



Cite this: DOI: 10.1039/d2ce00476c

Crystal structure prediction of *N*-halide phthalimide compounds: halogen bonding synthons as a touchstone†

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

We address the crystal structure prediction problem by combining the evolutionary algorithm USPEX (used to predict sets of low-energy crystal structures) and the synthon approach (extracting preferable supramolecular synthons from the Cambridge Structural Database, CSD, and then indicating the most stable structure as well as the most likely metastable polymorphs among the predicted low-energy structures with synthons). This approach is applied to three compounds: *N*-chlorophthalimide (**ncp**), *N*-bromophthalimide (**nbp**), and *N*-iodophthalimide (**nip**). Database search has indicated the importance of halogen bonding $N-X\cdots O=C$ synthons ($X = Cl, Br, I$) as a touchstone. The predicted structures of **ncp** of rank 1 and **nbp** of rank 3 match the experimental ones. Our results imply that the synthon approach can successfully suggest the right structures amongst a handful of lowest-energy predicted structures. Two candidate crystal structures were suggested using the synthon approach for **nip**.

Received 5th April 2022,
 Accepted 8th July 2022

DOI: 10.1039/d2ce00476c

rsc.li/crystengcomm

Introduction

The goal of crystal structure prediction (CSP) is to produce a set of possible crystal structures of a compound by starting from a chemical diagram.^{1–4} In CSP calculations for molecular crystals, due to weakness of intermolecular interactions, there are many crystal structures with close energies; resolving such small energy differences is a challenge for existing theoretical methods, and the experimental structures are often latent among the 100 lowest-energy structures.^{5,6} The energy difference between low-energy structures is tiny, making it difficult to choose the right one. Complementary experimental data can help to determine the right crystal structure with more certainty.^{2,5–7} In practice, an input of theory is very important in structure determination from powder diffraction (SDPD) to obtain the structural model for further refinement⁸ or to help avoid incorrect structure models;⁹ One usually resorts to SDPD when single crystals of sufficient quality are not available.^{10,11}

Polymorphism in organic compounds also complicates the CSP calculations, where molecules adopt different crystal

structures that can be assessed experimentally.^{12–18} If an experimental compound is thermodynamically stable (“thermodynamic polymorph”), it can be found as the global energy minimum, as opposed to metastable polymorphs that correspond to local energy minima. To find all possible polymorphs of a compound among the produced crystal structures, it is necessary to consider both kinetic and thermodynamic factors.^{5,19} In computational methods, only thermodynamic factors are considered and direct access to kinetic factors is very hard.²⁰

Information about kinetic factors implicitly exists in crystal structure databases of polymorphs, such as the Cambridge Structural Database (CSD), from which the repetitiveness of supramolecular interactions (synthons) can be extracted. Intuitively, rapidly crystallizing metastable polymorphs are likely to contain such synthons,²¹ but since synthons are based on the most important intermolecular interactions, they are also often found in stable polymorphs. Systematic analysis of large datasets (such as those found in the CSD) allows one to indicate likely polymorphs, whether kinetic or thermodynamic.²² Indeed, Desiraju emphasized that synthons encapsulate both kinetic and thermodynamic features.^{23–25} If a pattern is seen often enough in the experimental crystal structures (CSD), it can be assumed that this pattern is kinetically favoured (chemical model) and likely to exist in other crystal structures with a similar skeleton of molecules to reproduce a suitable arrangement in the crystal packing (geometric model).²⁶ Often, molecular crystals have small energy differences between different

^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

† Electronic supplementary information (ESI) available. CCDC 2159574. For ESI and crystallographic data in CIF or other electronic format see DOI: <https://doi.org/10.1039/d2ce00476c>

polymorphs, and theoretical calculations may have difficulty resolving these small differences – in these cases, synthons can indicate the most likely ground state. The synthon approach to crystal structure prediction helps to take into account complexities inherent in the crystallisation process that result in forming various polymorphs of compounds with kinetic and thermodynamic factors.^{19,20,23,27,28} In order to guide the search for a significant structural interaction in the crystal structure prediction, Gilmore *et al.* have proposed cluster analysis and dendrogram plots based on CSD data. Cluster analysis has been used to investigate the probability and geometry of intermolecular interactions in a class of related compounds and apply them to the desired molecule.²⁹

Here, we perform classification and analysis of intermolecular interactions using the clustering methods implemented in dSNAP software.^{30,31} In this work, we combine cluster analysis and synthon approaches, focusing on halogen bonding synthons, which leads to a simpler and faster study. The CSP calculations for *N*-halide phthalimides (*N*-chlorophthalimide (**nep**), *N*-bromophthalimide (**nbp**), and *N*-iodophthalimide (**nip**), the chemical structure of which is shown in Fig. 1) were performed using the evolutionary algorithm USPEX.^{32–35} The energy difference between the lowest-energy structures is very small, and the structures produced cannot be reliably ranked. The investigation of the generated crystal structures shows that the main distinction is in the geometry and type of the intermolecular interaction synthon. Understanding the type and geometry of a possible intermolecular interaction in the crystal structure will be useful for detecting plausible candidate structures. This can be achieved by applying crystal engineering concepts to the study of experimental crystal structures in the CSD. Regarding this, the interaction sites of *N*-halide phthalimides were determined to be aromatic CH, CO, and N–X groups. These sites have a potential to form intermolecular interaction synthons by hydrogen bonding (CH $\cdots$ O/X) and type I/II halogen bonding (X $\cdots$ X/O) in the crystal structure.

To understand which interactions can be observed in similar experimental structures (based on the probability and geometry of interactions), cluster analysis was performed on the basis of the synthon approach. Hydrogen and halogen bonds, which are directional interactions and can be easily distinguished by the bond angle, were used in the cluster analysis. The C–H $\cdots$ X hydrogen bond has geometries with a C–H $\cdots$ X angle of about 150–180° (for a weak interaction, the

C–H $\cdots$ X angle (X = N, Cl, and O) is smaller than 130°).^{36,37} In type I halogen bonding, close contacts of a halogen atom with electrophiles are observed at about 90–120°. Meanwhile, in type II halogen bonding, a halogen atom comes into close contact with nucleophiles approximately along the covalent bonds at about 180°.^{38–41} To classify the intermolecular interactions, CSD search was performed and its results were analysed as a dendrogram using dSNAP software. Moreover, the crystal structure of **nbp** was determined using the SDPD method, and the results were compared with a previously reported structure obtained using single-crystal X-ray analysis⁴² and with the CSP structures to confirm the methodology in this study.

