies instances where an MTP extrapolates beyond its training domain and incorporates those extrapolated configurations into the training dataset. To evaluate structures in this model, the D-optimality criterion was applied,⁴⁷ which measures the degree of extrapolation and determines how far a configuration deviates from the training set. Structures with high extrapolation indicate regions of the configuration space that are not well-covered with the current training set, highlighting areas requiring additional data for improved model accuracy. In the first step of relaxation, 98 structures with the highest degree of extrapolation (D-optimality) were selected. These structures were treated using DFT calculations and subsequently included in the training set. This iterative approach ensures that the training set adequately covers. Diverse configurations improve the robustness of the MTP model.

Subsequently, during the active learning cycles, the machine learning model assigned an extrapolation grade to each structure, which quantified how far the configuration was from the current training domain. The 98 structures with the highest extrapolation degree (D-optimality), automatically identified by the MLIP package, were selected for further DFT calculations. These updated structures are then utilized in active learning computations, and all configurations are relaxed using the revised training set. In the initial stage, structures are selected randomly; however, in subsequent iterations, configurations with the highest degree of extrapolation are prioritized for inclusion in the training set. Through eight iterations of training, eight machine-learning potentials with the same functional form

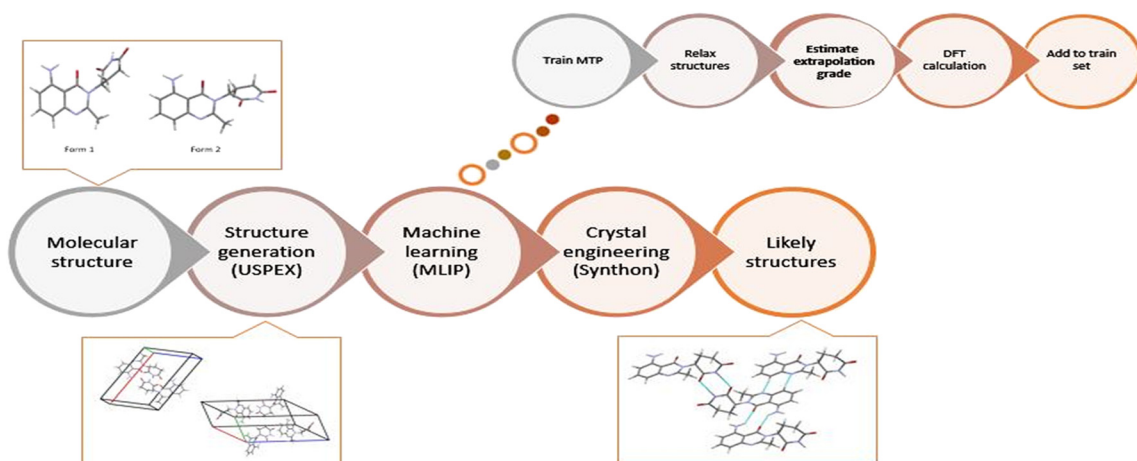


Fig. 3 Workflow for crystal structure prediction (CSP) and the machine learning scheme.

Table 1 Active learning information of 8 trains with active learning

Train number	Extrapolate structures	Extrapolation grade	Train data set
1	0	10 254.61	500
2	98	1039.08	598
3	98	520.96	696
4	98	41.75	794
5	98	9.71	892
6	98	4.61	990
7	98	1.86	1088
8	98	1.42	1186

but distinct parameter values were generated. In the initial stage, the MTP parameters are assigned either randomly or based on prior structural information. In the subsequent stages, these parameters are iteratively refined through an active learning process.

The reliability of the active learning loop is indicated when the extrapolation grade falls within the range of 1 to 3. Under these conditions, the errors remain consistent, and the uncertainty in these errors is minimal, reflecting the quality of the training process. The evolution of the MLIP model during training is summarized in Table 1. To evaluate the reliability of the ML potential, the RMSE errors were assessed. The RMSE error ranged from 1.5 to 0.3 meV per atom observed for specific iterations.⁸⁴ These errors are within acceptable thresholds, validating the accuracy of the MTP model for the present study.

