Research Article

Searching new cocrystal structures of CL-20 and HMX via evolutionary algorithm and machine learning potential

Zhong-Hao Ye[1], Feng Guo[1] (iD), **Chuan-Guo Chai[2], Yu-Shi Wen[2], Zheng-Rong Zhang[3], Heng-Shuai Li[1], Shou-Xin Cui[1], Gui-Qing Zhang[1], Xiao-Chun Wang[1]**

[1]School of Physical Science and Information Technology, Liaocheng University, Liaocheng 252059, Shandong, China.
[2]Institute of Chemical Materials, China Academy of Engineering Physics (CAEP), Mianyang 621900, Sichuan, China.
[3]Department of Physics, Lanzhou City University, Lanzhou 730070, Gansu, China.

*****Correspondence to:** Dr. Feng Guo, School of Physical Science and Information Technology, Liaocheng University, No.1 Hunan Road, Liaocheng 252059, Shandong, China. E-mail: gfeng.alan@foxmail.com

Abstract

In this work, we report the discovery of energy cocrystals using an efficient iterative workflow combining an evolutionary algorithm and a machine learning potential (MLP). The compound 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20) has attracted significant attention owing to its higher energy density than traditional energetic materials. However, the higher sensitivity has limited its applications. An important way to reduce its sensitivity involves cocrystal engineering with traditional explosives. Many cocrystal structures are expected to be composed of these two components. We developed an efficient iterative workflow to explore the phase space of CL-20 and 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane (HMX) cocrystals using an evolutionary algorithm and an MLP. The algorithm was based on the Universal Structure Predictor: Evolutionary Xtallography (USPEX) software, and the MLP was the reactive force field with neural networks (*ReaxFF-nn*) model. A set of high-density cocrystal structures was produced through this workflow; these structures were further checked via first-principles geometry optimizations. After careful screening, we identified several high-density cocrystal structures with densities of up to 1.937 g/cm^3 and HMX:CL-20 ratios of 1:1 and 1:2. The influence of hydrogen bonds on the formation of high-density cocrystals was also discussed, and a roughly linear relationship was found between energy and density.

Keywords: Crystal search, machine learning, *ReaxFF-nn*, evolutionary algorithm

INTRODUCTION

Crystal structure prediction (CSP) techniques enable the design of new crystal structures with a given chemical composition. However, the high computational costs of density functional theory (DFT) calculations limit their applications in molecular systems, whose unit cells usually contain dozens or even hundreds of atoms. Recently developed machine learning potentials (MLPs), such as the Gaussian approximation potential (GAP)[1–6], moment tensor potential (MTP)[7–9], and DeepMD-kit package[10–13], can perform DFT-level calculations at the computational cost of classical force fields. CSP methods, such as Universal Structure Predictor: Evolutionary Xtallography (USPEX)[14,15] and Crystal structure AnaLYsis by Particle Swarm Optimization (CALPSO)[16], combining MLPs[17–23] will undoubtedly become the main approaches in the search for new materials. The work by Podryabinkin[24] on CSP combining an MLP and an evolutionary algorithm showed that an accuracy of 11 meV/atom could be achieved in the known structure predictions. It is clear that MLPs[25] can accelerate CSP by replacing most DFT calculations. Although attempts[26,27] have been made to use the classical force field as a local optimizer for structure minimization, the parameter set used by the force field cannot be dynamically adjusted according to the search results and, therefore, cannot guarantee the predicted structure to be the best one. In contrast, machine learning-based methods can provide predictions with quality close to that of first-principles calculations at the computational costs of classical techniques and thus can deal with large-scale[28] systems. In addition, the parameter set can be dynamically updated through a *search–check–train–search* workflow developed by our group. A similar workflow is used in the active learning procedure used to train deep-learning potential models, such as the Deep Potential GENerator (DP-GEN)[29] and Fast Learning of Atomistic Rare Events (FLARE)[30] software, which contains three *Exploration–Labeling–Training* iterative steps. In the *Exploration* step, a sample of trajectories is generated using methods such as molecular dynamics (MD) simulations and randomly sampling the crystal phase space. The *search* step in our workflow is similar to the *Exploration* of the crystal space group by the USPEX software. In the *Labeling* step, and correspondingly in our *check* step, the *ab initio* energies are calculated for the configurations generated in the previous step, and the configurations with a large variance are added to the training dataset. In the last step, the model is trained, and a new parameter set is generated. In this work, we combined MLP and evolutionary algorithm to identify new cocrystal structures of 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20)[31] and 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane (HMX)[32], aiming to find an optimal balance between their energy and safety. CL-20 is known for its high detonation power; however, significantly reducing its sensitivity is a great challenge. Effectively addressing this problem involves combining CL-20 with other low-sensitivity materials, producing more attractive explosive cocrystal materials[33–37]. A highly promising CL-20:HMX (2:1) cocrystal was synthesized by Bolton *et al.*[38]; the cocrystal exhibits superior energetics to HMX while maintaining comparable impact sensitivity. Many cocrystal structures composed of CL-20 and HMX are expected to exist because of the large number of crystal structures the single components can form.

As thousands of crystal structures will need to be optimized in one search, using the DFT method as the local optimizer will require years to complete. The machine learning-based approach will speed up the search and provide similar accuracy, as the parameter set is dynamically updated after the search is corrected by DFT calculations. The MLP will "*learn*" the interactions between molecules and give more accurate predictions. A workflow combining the USPEX evolutionary algorithm and the *ReaxFF-nn* MLP was developed to explore the configuration space of CL-20 and HMX. Using this approach, we report the discovery of high-density cocrystal structures of HMX and CL-20 with 1:1 and 1:2 ratios. Furthermore, there are discussions on whether C–H. . . O hydrogen bond interactions[39] exist in crystal structures; hence, the role of C–H. . . O hydrogen bond interactions in the cocrystal structures is discussed by analyzing hydrogen bond energies.

METHODS

ReaxFF-nn machine learning potential

The *ReaxFF* reactive force field, originally introduced by van Duin *et al.*[40], has been extensively used[41,42] in reactive MD simulations[43–51]. We developed[52] an MLP model labeled *ReaxFF-nn* based on *ReaxFF* with neural networks for bond orders in previous work[53]. This significantly improved various calculations, such as equations of states, reaction energies, and potential energy surfaces. The total energies for *ReaxFF* and *ReaxFF-nn* are respectively given in:

$$E_{system}^{ReaxFF} = E_{bond} + E_{lp} + E_{over} + E_{under} + E_{val} + E_{pen} + E_{coa} +$$
$$E_{tors} + E_{conj} + E_{Hbond} + E_{vdWaals} + E_{Coulomb}, \tag{1}$$

$$E_{system}^{ReaxFF-nn} = E_{bond}^{nn} + E_{val} + E_{tors} + E_{Hbond} + E_{vdWaals} + E_{Coulomb}. \tag{2}$$

where E_{bond}, E_{lp}, E_{over}, E_{under}, E_{val}, E_{pen}, E_{coa}, E_{tors}, E_{conj}, E_{Hbond}, $E_{vdWaals}$ and $E_{Coulomb}$ correspond to the bond, lone pair, over coordinate, under coordinate, valence angle, penalty, three-body conjugation, torsion rotation barrier, four-body conjugation, hydrogen bond, van der Waals and Coulomb energy terms, respectively[40].