Methods

Crystal structure prediction

The molecular structure geometry was optimised by density functional theory (DFT), using the Gaussian 09 program⁴³ and the B3LYP/6-311g (p, d) basis set. The evolutionary algorithm USPEX was used to search for low-energy crystal structures of *N*-halide phthalimides (**nep**, **nbp**, and **nip**); we used the USPEX 10.1 code.^{32–35} In the USPEX calculations, each generation contained 50–100 structures. The first generation was produced using a random symmetric structure generator and using the most common space groups $P\bar{1}$, $P2_1/c$, $P2_12_12_1$, $P2_1$, $Pbca$, $C2/c$, $Pna2_1$, $Pnma$, $Pca2_1$, Cc , $C2$, $P1$, Pc , $P4_1$, and $P4_12_12_1$, and one molecule in the asymmetric unit ($Z' = 1$). The lowest-energy (60%) structures were used to produce the new generation. The second and all subsequent generations consisted of 50 structures, which were produced from low-energy ones by variation operators (50% by heredity, 10% by soft mutation, 10% by rotational mutation), and 30% of each generation were produced by a random symmetric structure generator. All generated structures were relaxed at $T = 0$ K and ambient pressure, and the enthalpy was used as fitness. All structures were optimised using the GULP code⁴⁴ with the Dreiding force field,⁴⁵ and the atomic charges were determined using the QEq method (Fig. S5) (ESI†).⁴⁶ At the post-processing step, we selected low-energy structures, removed duplicate structures (identified using the packing similarity search in Mercury software based on RMSD₁₅ criteria),⁴⁷ and reoptimized all structures using the PBE-D3 method,⁴⁸ (as well as PBE-MBD and PBE0-MBD methods⁴⁹ for the ten lowest-energy structures), as implemented in VASP^{50–52} within the framework of the projector augmented wave (PAW)^{53,54} method. A kinetic energy cutoff of 700 eV, Brillouin zone sampling with a k -point grid of $2\pi \times 0.06 \text{ \AA}^{-1}$ resolution, and convergence criteria of 1×10^{-5} eV per atom for total energies and 5×10^{-3} eV $\text{\AA}^{-1}$ for forces were used. The lowest-energy structures of **nep**, **nbp**, and **nip** are shown in ESI† Tables S1–S3. We also performed zero-point energy (ZPE) calculations for the top-3 structures of **nbp** and **nip**. The

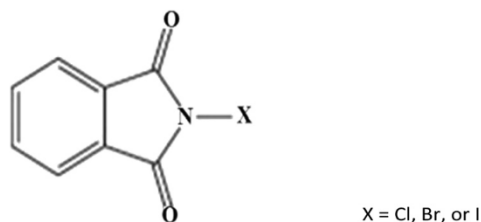


Fig. 1 Chemical structure of *N*-halide phthalimide compounds.

results showed that ZPE changes energy differences by a negligibly small amount (see Table S4†).

CSD search and cluster analysis

To investigate the probability of interactions and the synthon geometry in experimental structures of compounds similar to the compounds studied here, CSD analysis of fragments I and II was performed (Fig. 2). In the CSD search, we selected structures with R factor ≤ 0.07 , no disorder, no errors in the structure, no polymerization of molecules, and no ions (CSD version 5.38, updated November 2021).⁵⁵ In fragment I, the bond angle and length were specified for $N-7A\cdots QB$, where 7A is any halogen atom and $QB = F, Cl, Br, I, O$, or H (Fig. 2a). The number of atoms bonded to 7A was set to 1 (T1), and the cutoff for this interaction was set as the sum of van der Waals radii. In the database search for fragment I, 42 hit structures were found which have 53 hit fragments. The number of hit fragments is larger than the number of hit structures because some of the hits present several interaction cases. The obtained structures were classified on the basis of similarity in bond lengths and angles using dSNAP software.⁵⁶ In addition, fragment II was defined to study the interaction geometry of $N-X\cdots O=C$ separately for $X = Cl, Br, I$ (Fig. 2b). The interaction between the halogen atom and the carbonyl group was specified (the $X\cdots O$ length, $N-X\cdots O$ angle θ_1 , and $X\cdots O=C$ angle θ_2), the N atom was set to have three bonded atoms (T3), and the cutoff for this interaction was set as the sum of van der Waals radii.⁵² In fragment II, there were 20, 9, and 7 hit structures, with 26, 10, and 7 hit fragments for Cl, Br , and I , respectively. The histograms of angles θ_1 ($N-X\cdots O$) and θ_2 ($X\cdots O=C$) are shown in Fig. S1–S4 (ESI†).

The results of cluster analysis are shown as a dendrogram in Fig. 3. The horizontal axis shows the hits (experimental structures). The vertical axis indicates the similarity level between the hit structures in the classification, restricted by the cut level line. The choice of dendrogram cut level, which defines the number of clusters, is a critical issue in any

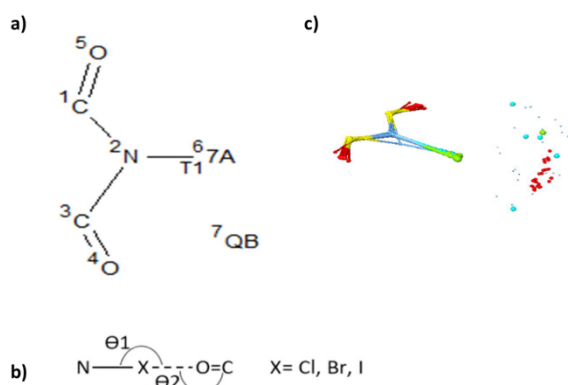


Fig. 2 (a) Fragment I and (b) fragment II of an N -halide phthalimide molecule. (c) Representation of structures including various interactions with fragment I obtained from CSD search before clustering.

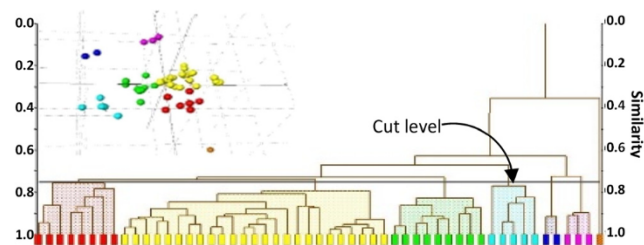


Fig. 3 Dendrogram and MMDS plot (inset) for interactions between sites 7A (halogen atom) and QB (halogen, O, and H atoms). Cluster A (red) contains 8 hits, cluster B (yellow) – 25, cluster C (green) – 9, cluster D (light blue) – 5, cluster E (dark blue) – 2, cluster F (magenta) – 3, and cluster G (brown) – 1.

cluster analysis. In classification methods, estimating the number of clusters is an unsolved problem.²¹ The choice of cut level depends on the level of detail required from the analysis, so one has to take it based on pragmatic factors; here we took it as 0.755. These settings produce clusters that represent a specific type of interaction based on the bond angle. The metric multidimensional scaling (MMDS) plot and the dendrogram agree. More details are presented in the ESI†