The one hundred lowest-energy structures for form I and form II were selected for DFT calculation in the VASP package. Previous studies have demonstrated that the PBE-D3 method yields very accurate and reliable results for organic pharmaceutical compounds.^{77,85} These computations employed PBE-D3, with a Γ -centered k -point grid with reciprocal space resolution of $2\pi \times 0.06 \text{ \AA}^{-1}$. A kinetic energy cutoff of 700 eV was applied. The details of the lowest-energy structures for forms I and II are provided in Table S3. Fig. 4 illustrates the relationship between the energy calculated using PBE-D3 and the density of the predicted structures, within 20 kJ mol⁻¹ of the lowest energy.

Powder X-ray diffraction (PXRD)

To identify which of these structures matches the experimental structure, the experimental PXRD pattern of avadomide (from patent AU 2017225168 B2) was used for comparison. The empirical data on avadomide are limited, and the quality of the information available in the patent is relatively low.

PXRD patterns of low-energy candidate structures were examined and it became clear that no structure exhibited perfect similarity to the reference model. The PXRD pattern of the structure ranked 4th in conformational form I showed the closest resemblance to the experimental pattern. This observation led us to hypothesize that this particular structure might be the observed polymorph. Alternatively, the PXRD pattern might reflect a mixture of two polymorphic forms, making interpretation more complex. Previous studies demonstrate that a significant number of compounds—estimated between 15% and 45%—may not crystallize in their thermodynamically most stable form, as access to this form can be restricted by kinetic hindrance.^{86,87} Fig. S5 in the SI illustrates the comparison between the PXRD pattern of this candidate structure and the experimental PXRD pattern.

Crystal engineering

Energy is the primary factor in evaluating the obtained structures. CSP studies and blind tests have demonstrated that modern computational methods can almost accurately calculate the energy ranking of polymorphs.^{87–91} Both thermodynamics and kinetics should be considered when selecting likely structures. As previously noted, direct access to kinetic information is challenging, but applying crystal engineering concepts offers a way to obtain this information indirectly.^{49–51} For this purpose, the synthon approach was employed to analyze structures within 20 kJ mol⁻¹ of the lowest-energy structure (Table S3).

In addition, the ten lowest-energy structures relative to the most stable structure were meticulously analyzed using Mercury software, based on the crystal engineering (Table 2). The main goal is to identify the most likely structure based

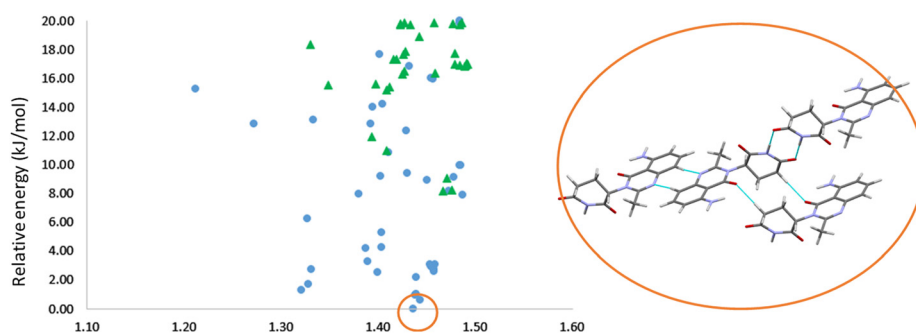


Fig. 4 Relative energy (ΔE (PBE + D3), kJ mol⁻¹) plotted against the density of the predicted structures of avadomide and first rank structure of avadomide.