In the current form, as given in Equation (2), neural networks are employed to compute short-range interactions, while long-range interactions, such as Coulomb and van der Waals terms, still use the same formula as *ReaxFF*. Because most MLPs use prespecified static point charges, they focus on short-range interactions[54–56] and have difficulties dealing with long-range Coulomb interactions, which remain a major challenge for these potentials at present. An exception is represented by *ReaxFF*, and, similarly, *ReaxFF-nn*, which use charges dynamically calculated by the electronegativity equilibrium[57] method, and further related approaches are expected to be developed in the future[58].

The corrected bond order *BO ReaxFF* and *ReaxFF-nn* is respectively calculated by:

$$BO^{ReaxFF} = BO' \cdot f_1(\Delta_i', \Delta_j') \cdot f_4(\Delta_i', BO_{ij}') \cdot f_5(\Delta_j', BO_{ij}'), \tag{3}$$

where the f_1, f_4, and f_5 terms are given by the *ReaxFF* formula in Ref.[40],

$$BO^{ReaxFF-nn} = BO' \cdot f_{nn}(\Delta_i', BO_{ij}', \Delta_j'), \tag{4}$$

f_{nn} used by a hidden-layer neural network is calculated by:

$$f_{nn}(x) = \sigma(w_o \cdot \sigma(w_h \cdot \sigma(w_i \cdot x + b_i) + b_h) + b_o). \tag{5}$$

where $x = (\Delta_i', BO_{ij}', \Delta_j')$ represents the input vector, Δ_i' is the sum of the BO_{ij}' terms of $atom_i$, and BO_{ij}' is the uncorrected bond order, as defined in Ref.[40]; moreover, σ is the logistic activation function $\sigma(x) = (1 + \exp(-x))^{-1}$, while w and b are the weights and biases of the neural networks, respectively.

The bond energies employed by *ReaxFF* and *ReaxFF-nn* are respectively given in:

$$E_{bond_{ij}}^{ReaxFF} = D_e^{\sigma} \cdot BO_{ij}^{\sigma} \cdot \exp[p_{be_1}(1 - BO_{ij}^{\sigma})^{p_{be_2}}] - D_e^{\pi} \cdot BO_{ij}^{\pi} - D_e^{\pi\pi} \cdot BO_{ij}^{\pi\pi},$$

(6)

$$E_{bond_{ij}}^{ReaxFF-nn} = D_e^{\sigma} \cdot f_{nn}^e(BO_{ij}^{\sigma}, BO_{ij}^{\pi}, BO_{ij}^{\pi\pi}).$$

(7)

where D_e^{σ}, D_e^{π}, $D_e^{\pi\pi}$, p_{be_1} and p_{be_2} are adjusting parameters, and BO_{ij}^{σ}, BO_{ij}^{π}, and $BO_{ij}^{\pi\pi}$ are bond order components.

In this case, f_{nn}^e represents a hidden-layer neural network that calculates the bond energy. The input vector for this network is the uncorrected bond order $BO_{ij} = (BO_{ij}^{\sigma}, BO_{ij}^{\pi}, BO_{ij}^{\pi\pi})$. The network output, multiplied by a scalable parameter D_e^{σ}, gives the bond energy of the pair of atoms $atom_i$ and $atom_j$.

Training and testing the potential model
The quality of the dataset used for the MD interaction potential plays a crucial role in determining its accuracy. To ensure the complete sampling of the phase space of molecular structures such as CL-20, 2,4,6-trinitrotoluene (TNT), 1,1-diamino-2,2-dinitroethylene (FOX-7), cyclotrimethylene trinitramine (RDX) and HMX, we manually collected a large amount of data using methods including *ab initio* MD simulations, stretching of specific chemical bonds, scanning covalent bond angles, and equations of state of a crystal structure. After DFT calculations, these data were stored in an atomic simulation environment (ASE)[59] *Trajectory* object, which can be interpreted as a time series containing many ASE *Atoms* objects moving in the configurational phase space. This data format is easy to prepare because the ASE package includes many DFT calculators, such as Vienna *ab initio* simulation package (VASP), Spanish Initiative for Electronic Simulations with Thousands of Atoms (SIESTA), and Quantum Espresso (QE). After performing calculations through these calculators, the *Trajectory* file can be generated automatically. One can also perform single-point energy calculations and gather the trajectory file after the calculations. The training function accepts a *dataset* variable, which is a Python *dict* object that contains all trajectory file names; moreover, the *I-ReaxFF* package can generate input information for the training function, such as collecting all bonds and all valence angles of the same type into variables that store this information. More details on the data preparation can be found in the documentation and examples of our *I-ReaxFF* package[53,60].

Figure 1A-C compares the lattice energies calculated by DFT and *ReaxFF-nn* as a function of the density for the β-HMX, ε-CL-20, and 1:1 HMX:CL-20 cocrystal. Scaling the box size in three directions changed the density from 1.4 to 2.4 g/cm^3. Figure 1D and E shows the energies *vs.* the N–NO$_2$ bond distance calculated by DFT and *ReaxFF-nn*. Figure 1F compares the lattice energies of a batch of 100 cocrystal structures of 1:1 HMX:CL-20 generated by the USPEX software. The results demonstrate a good agreement between our *ReaxFF-nn* potential model and the DFT calculations. Although active learning algorithms[29,61,62] to collect data automatically are available, the application to molecular systems requires further tests, and we also plan to develop an uncertainty-driven active learning algorithm based on the Z-matrix.