SDPD method

A powder microcrystalline sample of **nbp** was loaded into an aluminium sample holder which was rotated during data collection to improve particle statistics and to minimise the preferred orientation effects. The indexing of the powder diffraction pattern at low angles using DICVOL04 software⁵⁷ suggested an orthorhombic system with approximate cell parameters $a = 22.676 \text{ \AA}$, $b = 6.687 \text{ \AA}$, and $c = 5.245 \text{ \AA}$ [$M(14) = 23.4$; $F(14) = 36.2(0.0138, 28)$]. The best space group, estimated on the basis of the systematic absences, was $P2_12_12_1$. The structure was resolved using DASH software⁵⁸ with the simulated annealing technique. The final refinements were performed using the GSAS/EXPGUI software suite;⁵⁹ soft restraints were applied to the bond lengths and angles, with additional planarity restraints. The peak shapes were best described using the Thompson–Cox–Hastings formulation of the pseudo-Voigt function.⁶⁰ The background was described using the shifted first Chebyshev function with 36 points regularly distributed over the entire 2θ range. The preferred orientation was modelled using the spherical harmonics description⁶¹ with eight coefficients. The indexing figures of merit, crystallographic data, and processing parameters for **nbp** are presented in Table S7 (ESI†). The experimental, calculated, and difference patterns are shown in Fig. S6 (ESI†).

Results and discussion

The prediction of the N -halide phthalimide crystal structures was started with N -bromophthalimide (**nbp**). The search for the possible crystal structures was performed in a set of the most common space groups with one molecule in the

asymmetric unit. After 100 generations, 7000 structures were produced. The structures within 50 kJ mol^{-1} of the global minimum were selected. Structures with an identical space group with RMSD_{15} smaller than 0.2 were considered identical. The resulting set of structures was reoptimised using the PBE-D3 functional in the VASP code. We thus identified and reoptimized 20 structures within 15 kJ mol^{-1} of the most stable structure. The low-energy crystal structures belong to $P2_1/c$, $P2_12_12_1$, $P2_1$, and $C2$ space groups. Then, a similar procedure was used for **nbp**. In 30 generations, possible crystal structures with space groups $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, and $P2_12_12_1$ were produced, with $Z' = 1$. After reoptimization, 35 low-energy structures were obtained within 15 kJ mol^{-1} of the most stable one.

To predict the crystal structure of **nip**, 1500 structures were produced with space groups $P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pbca$, $Pna2_1$, $Pnma$, $Pca2_1$, Pc , $P4_1$, and $P4_12_12_1$, with $Z' = 1$, in 35 generations. After reoptimisation, 19 low-energy structures were obtained within 15 kJ mol^{-1} of the most stable one. The low-energy structures belong to $P2_1$, $P2_1/c$, $P2_12_12_1$, and $P4_1$ space groups.

The lowest-energy structures of **nbp**, **nbp**, and **nip** are presented in ESI† Tables S1–S3. We selected the PBE + D3 ranking for CSP landscapes since it includes all structures of interest (within 15 kJ mol^{-1} of the most stable structure). The plots of the PBE + D3 energy versus the density of the predicted structures are shown in Fig. 4. To get an idea of the sensitivity of the results to the computational method, for the ten lowest-energy structures, we performed relaxations and energy calculations with the much more expensive methods of PBE + MBD and PBE0 + MBD.⁴⁹ The relative energies of the ten lowest-energy predicted structures of **nbp**, **nbp**, and **nip** calculated with the methods mentioned are shown in Fig. 5 and Table S5†. The final rankings of the ten lowest-energy structures are based on PBE0 + MBD energies.

The cluster analysis for possible intermolecular interactions of fragment I in the CSD led to the outlining of seven clusters (Fig. 3 and Table S6†). Cluster A (red) contains 8 hits, cluster B (yellow) – 25 (the highest frequency), cluster C (green) – 9, cluster D (light blue) – 5, cluster E (dark blue) –

2, cluster F (magenta) – 3, and cluster G (brown) – 1. The characteristic geometries of these seven clusters are summarised in Table S6†. Cluster A corresponds to a close contact between X (halogen) and H atoms resulting in a hydrogen bond. According to the geometrical parameters of hydrogen bonds (the CH moiety as a donor),³³ this cluster includes four cases with $\text{CH}\cdots\text{X}$ bond angles from 140 to 166° and $\text{N-X}\cdots\text{H}$ from 103 to 109° . Cluster B (25 hits) corresponds to a close contact between X (halogen) and O atoms (22 cases) and $\text{X}\cdots\text{H}$ (3 cases). 22 structures have type II halogen bonding, with $\text{N-X}\cdots\text{O}$ bond angles from 153 to 179° . Cluster C corresponds to a close contact between halogen and H atoms (8 cases) forming a hydrogen bond, and in one case between two halogen atoms resulting in a halogen bond. These eight cases have weak hydrogen bonding, with $\text{CH}\cdots\text{X}$ bond angles from 117 to 135° and $\text{N-X}\cdots\text{H}$ bond angles from 100 to 133° . Cluster D structures have type II halogen bonding $\text{N-X}\cdots\text{X-N}$, with bond angles θ_1 from 153 to 166° and θ_2 from 92 to 135° (2 cases). In the other cases, there is no significant intermolecular interaction. Cluster E corresponds to a close contact between two halogen atoms which leads to type I halogen bonding. Cluster F has three hits, one of which has type I halogen bonding $\text{X}\cdots\text{O}$, whereas in others there is no significant intermolecular interaction. In cluster G, there is only one hit – type II halogen bonding $\text{X}\cdots\text{O}$. Overall, the cluster analysis of fragment I revealed that 70% of the intermolecular interactions occur *via* halogen bonding, with the rest of them resulting from hydrogen bonding. The results showed that halogen atoms interact with oxygen atoms more frequently (85% of cases are $\text{X}\cdots\text{O}$ bonds) than with other halogen atoms (15% of cases are $\text{N-X}\cdots\text{X}$ bonds). In addition, 55% of halogen bonding interactions are of type II halogen bond.