Table 2 List of low-energy structures of avadomide. The structures were optimized using PBE-D3

Rank	Volume (Å ³)	Conformation	Density (g cm ⁻³)	SPG	Z	Relative energy (kJ mol ⁻¹)	Synthons I(I')II(II')III
1	662.15	1	1.44	<i>P</i> $\bar{1}$	2	0.00	+++
2	1317.48	1	1.44	<i>P</i> 2 ₁ / <i>c</i>	4	0.60	+--
3	1321.98	1	1.44	<i>P</i> 2 ₁ / <i>c</i>	4	0.91	+-
4	1321.63	1	1.44	<i>P</i> 2 ₁ / <i>c</i>	4	1.01	+-
5	2878.01	1	1.32	<i>Pbca</i>	8	1.25	---
6	2863.29	1	1.33	<i>Pbcn</i>	8	1.67	+-
7	660.46	1	1.44	<i>P</i> $\bar{1}$	2	2.17	+++
8	1358.69	1	1.40	<i>Pna</i> 2 ₁	4	2.48	---
9	651.95	1	1.46	<i>P</i> $\bar{1}$	2	2.60	+-
10	2857.34	1	1.33	<i>Pbca</i>	8	2.71	---

Structural synthons were analyzed in three dimensions. A '+' sign indicates the presence of a synthon in that dimension, while a '-' sign indicates its absence.

on the experimental synthons in the CSD within the top-ranked structures.

In Fig. 1, specific atoms of the avadomide molecule are labelled to define the synthons.

Oxygen atoms O1 and O2, along with the imide group in the piperidinedione ring, can participate in hydrogen bonding. This interaction commonly occurs as an N-H...O hydrogen bond, forming an R₂²(8) synthon with the imide group of a neighboring molecule. This interaction is referred to as synthon I. The imide group, with the oxygen atom O2, can additionally participate in a C-H...O hydrogen bond, forming an R₂²(8) synthon through interactions with the oxygen atom of the imide group and the aliphatic hydrogens of the piperidinedione ring in a neighbouring molecule. This interaction is referred to as synthon I'.

The oxygen atom O3 of the carbonyl group in the quinazoline fragment commonly engages in a C-H...O hydrogen bond with C-H groups, referred to as synthon II. Additionally, this oxygen atom can participate in an N-H...O hydrogen bond with N-H groups, which is designated as synthon II'.

Nitrogen atom N4 in the quinazoline fragment typically interacts through a C-H...N hydrogen bond with a C-H group as hydrogen donor, forming an R₂²(8) synthon. This interaction is referred to as synthon III.

The low-energy structures of avadomide were identified in the space groups *P* $\bar{1}$, *P*2₁/*c*, *Pbcn*, and *Pbca* in conformational form I. The first-ranked structure, possessing space group *P* $\bar{1}$ with *Z'* = 1, is illustrated in Fig. 4. In this structure, the N-H...O hydrogen bond involving the oxygen atom (O2) of the imide group in the piperidinedione ring, together with the imide group of the piperidinedione ring in a neighbouring molecule, forms an R₂²(8) synthon. This synthon is characterized by a bond distance of 1.81 Å (H...A) and a bond angle of 178.87° (D-H...A). This synthon, frequently observed in CSD structures, is referred to as synthon I. The oxygen atom (O3) of the carbonyl group (C=O) in the quinazoline fragment forms a C-H...O hydrogen bond with a C-H group, corresponding to synthon II. Synthon II is characterized by a hydrogen bond distance of 2.09 Å (H...A) and a bond angle of 139.54° (D-H...A). Additionally, the nitrogen atom (N4) in the quinazoline fragment forms a C-H...N hydrogen bond with a C-H group, resulting in an R₂²(8) synthon. This synthon, commonly observed in the CSD, is referred to as synthon III. This synthon is characterized by a bond distance of 2.36 Å (H...A) and a bond angle of 174.36° (D-H...A) (see Fig. 5 (left)). These synthons play a crucial role in expanding the molecular structure in three dimensions, thereby contributing to the overall stability.