After data collection, the *ReaxFF-nn* model was trained using the *I-ReaxFF* software package, and the overall precision of this dataset was approximately 94.0%, as calculated by:

$$Accu = 1.0 - \frac{\sum |E_{DFT} - E_{ReaxFF-nn}|}{\sum (E_{DFT} - E_{ZPE})},$$
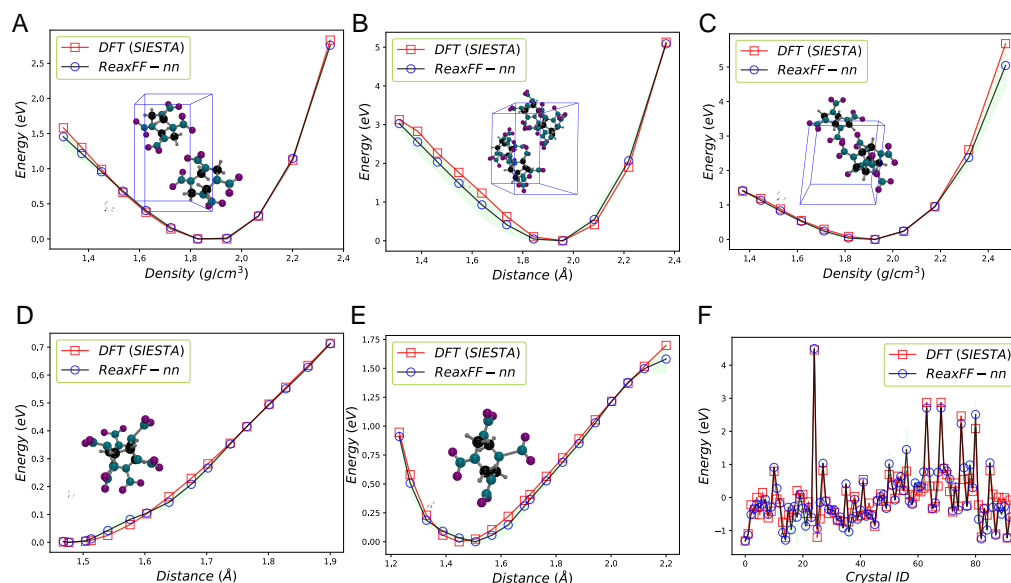
(8)

where E_{ZPE} represents the zero-point energy.

Figure 1. Lattice energy as a function of density of (A) β-HMX, (B) ε-CL-20, and (C) 1:1 HMX:CL-20 cocrystal; Energy of molecular (D) HMX and (E) CL-20 as a function of N–NO_2 bond distance; (F) Comparison of DFT- and *ReaxFF-nn*-calculated lattice energies of 1:1 cocrystal structures randomly generated by USPEX. HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; DFT: density functional theory; *ReaxFF-nn*: reactive force field with neural networks; USPEX: Universal Structure Predictor: Evolutionary Xtallography.

Table 1. The comparison of lattice constants of ε-CL-20 and β-HMX from experiment, DFT calculated values, and GULP geometry optimization with potential *ReaxFF-nn*

	a(Å)	b(Å)	c(Å)	β(°)
ε-CL-20, Exp. [27,65]	8.852	12.556	13.386	106.82
ε-CL-20, Exp. [66,67]	8.837	12.554	13.289	106.87
ε-CL-20, Exp. [68]	8.863	12.593	13.395	106.92
ε-CL-20, DFT-D [69]	8.871	12.557	13.477	106.62
ε-CL-20, *ReaxFF-nn*	8.989	12.401	13.428	105.90
β-HMX, Exp. [70]	6.526	11.037	7.364	102.67
β-HMX, Exp. [71]	6.530	11.030	7.350	102.69
β-HMX, Exp. [72]	6.540	11.050	7.371	102.83
β-HMX, DFT-D [69]	6.593	10.860	7.406	102.84
β-HMX, *ReaxFF-nn*	6.434	10.582	7.876	103.40

CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; DFT: density functional theory; GULP: General Utility Lattice Program.

To further evaluate the reliability of the *ReaxFF-nn* potential model, we optimized the crystal structures of β-HMX (chair conformation [63]) and ε-CL-20 (the most stable crystal form of CL-20 [27,38,64]). A comparison of the calculated lattice constants with the experimental data and the first-principles values is presented in Table 1. The analysis shows that the calculated values obtained using *ReaxFF-nn* agree well with the experimental and first-principles data.

RESULTS AND DISCUSSION

Molecular CSPs were carried out using the USPEX package [14,15,73,74]. This package is based on the evolutionary algorithm developed by Oganov, Glass, Lyakhov, and Zhu. Initially, a group of 100 structures was randomly generated from the appropriate crystal space group. In subsequent generations, the total population of 100 structures was maintained after operations including *heredity*, *randomization*, *permutation*, *rotation*,
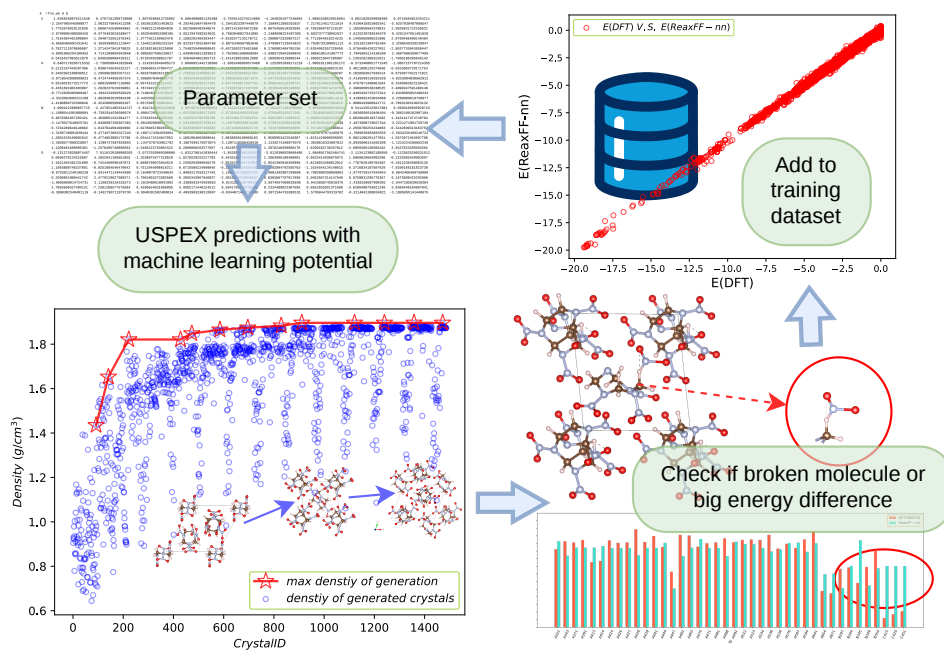
Figure 2. Schematic illustration of four-step iterative workflow combining USPEX and *ReaxFF-nn* for searching the appropriate cocrystal structures and training the MLP. USPEX: Universal Structure Predictor: Evolutionary Xtallography; *ReaxFF-nn*: reactive force field with neural networks; MLP: machine learning potential.

and *softmutation*. The optimal structures of the previous generation were kept as *keptbest* [15] structures. Each structure underwent two local relaxation steps using the General Utility Lattice Program (GULP) package [75] and the *ReaxFF-nn* potential. The fitness of each structure was determined based on the corresponding energy calculated by GULP. Following the principles of biological evolution, structures with lower fitness were then eliminated. Figure 2 illustrates the main search workflow comprising four stages. First, the initial parameter set is generated by training the model against the existing dataset. Second, the USPEX algorithm is combined with the *ReaxFF-nn* MLP (structural relaxation) to search for cocrystals for specific needs, generating a set of crystal structures. Third, these crystal structures are checked by comparing the corresponding *ReaxFF-nn*- and DFT-calculated energies. Finally, structures with a very large difference between the calculated energies are added to the training dataset. These steps are repeated until a small difference between the *ReaxFF-nn* and DFT results or new cocrystal structures meeting the selected requirements have been found.