Careful examination of the obtained CSD structures containing fragment I (CSD-structure hereafter for short) showed two types of synthon geometries, the $\text{N-X}\cdots\text{O}$ dimer (only 2 cases) and catemer (all remaining 40 cases) – see Fig. 6. The short-range synthons of halogen bonding in most cases expanded with the participation of the weak hydrogen bonding interactions of $\text{C-H}\cdots\text{O/X}$. For instance, the CSD

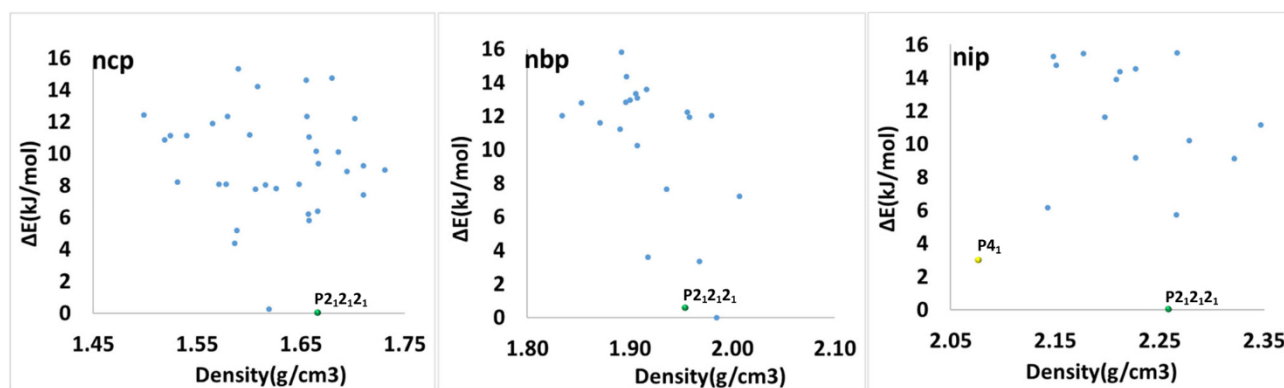


Fig. 4 Relative lattice energy plotted (ΔE (PBE + D3), kJ mol^{-1}) against the density of the predicted structures of **nbp**, **nbp**, and **nip**. The selected structures and the polymorphic forms of **nip** are shown in green and yellow, respectively, with the rank numbers (see Table S5†).

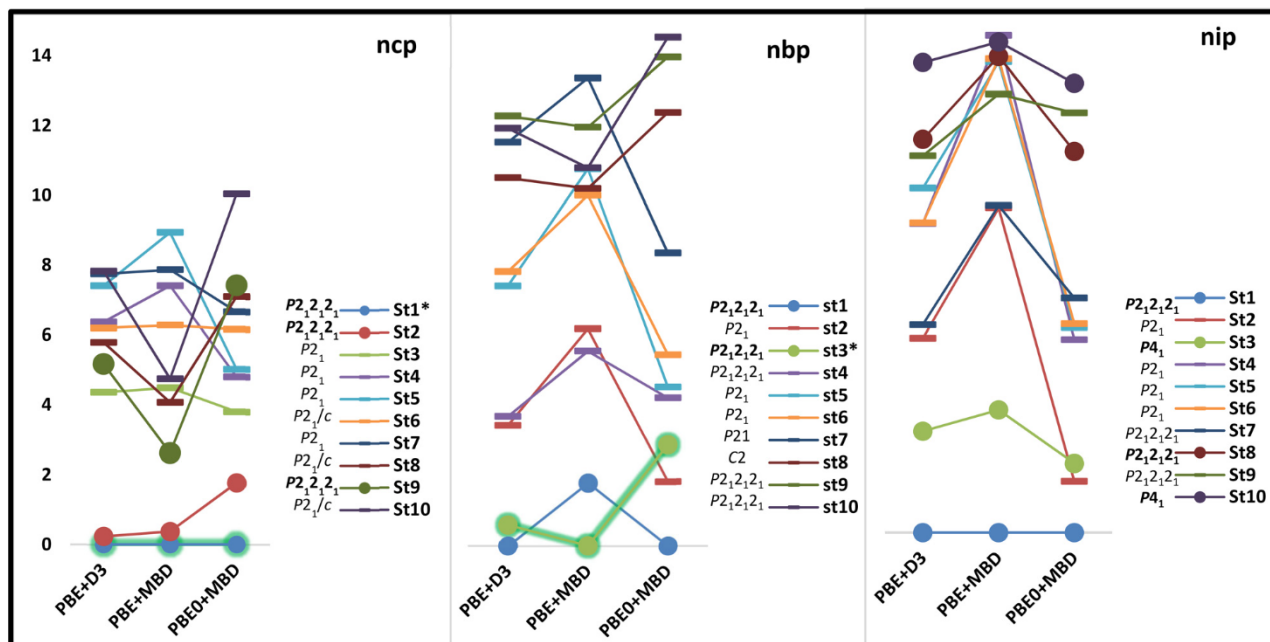


Fig. 5 The relative energies of the ten lowest-energy predicted structures of **ncp**, **nbp**, and **nip** were calculated with PBE + D3, PBE + MBD, and PBE0 + MBD.

refcodes CSUCIM for the dimer synthon and YOPJUI for the catemer synthon are shown in Fig. 6b. On the basis of cluster analysis results, we can expect that the N-X $\cdots$ O halogen bonding catemer synthon, as a repetitive synthon of the halide-imide skeleton, may be further observed in polymorphic structures expanded by hydrogen bonding of C-H $\cdots$ O/X interactions.

In addition, the N-X $\cdots$ O=C intermolecular interaction was studied in detail using the CSD structure as fragment II. The interaction parameters between the halogen atoms and O=C were defined as the X $\cdots$ O bond length, N-X $\cdots$ O angle θ_1 , and X $\cdots$ O=C angle θ_2 (Fig. 2b). For N-Cl $\cdots$ O=C, in 84.6% of the hits, the angles θ_1 and θ_2 are in the range 154–178° and 128–152°, respectively, the orientation of the chlorine atoms is related to the halogen interaction (the sigma hole of the halogen atom is involved), and the oxygen of the carbonyl group acts as a nucleophile. In 11.6% of the hits, θ_1 and θ_2 are approximately equal (with a difference of

1.5 to 3°). One hit has $\theta_1 = 161^\circ$ and $\theta_2 = 171^\circ$. All hit structures have a catemer synthon, except for two which have a dimer synthon. In the case of N-Br $\cdots$ O=C, the angles θ_1 and θ_2 are in the range 168–179° and 135–149°, respectively, and this orientation is related to the halogen interaction. All the hit structures have a catemer synthon. For N-I $\cdots$ O=C, seven hits were obtained, six of which have the angles θ_1 and θ_2 in the range 171–176° and 122–147°, respectively. In one case, $\theta_1 = 151^\circ$ and $\theta_2 = 147^\circ$, and all the hit structures have a catemer synthon.