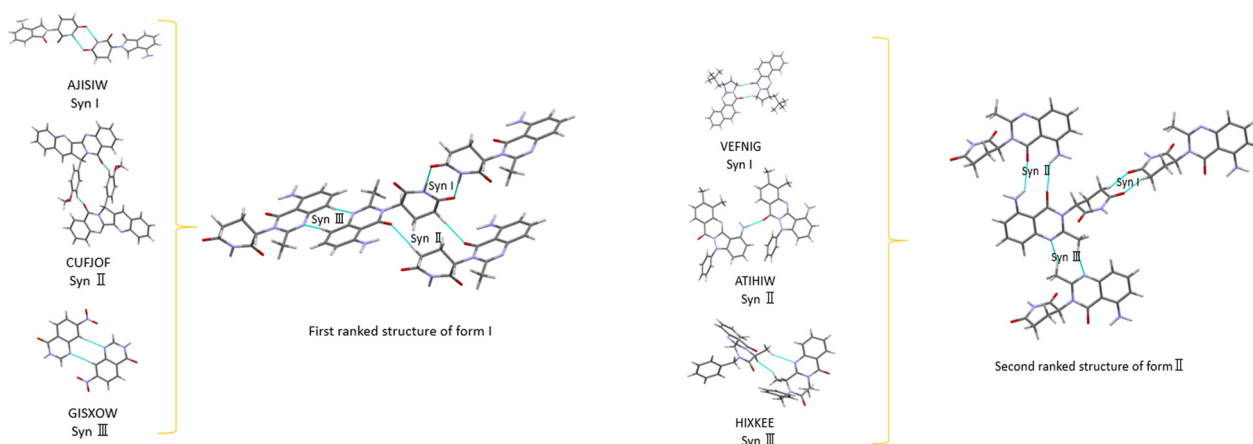


Fig. 5 Synthon details of the first ranked structure of form I (left) and second ranked structure of form II (right).

The structures of conformational form II, as listed in the ranking table, were analyzed using crystal engineering principles, with a specific focus on their synthons (Table S3). Notably, the second-ranked structure of conformational form II (overall rank: 20), which crystallizes in the space group $P\bar{1}$ with $Z' = 1$, exhibits three prominent synthons, emerging as a promising candidate from a crystal engineering perspective (Fig. 5 (right)). In this structure, the oxygen atom (O2) of the imide group in the piperidinedione ring forms a C–H $\cdots$ O hydrogen bond with the hydrogens of the piperidinedione ring in a neighbouring molecule, represented by the $R_2^2(8)$ synthon. This $R_2^2(8)$ interaction is characterized by a bond distance of 2.3 Å (H $\cdots$ A) and a bond angle of 125.34° (D–H $\cdots$ A). A review of the interaction behaviour of the imide group within the piperidinedione ring in the top-ranked structures of conformational form II indicates a transition in the dominant hydrogen bonding. The commonly observed N–H $\cdots$ O hydrogen bonding shifts to a C–H $\cdots$ O hydrogen bond in these structures. This change in interaction behaviour reflects a structural adaptation, likely driven by steric hindrance, resulting in the formation of an alternative hydrogen-bonding pattern, designated as synthon I'. Oxygen atom 3 (O3) of the carbonyl group in the quinazoline fragment participates in an N–H $\cdots$ O hydrogen bond, which is referred to as synthon II'. Synthon II' is characterized by a bond distance of 2.46 Å (H $\cdots$ A) and a bond angle of 172.94° (D–H $\cdots$ A). Nitrogen atom N4 in the quinazoline fragment participates in a C–H $\cdots$ N hydrogen bond, forming an $R_2^2(8)$ synthon. This synthon is characterized by a hydrogen bond distance of 2.57 Å (H $\cdots$ A) and a bond angle of 132.54° (D–H $\cdots$ A), corresponding to synthon III (Fig. 5 (right)). The refcodes for similar experimental structures are provided in Table S4.

Analysis of crystal structures revealed that the first-ranked structure, conformational form I, is characterized by three dominant synthons. This observation underscores its thermodynamic stability and aligns with key principles of crystal engineering, suggesting that form I is likely the dominant form under experimental conditions. The structure ranked 20th, originating from conformational form II, also exhibits three notable synthons. This structure has an energy difference of approximately 8.12 kJ mol^{−1} when compared to the first-ranked structure. The relatively small energy difference (less than 10 kJ mol^{−1}) between these two structures, combined with their identical space group and the presence of three confirmed synthons, provides strong evidence that the 20th-ranked structure is a feasible conformational polymorph of the first-ranked structure. Furthermore, this finding implies that the 20th-ranked structure represents a metastable polymorph of avadomide.