Prediction of 1:2 HMX:CL-20 cocrystal configurations

After several searches, we obtained some cocrystal structures with lower enthalpy (potential energy) and higher density. However, to acquire densely packed cocrystals, we set a small pressure in the search process (p = 0.1 GPa). The obtained structures had to be optimized (relaxed) to zero pressure. The optimization results using the *ReaxFF-nn* and DFT (SIESTA [76]) method with the Perdew-Burke-Ernzerhof (PBE) functional and a double-ζ basis set at zero pressure are shown in Supplementary Table 1. Furthermore, an experimentally synthesized cocrystal structure was optimized by DFT (SIESTA) to serve as a reference, as shown in the first row of Supplementary Table 1.

Supplementary Table 1 reveals a good agreement between the cocrystal structures optimized by *ReaxFF-nn* and DFT (SIESTA), demonstrating the high performance of our MLP model. Combined with the data from the structures in Supplementary Table 1, the enthalpy and density of the cocrystal structures are shown as bar graphs in Figure 3. The enthalpy and density values calculated by SIESTA and *ReaxFF-nn* are shown in red
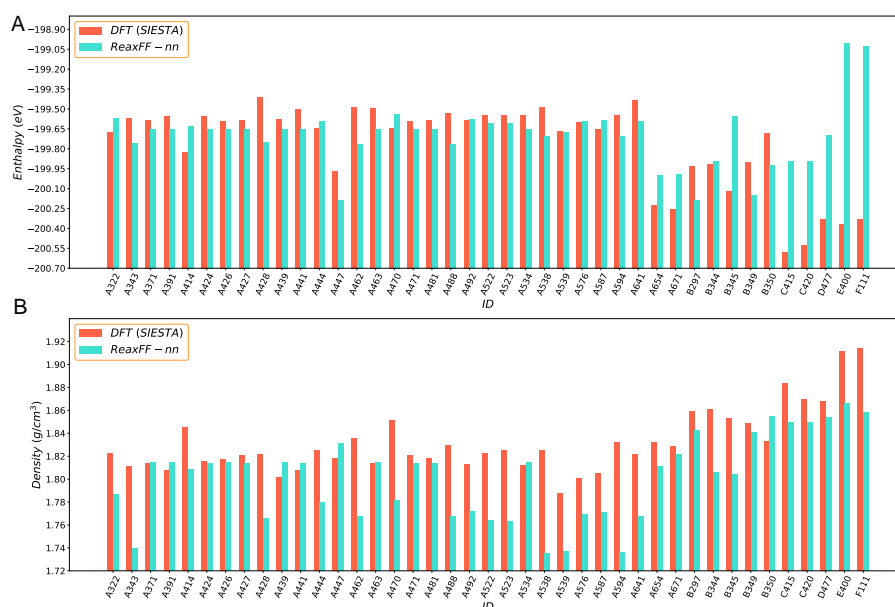
Figure 3. (A) Enthalpy and (B) density *vs.* crystal index from USPEX of 1:2 HMX:CL-20 optimized by DFT (SIESTA) and MLP (*ReaxFF-nn*). USPEX: Universal Structure Predictor: Evolutionary Xtallography; HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; DFT: density functional theory; SIESTA: Spanish Initiative for Electronic Simulations with Thousands of Atoms; MLP: machine learning potential; *ReaxFF-nn*: reactive force field with neural networks.

Table 2. The comparison of the enthalpy, density, and lattice constants from DFT (SIESTA), vdW-DF, and DFT-D3 (VASP) calculations of the top predicted 1:2 HMX:CL-20 cocrystal structures, respectively

ID	*Method*	Enthalpy (eV)	$\rho(g \cdot cm^{-3})$	$a(\text{Å})$	$b(\text{Å})$	$c(\text{Å})$	$\alpha(°)$	$\beta(°)$	$\gamma(°)$
	PBE	-200.580	1.884	11.026	12.052	8.251	84.28	90.71	71.44
C415	vdW-DF	-199.732	1.921	10.901	12.014	8.206	83.94	91.13	71.74
	PBE-D3	-648.675	1.902	11.095	12.146	8.120	83.39	91.06	70.63
	PBE	-200.532	1.870	10.866	12.128	8.311	85.03	89.12	72.60
C420	vdW-DF	-199.725	1.923	10.829	12.028	8.231	83.92	90.56	71.95
	PBE-D3	-648.699	1.871	11.090	12.198	8.198	83.66	90.12	70.89
	PBE	-200.336	1.868	10.858	12.184	8.323	84.55	88.82	71.93
D477	vdW-DF	-199.460	1.920	10.874	11.596	8.656	79.77	82.86	71.18
	PBE-D3	-648.573	1.905	10.951	11.759	8.615	79.19	82.90	70.00
	PBE	-200.373	1.912	10.771	12.231	8.406	100.04	83.61	71.59
E400	vdW-DF	-199.361	1.954	10.882	11.950	8.313	99.59	83.25	71.88
	PBE-D3	-648.935	1.934	10.891	12.210	8.258	99.28	82.96	71.03
	PBE	-200.329	1.914	11.234	12.136	8.268	98.71	82.17	68.92
F111	vdW-DF	-199.260	1.991	11.239	11.751	8.056	98.00	81.73	71.38
	PBE-D3	-648.976	1.937	11.244	12.145	8.099	98.14	82.10	69.64

DFT: Density functional theory; SIESTA: Spanish Initiative for Electronic Simulations with Thousands of Atoms; vdW-DF: van der Waals Density Functional; VASP: Vienna *ab initio* simulation package; HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; PBE: Perdew-Burke-Ernzerhof.

and turquoise, respectively.

In order to confirm the findings of the search workflow, we chose the five most significant cocrystal configurations with a 1:2 ratio and further optimized them using the van der Waals Density Functional (vdW-DF)[77] method and DFT-D3 method with the PBE functional and an energy cutoff of 600 eV. These DFT-D3 calculations were carried out using the VASP[78] software. The results are detailed in Table 2.