With this knowledge, we investigated the structures obtained using CSP (CSP-structures hereafter for short) which are within 15 kJ mol $^{-1}$ of the most stable structures of the N-halide phthalimides. Interestingly, for **ncp**, among the 35 CSP-structures, 82.9% showed a halogen bonding synthon, with the angles θ_1 and θ_2 in the range 152–168° and 104–149°, respectively. Meanwhile, only 5.7% of these structures have a chalcogen bond, with the angles θ_1 and θ_2 in the

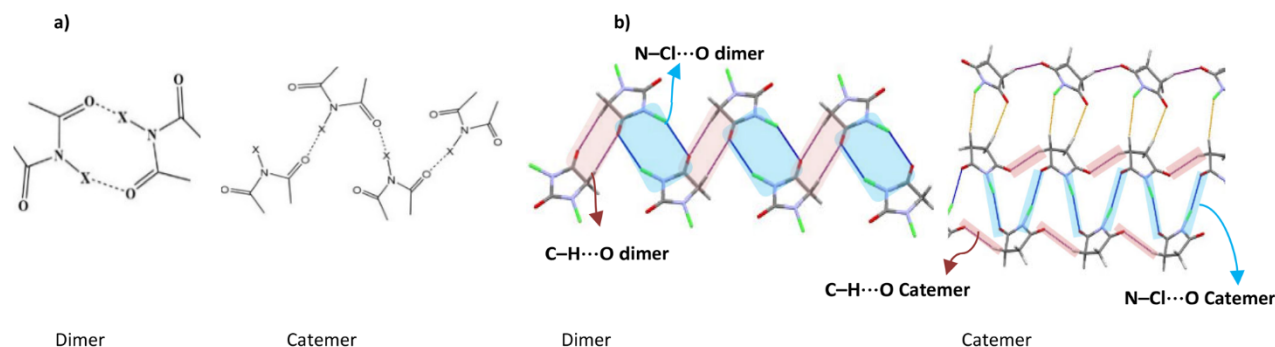


Fig. 6 (a) N-X $\cdots$ O dimer and catemer synthons, and (b) their respective CSD refcodes of CSUCIM (left) and YOPJUI (right).

Table 1 Presence of structures with a N-X...O=C contact in the CSD and CSP sets and their geometric parameters of interactions

X	CSD-structures		Angle range ^b		CSP-structures		Angle range ^b	
	Percentage	Interaction type	θ_1	θ_2	Percentage	Interaction type	θ_1	θ_2
Cl	84.6	Halogen bond	154–178	128–152	82.9	Halogen bond	152–168	104–149
	11.6 ^a	$\theta_1 \cong \theta_2$	142–161	140–162				
	3.8	θ_1 and $\theta_2 > 160^\circ$	161	171				
Br	100	Halogen bond	168–179	138–149	100	Halogen bond	153–171	102–147
I	87.5	Halogen bond	171–176	122–145	94.7	Halogen bond	153–176	116–153
	12.5	Halogen bond	151	147				

^a There are three hits where θ_1 and θ_2 are equal to (160.9°, 162.45°), (158.88°, 155.98°), and (141.72°, 139.93°). ^b The angles are reported as integers.

range 70–71° and 170–172°, respectively. In the case of **nbp**, there are 20 CSP-structures, all of them having halogen bond interactions, with the angles θ_1 and θ_2 in the range 153–171° and 102–147°, respectively. Similarly, for **nip**, there are 19 CSP-structures, and 94.7% of them are involved in halogen bond interactions, with θ_1 and θ_2 angles ranging from 153 to 176° and from 116 to 153°, respectively.

We compared the interactional behaviours of the CSD-structures and CSP-structures by means of the presence of the halogen bonding synthon as a criterion (Table 1). The

experimental and theoretical structures are in close agreement with each other. In both sets, 85–100% of the hits are related to the halogen catemer synthon. All CSP-structures of **nip** have a halogen interaction, and low-energy structures have space groups $P2_1$, $P2_12_12_1$, and $P4_1$. Interestingly, CSD-structures containing a N-I...O=C contact with refcodes of AKIXUP, PAKBUB, ZZZVCQ01, and PAKBEL are present in the set of nearly similar space groups of $P\bar{1}$, $P2_12_12_1$, $P4_1$, and $P4_12_12_1$, respectively (Fig. 7a). One can find that the orientations of molecules that form a plausible

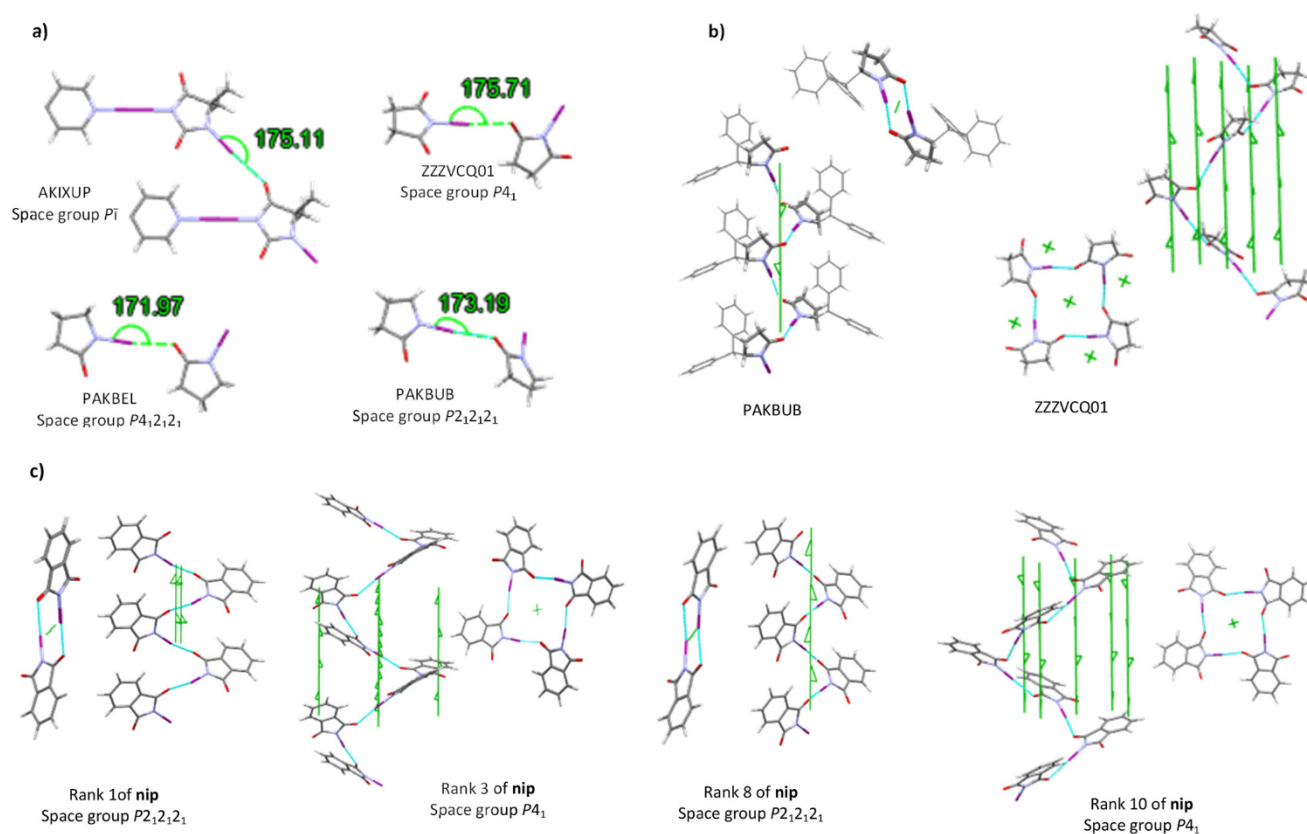


Fig. 7 (a) N-I...O=C halogen bond in CSD structures with fragment I, (b) the arrangement of molecules in CSD refcodes ZZZVCQ01 and PAKBUB around the screw axes 4_1 and 2_1 to form the halogen bond, and (c) the halogen bonding synthons and the arrangement of molecules in the lowest-energy predicted structures of **nip** of ranks 1, 3, 8, and 10.