Conclusions

In this study, avadomide was selected as a model compound for polymorph screening due to its significance in the pharmaceutical industry, where the identification and control of polymorphic forms are essential for ensuring drug efficacy,

stability, and intellectual property protection. The crystal structures of avadomide were predicted using the USPEX code combined with a machine learning potential (MTP) model. The integration of ML with traditional computational methods has shown significant potential in advancing CSP. By leveraging ML methods, such as MTP, alongside DFT methods, more accurate and efficient predictions of stable and metastable structures can be achieved. This method offers a cost-effective alternative to DFT while maintaining sufficient computational efficiency. The results of this study emphasize the importance of applying ML methods in conjunction with the concepts of crystal engineering to overcome challenges such as polymorphism and identification of kinetically favoured polymorphs. By utilizing USPEX, trial structures of avadomide were generated based on two lowest-energy conformations of the molecule, conformational states I and II. Duplicate structures were eliminated, and the low-energy structures within 60 kJ mol^{−1} of the most stable generated structure were selected for further analysis. These structures were subsequently optimized and ranked using MTP machine learning interatomic potential. In the final optimization steps, the PBE-D3 method was used for the final structures, ensuring accurate and efficient calculations. The lowest-energy structures were analyzed for the presence of synthons using Mercury software, providing a detailed understanding of crystal packing. Synthons of each structure were compared to those of the experimental structures, and any matches were considered. One candidate structure showed partial agreement with the PXRD pattern reported in the patent; however, the low quality of the experimental PXRD data limited the certainty of a definitive structure assignment.

In addition to this structure, synthon-based analysis and evaluation of low-energy structures led to the identification of two other plausible polymorphs. Although these alternative forms have not yet been observed experimentally, their predicted stability and structural features suggest they could be accessible through future experimental efforts.

Yet it is possible that some low-energy structures have not been identified, which is expected given the limited number of space groups explored. Furthermore, there may still be unidentified structures, as polymorph screening has been performed on only two conformations.

Conflicts of interest

There are no conflicts to declare.

Data availability

The data supporting this article have been included as part of the supplementary information (SI).

Supplementary information: the SI includes CSD comparisons, energies, CSP optimization, PXRD and top 10 predicted CIFs. See DOI: <https://doi.org/10.1039/d5ce00849b>.

Acknowledgements

A. S. and F. B. respectfully acknowledge the Ferdowsi University of Mashhad Research Council for the financial support of project code number 3/57777. We thank the Russian Science Foundation (grant 19-72-30043) for supporting this work. The authors acknowledge the Cambridge Crystallographic Data Centre for the access to CSD Enterprise.