Based on the information on the DFT-D3-optimized cocrystal structures from Table 2, the five cocrystal struc-
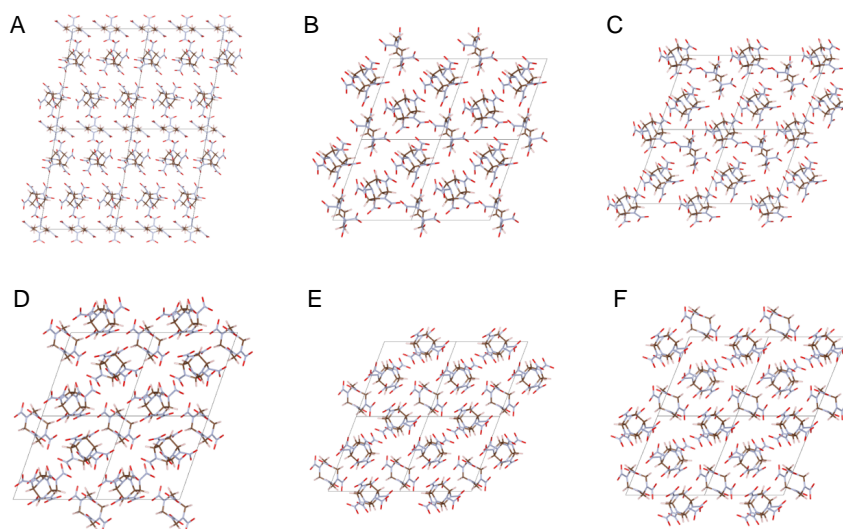
Figure 4. (A) Experimentally synthesized cocrystal structure[38], and (B–F) 1:2 HMX:CL-20 cocrystal structures after DFT-D3 structure optimization at zero pressure. The (B–F) structures correspond to IDs *C415*, *C420*, *D477*, *E400*, and *F111* in Table 2, respectively. HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; DFT: density functional theory.

tures with lower energy and higher density and the structure experimentally synthesized by Bolton *et al.*[38] are shown in Figure 4. The optimized cocrystal structures in this picture show the characteristic alternating arrangements of monolayer HMX and bilayer CL-20, consistent with the cocrystal arrangement of the experimentally synthesized structure.

Among the structures with the predicted 1:2 HMX:CL-20 ratio obtained from USPEX and MLP, the structure *F111* shown in Figure 4F had the highest density of 1.914 g/cm^3 after DFT optimization and 1.937 g/cm^3 after DFT-D3 optimization. Furthermore, the *F111* cocrystal also exhibited the lowest lattice energy.

Prediction of 1:1 HMX:CL-20 cocrystal configurations

A search of 1:1 HMX:CL-20 cocrystal structures was carried out using USPEX and *ReaxFF-nn* through the *search–check–train–search* iterative loop, as illustrated in Figure 2. A low pressure was applied to the candidate cocrystal cells; in the next step, the generated structures were optimized using the DFT (SIESTA) method with the PBE functional and a double-ζ basis set, as well as with the *ReaxFF-nn* potential to zero pressure. Supplementary Table 2 shows the structural information on the optimized cocrystals with lower enthalpy and higher density for the 1:1 ratio at zero pressure. The enthalpy and density of the cocrystal structures after optimization, obtained using *ReaxFF-nn* and DFT, are shown in Figure 5, which enables us to identify structures with lower enthalpy and higher density.

Similarly, to confirm the findings of the search workflow, we chose the six most significant cocrystal configurations with a 1:1 ratio and further optimized them using the vdW-DF method and DFT-D3 (VASP) method with the PBE functional and an energy cutoff of 600 eV. The results are detailed in Table 3. We selected the better cocrystal structures calculated by DFT-D3, as shown in Figure 6. The optimized structures in Figure 6 show characteristic cocrystal structures with alternating arrangements of HMX and CL-20 monolayers. Theoretically, this is consistent with the characteristic arrangement of cocrystal structures.

The results show the *D558* and *D559* structures had the highest density of 1.925 g/cm^3 after DFT (SIESTA) op-
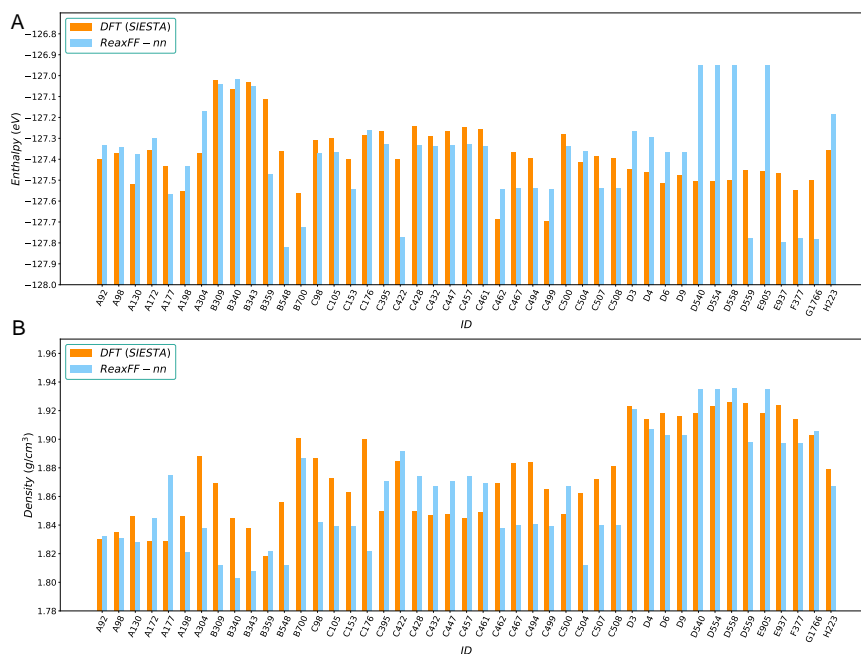
Figure 5. Comparison of DFT (SIESTA)- and *ReaxFF-nn*-calculated (A) enthalpy and (B) density values *vs.* crystal index of cocrystal structures with 1:1 HMX:CL-20 ratio generated by USPEX. DFT: Density functional theory; SIESTA: Spanish Initiative for Electronic Simulations with Thousands of Atoms; *ReaxFF-nn*: reactive force field with neural networks; USPEX: Universal Structure Predictor: Evolutionary Xtallography.