Table 2 Lattice parameters of the predicted and experimental structures for **ncp**, **nbp**, and **nip**

CSD refcode/structure rank	Space group	<i>a</i> , Å	<i>b</i> , Å	<i>c</i> , Å	α , °	β , °	γ , °	<i>V</i> , Å ³	<i>d</i> , g cm ⁻³	RMSD ₁₅
Exp-WEZVIG	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	5.728(1)	6.275(1)	20.725(2)	90	90	90	744.923	1.620	—
Rank 1-ncp	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	5.661	6.138	20.830	90	90	90	723.825	1.660	0.19
Exp-SDPD	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	5.217(27)	6.661(24)	22.633(9)	90	90	90	786.611	1.900	—
Exp-VELWIV	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	5.235(3)	6.661(4)	22.648(13)	90	90	90	789.738	1.900	—
Rank 3-nbp	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	5.133	6.483	23.079	90	90	90	768.096	1.946	0.29
Rank 1-nip	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	23.704	6.604	5.129	90	90	90	802.841	2.262	—
Rank 3-nip	<i>P</i> ₄ ₁	8.252	8.252	12.826	90	90	90	873.315	2.079	—
Rank 8-nip	<i>P</i> ₂ ₁ ₂ ₁ ₂ ₁	7.389	8.379	13.363	90	90	90	827.346	2.195	—
Rank 10-nip	<i>P</i> ₄ ₁	7.753	7.753	13.664	90	90	90	821.239	2.211	—

directional halogen bond are related to the screw axes 4_1 and 2_1 as repetitive symmetry elements (Fig. 7b).

The rankings of the ten low-energy structures for each compound, based on various methods, are listed in Table S5.† In addition, the previously reported experimental crystal structure of **ncp**⁶² and **nbp**⁴² compounds and the structure of **nbp** determined in this work using the SDPD method are shown in Fig. S7.† Comparison of the CSP-structures with the experimental ones based on the RMSD₁₅ criteria revealed that the predicted structures match the experimental ones are in ranks 1 for **ncp** and 3 for **nbp** (Tables S5† and 2). The low-energy structure of **nbp** in rank 1 as a thermodynamically stable form is also close to the experimental structure based on the lattice energy, space group, density, and synthon

arrangement. The similarity index revealed that only the structure of rank 3 matches the experimental one. Interestingly, comparison of the bond angles in the halogen bonding synthons of the experimental structures with those of the ten lowest-energy of **nbp** shows that the rank 3 structure has the closest values to the experimental structure. This indicates both the kinetic and thermodynamic preference for synthons, so that the most abundant synthon is seen in the most stable (rank 1) structure of **ncp**, while the predicted metastable (rank 3) structure in **nbp** includes this synthon.

For **nip**, the geometry of the synthon is related to the structures of ranks 1, 3, 8, and 10 which are within the range of CSD-structures. It can be said more precisely that the

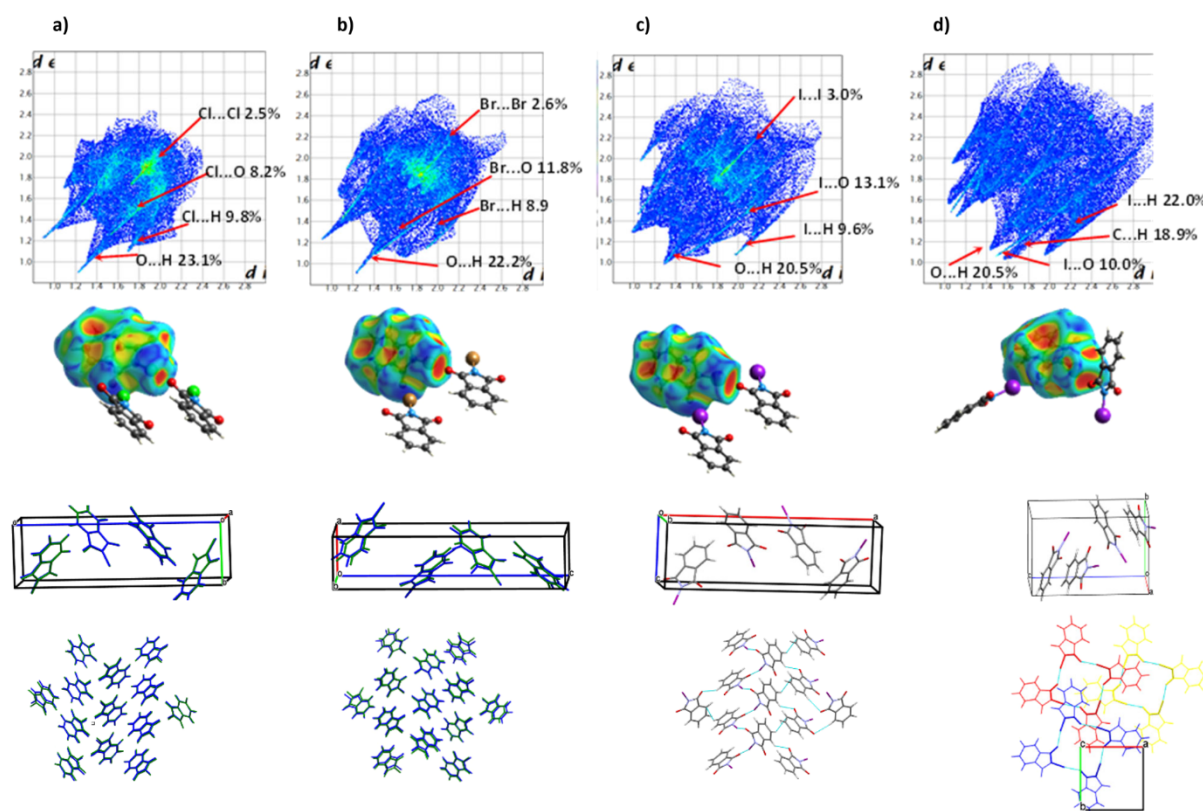


Fig. 8 2D Hirshfeld fingerprint plots, d_{norm} mapping of Hirshfeld surfaces, and crystal packing for the structures of (a) **ncp** of rank 1, (b) **nbp** of rank 3, and (c) **nip** of rank 1 that are isostructural (orthorhombic crystal system with the *P*₂₁₂₁₂₁ space group). (d) Suggested structure for **nip** of rank 3.