Notes and references

- 1 A. V. Joshi, *Machine learning and artificial intelligence*, CRC Press, 2020.
- 2 S. Zhong, K. Zhang, M. Bagheri, J. G. Burken, A. Gu, B. Li, X. Ma, B. L. Marrone, Z. J. Ren and J. Schrier, *Environ. Sci. Technol.*, 2021, **55**, 12741–12754.
- 3 D. Doudesis, K. K. Lee, J. Boeddinghaus, A. Bularga, A. V. Ferry, C. Tuck, M. T. Lowry, P. Lopez-Ayala, T. Nestelberger and L. Koechlin, *Nat. Med.*, 2023, **29**, 1201–1210.
- 4 A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao and T. Wongjirad, *Nature*, 2018, **560**, 41–48.
- 5 A. S. Walker and J. Clardy, *J. Chem. Inf. Model.*, 2021, **61**, 2560–2571.
- 6 K. A. Brown, S. Brittman, N. Maccaferri, D. Jariwala and U. Celano, *Nano Lett.*, 2019, **20**, 2–10.
- 7 K. Guo, Z. Yang, C.-H. Yu and M. J. Buehler, *Mater. Horiz.*, 2021, **8**, 1153–1172.
- 8 G. R. Schleder, A. C. Padilha, C. M. Acosta, M. Costa and A. Fazzio, *JPhys Mater.*, 2019, **2**, 032001.
- 9 E. Kiely, R. Zwane, R. Fox, A. M. Reilly and S. Guerin, *CrystEngComm*, 2021, **23**, 5697–5710.
- 10 C. H. Chan, M. Sun and B. Huang, *EcoMat*, 2022, **4**, 12194.
- 11 J. Graser, S. K. Kauwe and T. D. Sparks, *Chem. Mater.*, 2018, **30**, 3601–3612.
- 12 E. V. Podryabinkin, E. V. Tikhonov, A. V. Shapeev and A. R. Oganov, *Phys. Rev. B*, 2019, **99**, 064114.
- 13 M. Annadhasan, S. Basak, N. Chandrasekhar and R. Chandrasekar, *Adv. Opt. Mater.*, 2020, **8**, 2000959.
- 14 C. Yao, I. A. Guzei, Y. Jin, S. Ruan, G. Sun, Y. Gui, L. Wang and L. Yu, *Cryst. Growth Des.*, 2020, **20**, 7874–7881.
- 15 O. Egorova, R. Hafizi, D. C. Woods and G. M. Day, *J. Phys. Chem. A*, 2020, **124**, 8065–8078.
- 16 S. A. Tawfik, O. Isayev, M. J. Spencer and D. A. Winkler, *Adv. Theory Simul.*, 2020, **3**, 1900208.
- 17 J. P. Lommerse, W. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. Hofmann, F. J. Leusen, W. T. Mooij, S. L. Price and B. Schweizer, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2000, **56**, 697–714.
- 18 W. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Dzyabchenko, P. Erk, A. Gavezzotti, D. W. Hofmann, F. J. Leusen, J. P. Lommerse and W. T. Mooij, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 647–661.
- 19 A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk and A. Tkatchenko, *Chem. Sci.*, 2017, **8**, 4926–4940.
- 20 X. Li, X. Ou, B. Wang, H. Rong, B. Wang, C. Chang, B. Shi, L. Yu and M. Lu, *Commun. Chem.*, 2020, **3**, 152.
- 21 B. K. Hodnett and V. Verma, *Processes*, 2019, **7**, 272.
- 22 N. Krishna Kumar and R. Suryanarayanan, *Mol. Pharmaceutics*, 2022, **19**, 472–483.