Table 3. Comparison of enthalpy, density, and lattice constants from DFT (SIESTA), vdW-DF, and DFT-D3 (VASP) calculations of the top predicted cocrystal structures with 1:1 ratio, respectively

ID	*Method*	Enthalpy (eV)	$\rho(g \cdot cm^{-3})$	$a(\text{Å})$	$b(\text{Å})$	$c(\text{Å})$	$\alpha(°)$	$\beta(°)$	$\gamma(°)$
	PBE	-127.447	1.923	8.144	8.491	9.343	98.31	90.83	82.73
D3	vdW-DF	-127.416	1.933	8.120	8.457	9.363	98.51	90.71	82.75
	PBE-D3	-411.687	1.937	8.141	8.455	9.318	98.49	91.15	82.81
	PBE	-127.499	1.925	8.205	8.666	9.055	98.10	90.51	83.40
D558	vdW-DF	-127.521	1.927	8.217	8.645	9.061	98.04	90.38	83.15
	PBE-D3	-411.423	1.898	8.307	8.687	9.063	98.38	90.97	83.29
	PBE	-127.451	1.925	8.119	8.420	9.443	98.30	90.69	82.56
D559	vdW-DF	-127.415	1.933	8.137	8.379	9.442	98.39	90.50	82.22
	PBE-D3	-411.459	1.895	8.207	8.462	9.450	98.66	91.22	82.76
	PBE	-127.545	1.914	8.051	9.604	8.403	81.04	83.05	89.45
F377	vdW-DF	-127.591	1.930	8.071	9.637	8.293	80.96	82.72	89.63
	PBE-D3	-411.682	1.919	8.180	9.298	8.532	80.58	83.01	88.90
	PBE	-127.498	1.900	8.854	8.378	8.709	84.61	86.44	91.46
G1766	vdW-DF	-127.532	1.918	8.918	8.315	8.647	84.35	85.57	91.32
	PBE-D3	-411.547	1.906	8.953	8.319	8.654	84.16	86.57	91.41
	PBE	-127.354	1.879	11.690	8.364	6.709	91.57	96.94	85.42
H223	vdW-DF	-127.259	1.908	11.523	8.387	6.732	91.19	98.93	83.99
	PBE-D3	-411.620	1.921	11.448	8.282	6.827	91.40	99.53	83.95

DFT: Density functional theory; SIESTA: Spanish Initiative for Electronic Simula-tions with Thousands of Atoms; vdW-DF: van der Waals Density Functional; VASP: Vienna *ab initio* simulation package; PBE: Perdew-Burke-Ernzerhof.

timization. However, the DFT-D3 calculations with the VASP software gave a different result: the *D3* structure shown in Figure 6A exhibited the highest density of 1.937 g/cm^3 and the lowest lattice energy. To our knowledge, there have been no previous theoretical or experimental reports of 1:1 HMX:CL-20 cocrystal structures, and this is the first report of 1:1 HMX:CL-20 cocrystal structures with a density higher than that of the β-HMX (d = 1.901 g/cm^3)[70] crystal.
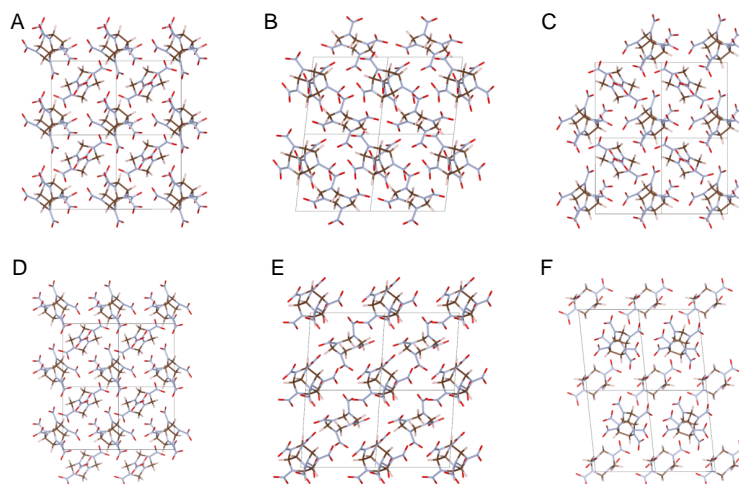
Figure 6. Cocrystal structures with 1:1 HMX:CL-20 ratio after DFT-D3 (VASP) optimization. Structures (A–F) correspond to IDs *D3*, *D558*, *D559*, *F377*, *G1766* and *H223* in Table 3, respectively. HMX: 1,3,5,7-tetranitro-1,3,5,7-tetraazacyclooctane; CL-20: 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane; DFT: density functional theory; VASP: Vienna *ab initio* simulation package.
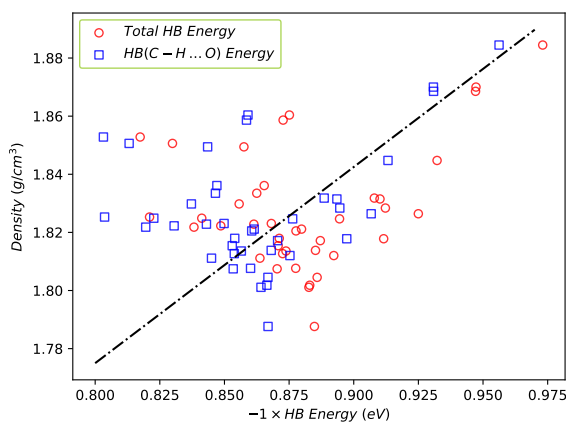

Figure 7. Relationship between hydrogen bond energy and density of predicted cocrystals. The hydrogen bond energy was calculated using the *ReaxFF-nn* potential. *ReaxFF-nn*: Reactive force field with neural networks.

Hydrogen bond interactions

Hydrogen bond interactions play an important role in the design of molecular crystals; in practice, we found that increasing the C–H...O hydrogen bond contribution to the total energy can apparently raise the rate of high-density cocrystals. This leads to the conclusion[79] that the C–H...O interaction can be employed to design molecular crystals. Figure 7 shows the relationship between hydrogen bond energy (calculated by the *ReaxFF-nn* potential) and crystal density. The figure shows that the C–H...O hydrogen bond energy represents the main part of the total hydrogen bond energy; moreover, an approximately linear relationship is observed between the hydrogen bond energy and the density of the cocrystals, with a correlation factor $\eta = 0.67$, which indicates that the C–H...O hydrogen bond energy is the key factor in the design of HMX/CL-20 cocrystals.