predicted crystal structure of **nip** closest to the CSD is of rank 3, with the $P4_1$ space group with $\theta_1 = 173^\circ$ and $\theta_2 = 135^\circ$. Interestingly, the structure of rank 1 is isostructural with **nep** and **nbp**. This structure has $\theta_1 = 170^\circ$, $\theta_2 = 132^\circ$, and an orthorhombic crystal system with the $P2_12_12_1$ space group. The structure of rank 10 is similar to that of rank 3 with $\text{RMSD}_{15} = 0.5$. The structure of rank 8 also exhibits the $P2_12_12_1$ space group, $\theta_1 = 172^\circ$, and $\theta_2 = 136^\circ$ (see Table 2). In Fig. 7c, the halogen bonding synthons and the molecular arrangement in the lowest-energy predicted structures of **nip** are of ranks 1, 3, 8, and 10 which are related to the screw axes 4_1 and 2_1 as repetitive symmetry elements in comparison with CSD-structures (refcodes ZZZVCQ01 and PAKBUB).

Regarding the recurrence of the halogen bonding synthon in the structures in this study as a robust unit in the crystal packing, we used this synthon as a criterion in selecting (or rejecting) the structures. On the other hand, the role of supramolecular synthons as structural units was previously mentioned as a kinetic factor in the formation of polymorphic structures (kinetic polymorphs).²³ These polymorphs regularly have structures with higher energy than those in the global energy minimum. However, we can suggest that two selected structures of **nip** (ranks 1 and 3) could be thermodynamic and kinetic polymorphic forms of this compound closest to the global minimum with suitable geometrical parameters of the halogen bond. The results indicate that both thermodynamic and kinetic effects of the supramolecular synthon are likely important in determining the crystal structure. Indeed, the importance of these effects appears to increase with the presence of directional interactions such as halogen bonds, which can lead to synthesizable metastable polymorphs. The lattice parameters of the experimental and predicted structures for **nep** and **nbp** and the polymorphic structures suggested for **nip** of ranks 1, 3, 8, and 10 are presented in Table 2.

To obtain more insight into the similarities and differences of crystal packings in the obtained structures, the intermolecular interactions were investigated using Hirshfeld surface analysis. The d_{norm} mappings of the Hirshfeld surfaces for the predicted structures are shown in Fig. 8. In these mappings, the sigma holes of the halogen atom engaged in the $\text{N-X}\cdots\text{O}=\text{C}$ interactions are indicated by red circles. In the 3D surfaces, the contacts whose lengths are equal to the sum of the van der Waals radii of the involved atoms are shown in white, whereas those having shorter and longer lengths are shown in red and blue, respectively. The fingerprint plots also show the relative quantitative contribution of each type of contact to the intermolecular interactions: for the $\text{X}\cdots\text{O}$ contacts in these compounds, the contributions for the Cl, Br, and I atoms are 8.2%, 11.8%, and 13.1%, respectively. Interestingly, the ascending order of these contributions is consistent with the increasing halogen polarisation and thus the strength of the halogen bonding. The $\text{X}\cdots\text{H}$ and $\text{O}\cdots\text{H}$ contacts indicate the role of hydrogen bonds in these compounds. The contribution of oxygen-

halogen interactions is three to four times larger than that of halogen-halogen interactions. These Hirshfeld surfaces clearly show the similarities of the crystal packings of **nep**, **nbp**, and rank 1 **nip**, whereas **nip** of rank 3 has a completely different crystal structure.

Conclusions

In this study, the evolutionary algorithm USPEX was used to predict the crystal structures of *N*-halide phthalimide compounds. Sets of low-energy structures (within 15 kJ mol⁻¹ of the most stable structure) were studied in detail. On the basis of the synthon approach, the most likely structures were selected from the low-energy predicted structures. In fact, we used the halogen bonding synthon as a known repetitive interaction unit for selecting (or rejecting) the target structures. The selection of the best candidate was carried out based on the best orientation of the $\text{O}=\text{C}$ moiety towards the sigma hole of the halogen atom (θ_1 and θ_2 angles), which led to a suitable chemical model and the subsequent formation of the target crystal structure. Out of the ten lowest-energy structures obtained, the structures of **nep** of rank 1 and **nbp** of rank 3 were found within 3 kJ mol⁻¹ of the most stable structures, showing close agreement with experiments. This methodology was used to predict the unknown structures of **nip** as two polymorphic structures of ranks 1 and 3. The results indicate that synthons, including both chemical and geometric recognition, can have an important role in determining the most likely structures among large sets of crystal structures produced in CSP calculations.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgements

Z. M. A. and A. S. respectfully acknowledge the Ferdowsi University of Mashhad Research Council for the financial support of project code number 3/47255. Z. M. A. and A. S. would like to thank the Iran National Science Foundation (INSF) for funding this work, through grant number 97012508. The authors thank the Cambridge Crystallographic Data Centre for the access to the CSD Enterprise. The work of A. R. O. was supported by the Russian Science Foundation (grant 19-72-30043).

Notes and references

- O. Egorova, R. Hafizi, D. C. Woods and G. M. Day, *J. Phys. Chem. A*, 2020, **124**, 8065–8078.
- M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber and O. Grassmann, *Nat. Commun.*, 2015, **6**, 7793–7799.
- S. L. Price, *Chem. Soc. Rev.*, 2014, **43**, 2098–2111.
- A. R. Oganov, C. J. Pickard, Q. Zhu and R. J. Needs, *Nat. Rev. Mater.*, 2019, **4**, 331.