- 23 Y. V. Torubaev and A. S. Samigullina, *Chem*, 2022, **4**, 196–205.
- 24 I. Ahmad, A. A. Ganie and A. A. Dar, *CrystEngComm*, 2020, **22**, 3933–3942.
- 25 Z. N. Baghan, A. Salimi, H. Eshtiagh-Hosseini and A. G. Oliver, *CrystEngComm*, 2021, **23**, 6276–6290.
- 26 F. Tavakoli-Quchani, A. Salimi and A. N. Pour, *CrystEngComm*, 2024, **26**, 4708–4719.
- 27 A. K. Nangia and G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2019, **58**, 4100–4107.
- 28 S. Saha, M. K. Mishra, C. M. Reddy and G. R. Desiraju, *Acc. Chem. Res.*, 2018, **51**, 2957–2967.
- 29 A. Samie and A. Salimi, *CrystEngComm*, 2019, **21**, 3721–3730.
- 30 J. Sarma and G. R. Desiraju, *Cryst. Growth Des.*, 2002, **2**, 93–100.
- 31 A. Samie, A. Salimi and J. C. Garrison, *CrystEngComm*, 2019, **21**, 6432–6445.
- 32 G. Bolla, B. Sarma and A. K. Nangia, *CrystEngComm*, 2021, 157–229.
- 33 D. D. Kara and M. Rathnanand, *Crystals*, 2022, **12**, 1337.
- 34 G. Sun, X. Liu, Y. A. Abramov, S. O. Nilsson Lill, C. Chang, V. Burger and A. Broo, *Cryst. Growth Des.*, 2021, **21**, 1972–1983.
- 35 S. L. Price and L. S. Price, *Polymorphism in the Pharmaceutical Industry: Solid Form and Drug Development*, 2018, pp. 133–157.
- 36 M. K. Dudek and K. Druzbicki, *CrystEngComm*, 2022, **24**, 1665–1678.
- 37 Z. N. Baghan, A. Salimi, H. Eshtiagh-Hosseini and A. G. Oliver, *CrystEngComm*, 2019, **21**, 2691–2701.
- 38 A. Banerjee, D. Jasrasaria, S. P. Niblett and D. J. Wales, *J. Phys. Chem. A*, 2021, **125**, 3776–3784.
- 39 F. Tavakoli-Quchani, A. Salimi, B. Notash, A. Samie and J. C. Garrison, *CrystEngComm*, 2023, **25**, 299–308.
- 40 J. Nyman and G. M. Day, *CrystEngComm*, 2015, **17**, 5154–5165.
- 41 R. J. Clements, J. Dickman, J. Johal, J. Martin, J. Glover and G. M. Day, *MRS Bull.*, 2022, **47**, 1054–1062.
- 42 A. R. Oganov, *Faraday Discuss.*, 2018, **211**, 643–660.
- 43 S. Wengert, G. Csányi, K. Reuter and J. T. Margraf, *Chem. Sci.*, 2021, **12**, 4536–4546.
- 44 C. Greenwell, J. L. McKinley, P. Zhang, Q. Zeng, G. Sun, B. Li, S. Wen and G. J. Beran, *Chem. Sci.*, 2020, **11**, 2200–2214.
- 45 I. Novoselov, A. Yanilkin, A. Shapeev and E. Podryabinkin, *Comput. Mater. Sci.*, 2019, **164**, 46–56.
- 46 M. Zeraati, A. R. Oganov, T. Fan and S. F. Solodovnikov, *Phys. Rev. Mater.*, 2024, **8**, 033601.
- 47 I. S. Novikov, K. Gubaev, E. V. Podryabinkin and A. V. Shapeev, *Mar. Life Sci. & Technol.*, 2020, **2**, 025002.
- 48 E. V. Podryabinkin and A. V. Shapeev, *Comput. Mater. Sci.*, 2017, **140**, 171–180.
- 49 Z. M. Abardeh, A. Salimi and A. R. Oganov, *CrystEngComm*, 2022, **24**, 6066–6075.