CONCLUSIONS

In this work, we report an evolutionary algorithm and machine learning workflow denoted as *search–check–train–search* to explore new cocrystal structures of HMX and CL-20, which are expected to have higher energy than β-HMX and lower sensitivity than ε-CL-20. Using the present workflow, several high-density HMX/CL-20 energetic cocrystal structures with 1:1 and 1:2 HMX:CL-20 ratios were found and validated *via* DFT and DFT-D3 calculations. The most valuable finding was the *D3* cocrystal structure, which had the highest density and lowest lattice energy among the 1:1 cocrystal structures. The density of *D3* (1.937 g/cm^3) was only slightly lower than that of 1:2 HMX:CL-20 (1.945 g/cm^3) reported by Bolton *et al.*[38] at room temperature. Another high-density cocrystal with a 1:1 ratio was *H223*, which had a similar lattice energy and a density of 1.921 g/cm^3. Among cocrystals with 1:2 HMX:CL-20 ratio, *F111* had the lowest lattice energy and the highest density of 1.937 g/cm^3. These high-density cocrystal structures can be accessed on the website[80]. In practice, we found that the C–H. . . O hydrogen bond interactions play an important role in the search for high-density cocrys-tals. Although these interactions are weak, with an equilibrium hydrogen bond length of 2.254 Å estimated through EOS calculations, they account for the vast majority of the total hydrogen bond energy of HMX/CL-20 cocrystals. An approximately linear relationship was found between the hydrogen bond energy and the density of cocrystal structures.

DECLARATIONS

Authors' contributions
Designed this project: Guo F
Complete all the calculations: Ye ZH
Provided technical and material support: Wen YS, Chai CG, Zhang ZR
Offered writing suggestions: Wang XC, Li HS, Zhang GQ, Cui SX

Availability of data and materials
The raw data (high-density cocrystals and parameter set for LAMMPS) are available from the website[80].

Conflicts of interest
All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

REFERENCES

1.   Bartók AP, Csányi G. Gaussian approximation potentials: a brief tutorial introduction. *Int J Quantum Chem* 2015;115:1051–7. DOI

2.   Dragoni D, Daff TD, Csányi G, Marzari N. Achieving DFT accuracy with a machine-learning interatomic potential: thermomechanics and defects in bcc ferromagnetic iron. *Phys Rev Mater* 2018;2:013808. DOI

3.   Bartók AP, Payne MC, Kondor R, Csányi G. Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons. *Phys Rev Lett* 2010;104:136403. DOI

4.   Szlachta WJ, Bartók AP, Csányi G. Accuracy and transferability of Gaussian approximation potential models for tungsten. *Phys Rev B* 2014;90:104108. DOI

5.   Botu V, Ramprasad R. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys Rev B* 2015;92:094306. DOI

6.   Li Z, Kermode JR, De Vita A. Molecular dynamics with on-the-fly machine learning of quantum-mechanical forces. *Phys Rev Lett* 2015;114:096405. DOI

7.   Novikov IS, Gubaev K, Podryabinkin EV, Shapeev AV. The MLIP package: moment tensor potentials with MPI and active learning. *Mach Learn Sci Technol* 2021;2:025002. DOI

8.   Shapeev AV. Moment tensor potentials: a class of systematically improvable interatomic potentials. *Multiscale Model Simul* 2016;14:1153–73. DOI

9.   Novikov IS, Suleimanov YV, Shapeev AV. Automated calculation of thermal rate coefficients using ring polymer molecular dynamics and machine-learning interatomic potentials with active learning. *Phys Chem Chem Phys* 2018;20:29503–12. DOI

10. Zeng J, Zhang D, Lu D, et al. DeePMD-kit v2: a software package for deep potential models. *J Chem Phys* 2023;159:054801. DOI

11. Li J, An Q. Quasiplastic deformation in shocked nanocrystalline boron carbide: grain boundary sliding and local amorphization. *J Eur Ceram Soc* 2023;43:208–16. DOI

12. Pitike KC, Setyawan W. Accurate Fe–He machine learning potential for studying He effects in BCC-Fe. *J Nucl Mater* 2023;574:154183. DOI

13. Zhai B, Wang HP. Accurate interatomic potential for the nucleation in liquid Ti-Al binary alloy developed by deep neural network learning method. *Comput Mater Sci* 2023;216:111843. DOI

14. Lyakhov AO, Oganov AR, Stokes HT, Zhu Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput Phys Commun* 2013;184:1172–82. DOI

15. Zhu Q, Oganov AR, Glass CW, Stokes HT. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystal Sect B Struct Sci* 2012;68:215–26. DOI

16. Wang Y, Lv J, Zhu L, Ma Y. CALYPSO: a method for crystal structure prediction. *Comput Phys Commun* 2012;183:2063–70. DOI

17. Yamashita T, Kanehira S, Sato N, et al. CrySPY: a crystal structure prediction tool accelerated by machine learning. *Sci Technol Adv Mater Methods* 2021;1:87–97. DOI

18. Li C, Wang C, Sun M, et al. Correlated RNN framework to quickly generate molecules with desired properties for energetic materials in the low data regime. *J Chem Inf Model* 2022;62:4873–87. DOI

19. Choudhary K, DeCost B, Chen C, et al. Recent advances and applications of deep learning methods in materials science. *npj Comput Mater* 2022;8:59. DOI

20. Hong C, Choi JM, Jeong W, et al. Training machine-learning potentials for crystal structure prediction using disordered structures. *Phys Rev B* 2020;102:224104. DOI

21. Chen WC, Schmidt JN, Yan D, Vohra YK, Chen CC. Machine learning and evolutionary prediction of superhard B-C-N compounds. *npj Comput Mater* 2021;7:114. DOI

22. Chen WC, Vohra YK, Chen CC. Discovering superhard B–N–O compounds by iterative machine learning and evolutionary structure predictions. *ACS Omega* 2022;7:21035–42. DOI

23. Kruglov IA, Yanilkin A, Oganov AR, Korotaev P. Phase diagram of uranium from *ab initio* calculations and machine learning. *Phys Rev B* 2019;100:174104. DOI

24. Podryabinkin EV, Tikhonov EV, Shapeev AV, Oganov AR. Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning. *Phys Rev B* 2019;99:064114. DOI

25. Bereznikova LA, Propad YV, Kruglov IA. Nitrogen phase diagram at high P–T conditions by the T-USPEX method. *J Phys Chem C* 2023;127:5683–8. DOI

26. Pakhnova M, Kruglov I, Yanilkin A, Oganov AR. Search for stable cocrystals of energetic materials using the evolutionary algorithm USPEX. *Phys Chem Chem Phys* 2020;22:16822–30. DOI

27. Wang C, Ni Y, Zhang C, Xue X. Crystal structure prediction of 2,4,6,8,10,12-Hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (CL-20) by a tailor-made OPLS-AA force field. *Cryst Growth Des* 2021;21:3037–46. DOI

28. Wespiser C, Mathieu D. Application of machine learning to the design of energetic materials: preliminary experience and comparison with alternative techniques. *Prop Explos Pyrotech* 2023;48:e202200264. DOI

29. Zhang Y, Wang H, Chen W, et al. DP-GEN: A concurrent learning platform for the generation of reliable deep learning based potential energy models. *Comput Phys Commun* 2020;253:107206. DOI