- 5 A. Dey, N. N. Pati and G. R. Desiraju, *CrystEngComm*, 2006, **8**, 751–755.
- 6 X. Li, X. Ou, B. Wang, H. Rong, B. Wang, C. Chang, B. Shi, L. Yu and M. Lu, *Commun. Chem.*, 2020, **3**, 152.
- 7 W. D. S. Motherwell, *Mol. Cryst. Liq. Cryst.*, 2001, **356**, 559–567.
- 8 A. G. Shtukenberg, Q. Zhu, D. J. Carter, L. Vogt, J. Hoja, E. Schneider, H. Song, B. Pokroy, I. Polishchuk, A. Tkachenko, A. R. Oganov, A. L. Rohl, M. E. Tuckerman and B. Kahr, *Chem. Sci.*, 2017, **8**, 4926–4940.
- 9 X. F. Zhou, A. R. Oganov, G. R. Qian and Q. Zhu, *Phys. Rev. Lett.*, 2012, **109**, 245503.
- 10 W. I. F. David and K. Shankland, *Acta Crystallogr., Sect. A: Found. Crystallogr.*, 2008, **64**, 52–64.
- 11 K. D. M. Harris and Y. Ch. Eugene, *Chem. Soc. Rev.*, 2004, **33**, 526–538.
- 12 I. Bier, D. O'Conno, Y. T. Hsieh, W. Wen, A. M. Hiszpanski, T. Y. Han and N. Marom, *CrystEngComm*, 2021, **23**, 6023–6038.
- 13 A. A. Aina, A. J. Misquitta and S. L. Price, *J. Chem. Phys.*, 2021, **154**, 094123.
- 14 T. Pawlak, I. Sudgen, G. Bujacz, D. Iuga, S. P. Brown and M. J. Potrzebowski, *Cryst. Growth Des.*, 2021, **21**, 3328–3343.
- 15 C. Greenwell and G. Beran, *Cryst. Growth Des.*, 2020, **20**, 4875–4881.
- 16 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.
- 17 S. Yang and G. M. Day, *J. Chem. Theory Comput.*, 2021, **17**, 1988–1999.
- 18 J. Kendrick, F. J. J. Leusen, M. A. Neumann and J. van de Streek, *Chem. – Eur. J.*, 2011, **17**, 10736–10744.
- 19 T. S. Thakur and G. R. Desiraju, *Cryst. Growth Des.*, 2008, **8**, 4031–4044.
- 20 J. R. P. Sarma and G. R. Desiraju, *Cryst. Growth Des.*, 2002, **2**, 93–100.
- 21 A. Collins, C. C. Wilson and C. J. Gilmore, *CrystEngComm*, 2010, **12**, 810–816.
- 22 B. Sandhu, A. McLean, A. S. Sinha, J. Desper and C. B. Aakeröy, *Chemistry*, 2021, **3**, 612–629.
- 23 A. Dey, M. T. Kirchner, V. R. Vangala, G. R. Desiraju, R. Mondal and J. A. K. Howard, *J. Am. Chem. Soc.*, 2005, **127**, 10545–10559.
- 24 G. R. Desiraju, *Angew. Chem., Int. Ed.*, 2007, **46**, 8342–8356.
- 25 S. Chakraborty and G. R. Desiraju, *CrystEngComm*, 2018, **20**, 2793–2805.
- 26 S. Saha and G. R. Desiraju, *J. Am. Chem. Soc.*, 2018, **140**, 6361–6373.
- 27 S. S. Singhab and T. S. Thakur, *CrystEngComm*, 2014, **16**, 4215–4230.
- 28 D. W. M. Hofmann, L. N. Kuleshova and M. Y. Antipin, *Cryst. Growth Des.*, 2004, **4**, 1395–1402.
- 29 G. Barr, W. Dong, C. J. Gilmore, A. Parkin and C. C. Wilson, *J. Appl. Crystallogr.*, 2005, **38**, 833–841.
- 30 A. Parkin, G. Barr, W. Dong, C. J. Gilmore and C. C. Wilson, *CrystEngComm*, 2006, **8**, 257–264.
- 31 A. Collins, A. Parkin, G. Barr, W. Dong, C. J. Gilmore and C. C. Wilson, *CrystEngComm*, 2007, **9**, 245–253.
- 32 Q. Zhu, A. R. Oganov, C. W. Glass and H. T. Stokes, *Acta Crystallogr., Sect. B: Struct. Sci.*, 2012, **68**, 215–226.
- 33 A. O. Lyakhov, A. R. Oganov, H. T. Stokes and Q. Zhu, *Comput. Phys. Commun.*, 2013, **184**, 1172–1182.
- 34 A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
- 35 A. R. Oganov, A. O. Lyakhov and M. Valle, *Acc. Chem. Res.*, 2011, **44**, 227.
- 36 E. Y. Tupikina, G. S. Denisov, A. S. Antonova and P. M. Tolstoy, *Phys. Chem. Chem. Phys.*, 2020, **22**, 1994–2000.
- 37 M. Castro, I. Nicolas-Vazquez, J. I. Zavala, F. Sanchez-Viesca and M. Berros, *J. Chem. Theory Comput.*, 2007, **3**, 681–688.
- 38 A. Emily, J. Cann, A. Decken and S. Eisler, *CrystEngComm*, 2017, **7**, 1024–1027.
- 39 P. Politzer, P. Lane, M. C. Concha, Y. Ma and J. S. Murray, *J. Mol. Model.*, 2007, **13**, 305–311.
- 40 P. Politzer, J. S. Murray and T. Clark, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7748–7757.
- 41 P. R. Varadwaj, A. Varadwaj and H. M. Marques, *Inorganics*, 2019, **7**, 40.
- 42 M. Eraković, V. Nemec, T. Lež, I. Porupski, V. Stilić and D. Cinić, *Cryst. Growth Des.*, 2018, **18**, 1182–1190.
- 43 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, O. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian 09 Revision D.01*, Gaussian Inc, Wallingford CT, 2013.
- 44 J. D. Gale and A. L. Rohl, *Mol. Simul.*, 2003, **29**, 291–341.
- 45 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
- 46 A. K. Rappe and W. A. Goddard III, *J. Phys. Chem.*, 1991, **95**, 3358–3363.
- 47 J. A. Chisholm and S. Motherwell, *J. Appl. Crystallogr.*, 2005, **38**, 228–231.
- 48 S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- 49 A. Tkatchenko, R. A. DiStasio Jr, R. Car and M. Scheffler, *Phys. Rev. Lett.*, 2012, **108**, 236402.
- 50 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1996, **54**, 11169–11186.

- 51 G. Kresse and J. Hafner, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1993, **47**, 558–561.
- 52 G. Kresse and J. Furthmüller, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **49**, 14251.
- 53 P. E. Blöchl, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1994, **50**, 17953.
- 54 G. Kresse and D. Joubert, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 1999, **59**, 1758–1775.
- 55 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.
- 56 G. Barr, W. Dong, C. J. Gilmore, A. Parkin and C. C. Wilson, *J. Appl. Crystallogr.*, 2005, **38**, 833–841.
- 57 A. Boulton and D. J. Louer, *J. Appl. Crystallogr.*, 2004, **37**, 724.
- 58 W. I. F. David, K. Shankland, J. van der Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *J. Appl. Crystallogr.*, 2006, **39**, 910.
- 59 A. Larson and R. Von Dreele, *General Structure Analysis System, LANL Report LAUR 86*, Los Alamos National Laboratory, Los Alamos, NM, 2000.
- 60 B. H. Toby, *J. Appl. Crystallogr.*, 2001, **34**, 210.
- 61 P. Thompson, D. E. Cox and J. B. Hastings, *J. Appl. Crystallogr.*, 1987, **20**, 79.
- 62 M. Ghassemzadeh, K. Harms, K. Dehnicke and J. Magull, *Z. Naturforsch., B: J. Chem. Sci.*, 1994, **49**, 506–512.