- 50 Z. Momenzadeh Abardeh, F. Bahrami and A. R. Oganov, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2024, **80**, 620–627.
- 51 Z. M. Abardeh, A. Salimi and A. R. Oganov, *CrystEngComm*, 2025, **27**, 5514–5526.
- 52 D. W. Rasco, K. P. Papadopoulos, M. Pourdehnad, A. K. Gandhi, P. R. Hagner, Y. Li, X. Wei, R. Chopra, K. Hege and J. DiMartino, *Clin. Cancer Res.*, 2019, **25**, 90–98.
- 53 J.-M. Michot, R. Bouabdallah, U. Vitolo, J. K. Doorduijn, G. Salles, A. Chiappella, P. L. Zinzani, F. Bijou, M. J. Kersten and R. Sarmiento, *Lancet Haematol.*, 2020, **7**, 649–659.
- 54 K. Hatake, T. Chou, T. Doi, Y. Terui, H. Kato, T. Hirose, S. Seo, M. Pourdehnad, Y. Ogaki and H. Fujimoto, *Cancer Sci.*, 2021, **112**, 331–338.
- 55 S. L. Price, *Faraday Discuss.*, 2018, **211**, 9–30.
- 56 M. J. Frisch, G. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci and G. Petersson, *Uranyl extraction by N, N-dialkylamide ligands studied by static and dynamic DFT simulations*, Gaussian Inc., Wallingford, 2009.
- 57 Q. Zhu, A. R. Oganov, C. W. Glass and H. T. Stokes, *Acta Crystallogr.*, 2012, **68**, 215–226.
- 58 A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.
- 59 A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 1–15.
- 60 A. R. Oganov, A. O. Lyakhov and M. Valle, *Acc. Chem. Res.*, 2011, **44**, 227–237.
- 61 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
- 62 J. D. Gale and A. L. Rohl, *Mol. Simul.*, 2003, **29**, 291–341.
- 63 A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.*, 1991, **95**, 3358–3363.
- 64 J. A. Chisholm and S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228–231.
- 65 C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens and M. Towler, *J. Appl. Crystallogr.*, 2020, **53**, 226–235.
- 66 A. Tkatchenko, R. A. DiStasio Jr., R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.
- 67 A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 073005.
- 68 B. Hourahine, B. Aradi, V. Blum, F. Bonafe, A. Buccheri, C. Camacho, C. Cevallos, M. Deshayé, T. Dumitrică and A. Dominguez, *J. Chem. Phys.*, 2020, **152**, 152–172.
- 69 M. Kubillus, T. Kubar, M. Gaus, J. Rezac and M. Elstner, *J. Chem. Theory Comput.*, 2015, **11**, 332–342.
- 70 X. Lu, M. Gaus, M. Elstner and Q. Cui, *J. Phys. Chem. B*, 2015, **119**, 1062–1082.
- 71 M. Gaus, X. Lu, M. Elstner and Q. Cui, *J. Chem. Theory Comput.*, 2014, **10**, 1518–1537.
- 72 M. Gaus, A. Goetz and M. Elstner, *J. Chem. Theory Comput.*, 2013, **9**, 338–354.
- 73 E. Podryabinkin, K. Garifullin, A. Shapeev and I. Novikov, *J. Chem. Phys.*, 2023, **159**, 1–9.
- 74 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169.
- 75 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558.
- 76 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251.
- 77 J. Moellmann and S. Grimme, *J. Phys. Chem. C*, 2014, **118**, 7615–7621.
- 78 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.
- 79 I. J. Bruno, J. C. Cole, P. R. Edgington, M. Kessler, C. F. Macrae, P. McCabe, J. Pearson and R. Taylor, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2002, **58**, 389–397.
- 80 R. A. Sykes, P. McCabe, F. H. Allen, G. M. Battle, I. J. Bruno and P. A. Wood, *J. Appl. Crystallogr.*, 2011, **44**, 882–886.
- 81 S. DeWitt, A. W. Czarnik and V. Jacques, *ACS Med. Chem. Lett.*, 2020, **11**, 1789–1792.
- 82 T. Asatsuma-Okumura, T. Ito and H. Handa, *Pharmaceuticals*, 2020, **13**, 95.
- 83 R. U. McVicker and N. M. O'Boyle, *J. Med. Chem.*, 2024, **67**, 2305–2320.
- 84 D. McDonagh, C.-K. Skylaris and G. M. Day, *J. Chem. Theory Comput.*, 2019, **15**, 2743–2758.
- 85 G. Liu, S.-H. Wei and C. Zhang, *J. Cryst. Growth*, 2022, **22**, 5307–5321.
- 86 M. A. Neumann and J. van de Streek, *Faraday Discuss.*, 2018, **211**, 441–458.
- 87 L. M. Hunnisett, N. Francia, J. Nyman, N. S. Abraham, S. Aitipamula, T. Alkhidir, M. Almehairbi, A. Anelli, D. M. Anstine and J. E. Anthony, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2024, **80**, 548–574.
- 88 G. M. Day, T. G. Cooper, A. J. Cruz-Cabeza, K. E. Hejczyk, H. L. Ammon, S. X. Boerrigter, J. S. Tan, R. G. Della Valle, E. Venuti and J. Jose, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2009, **65**, 107–125.
- 89 D. A. Bardwell, C. S. Adjiman, Y. A. Arnautova, E. Bartashevich, S. X. Boerrigter, D. E. Braun, A. J. Cruz-Cabeza, G. M. Day, R. G. Della Valle and G. R. Desiraju, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2011, **67**, 535–551.
- 90 A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell and R. Car, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 439–459.
- 91 L. M. Hunnisett, J. Nyman, N. Francia, N. S. Abraham, C. S. Adjiman, S. Aitipamula, T. Alkhidir, M. Almehairbi, A. Anelli and D. M. Anstine, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2024, **80**, 517–547.