30. Vandermause J, Torrisi SB, Batzner S, et al. On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events. *npj Comput Mater* 2020;6:20. DOI

31. Venkata Viswanath J, Venugopal KJ, Srinivasa Rao NV, Venkataraman A. An overview on importance, synthetic strategies and studies of 2,4,6,8,10,12-hexanitro-2,4,6,8,10,12-hexaazaisowurtzitane (HNIW). *Def Technol* 2016;12:401–18. DOI

32. Taylor JW, Crookes RJ. Vapour pressure and enthalpy of sublimation of 1,3,5,7-tetranitro-1,3,5,7-tetra-azacyclo-octane (HMX). *J Chem Soc Faraday Trans 1* 1976;72:723–9. DOI

33. Tariq QuN, Tariq MuN, Dong WS, Manzoor S, Arshad F, Zhang JG. Comparative studies of synthesis, performance, and applications of

recently developed CL-20 based co-crystals. *Cryst Growth Des* 2023;23:6974–87. DOI

34. Yang Z, Li H, Zhou X, Zhang C, et al. Characterization and properties of a novel energetic–energetic cocrystal explosive composed of HNIW and BTF. *Cryst Growth Des* 2012;12:5155–8. DOI

35. Xu H, Duan X, Li H, Pei C. A novel high-energetic and good-sensitive cocrystal composed of CL-20 and TATB by a rapid solvent/non-solvent method. *RSC Adv* 2015;5:95764–70. DOI

36. Liu N, Duan B, Lu X, et al. Preparation of CL-20/DNDAP cocrystals by a rapid and continuous spray drying method: an alternative to cocrystal formation. *Cryst Eng Comm* 2018;20:2060–7. DOI

37. Wang Y, Yang Z, Li H, et al. A novel cocrystal explosive of HNIW with good comprehensive properties. *Prop Explos Pyrotech* 2014;39:590–6. DOI

38. Bolton O, Simke LR, Pagoria PF, Matzger AJ. High power explosive with good sensitivity: a 2:1 cocrystal of CL-20:HMX. *Cryst Growth Des* 2012;12:4311–4. DOI

39. Corpinot MK, Bučar DK. A practical guide to the design of molecular crystals. *Cryst Growth Des* 2019;19:1426–53. DOI

40. van Duin ACT, Dasgupta S, Lorant F, Goddard WA. ReaxFF: A reactive force field for hydrocarbons. *J Phys Chem A* 2001;105:9396–409. DOI

41. Senftle TP, Hong S, Islam MM, et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Comput Mater* 2016;2:15011 DOI

42. Han Y, Jiang D, Zhang J, Li W, Gan Z, Gu J. Development, applications and challenges of ReaxFF reactive force field in molecular simulations. *Front Chem Sci Eng* 2016;10:16–38. DOI

43. Chenoweth K, van Duin ACT, Goddard WA. ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation. *J Phys Chem A* 2008;112:1040–53. DOI

44. Guo F, Zhang H, Hu HQ, Cheng XL. Effects of hydrogen bonds on solid state TATB, RDX, and DATB under high pressures. *Chin Phys B* 2014;23:046501. DOI

45. Liu J, Li X, Guo L, et al. Reaction analysis and visualization of ReaxFF molecular dynamics simulations. *J Mol Graphics Modell* 2014;53:13–22. DOI

46. Zhou T, Song H, Liu Y, Huang F. Shock initiated thermal and chemical responses of HMX crystal from ReaxFF molecular dynamics simulation. *Phys Chem Chem Phys* 2014;16:13914–31. DOI

47. Huang X, Guo F, Yao K, et al. Anisotropic hydrogen bond structures and orientation dependence of shock sensitivity in crystalline 1,3,5-tri-amino-2,4,6-tri-nitrobenzene (TATB). *Phys Chem Chem Phys* 2020;22:11956–66. DOI

48. Daksha CM, Yeon J, Chowdhury SC, Gillespie JW Jr. Automated ReaxFF parametrization using machine learning. *Comput Mater Sci* 2021;187:110107. DOI

49. Bu Y, Guo F, Li K, et al. High-temperature pyrolysis behavior and structural evolution mechanism of graphene oxide: a ReaxFF molecular dynamics simulation. *Appl Surf Sci* 2022;593:153451. DOI

50. Jiang J, Wang HR, Zhao FQ, Xu SY, Ju XH. Decomposition mechanism of 1,3,5-trinitro-2,4,6-trinitroaminobenzene under thermal and shock stimuli using ReaxFF molecular dynamics simulations. *Phys Chem Chem Phys* 2023;25:3799–805. DOI

51. Feng S, Guo F, Yuan C, et al. Effect of neutron irradiation on structure and decomposition of α-RDX: a ReaxFF molecular dynamics study. *Comput Theor Chem* 2023;1219:113965. DOI

52. Xue LY, Guo F, Wen YS, et al. ReaxFF-MPNN machine learning potential: a combination of reactive force field and message passing neural networks. *Phys Chem Chem Phys* 2021;23:19457–64. DOI

53. Guo F, Wen YS, Feng SQ, et al. Intelligent-ReaxFF: evaluating the reactive force field parameters with machine learning. *Comput Mater Sci* 2020;172:109393. DOI

54. Friederich P, Häse F, Proppe J, Aspuru-Guzik A. Machine-learned potentials for next-generation matter simulations. *Nat Mater* 2021;20:750–61. DOI

55. Schütt KT, Sauceda HE, Kindermans PJ, Tkatchenko A, Müller KR. SchNet - A deep learning architecture for molecules and materials. *J Chem Phys* 2018;148:241722. DOI

56. Yoo P, Sakano M, Desai S, Islam MM, Liao P, Strachan A. Neural network reactive force field for C, H, N, and O systems. *npj Comput Mater* 2021;7:9. DOI

57. Rappe AK, Goddard WA III. Charge equilibration for molecular dynamics simulations. *J Phys Chem* 1991;95:3358–63. DOI

58. Islam MM, Kolesov G, Verstraelen T, Kaxiras E, van Duin ACT. eReaxFF: a pseudoclassical treatment of explicit electrons within reactive force field simulations. *J Chem Theory Comput* 2016;12:3463–72. DOI

59. Larsen AH, Mortensen JJ, Blomqvist J, et al. The atomic simulation environment - a Python library for working with atoms. *J Phys: Condens Matter* 2017;29:273002. DOI

60. Guo F. I-ReaxFF: intelligent-reactive force field. Available from: https://github.com/fenggo/I-ReaxFF. [Last accessed on 15 May 2024]

61. Sivaraman G, Krishnamoorthy AN, Baur M, et al. Machine-learned interatomic potentials by active learning: amorphous and liquid hafnium dioxide. *npj Comput Mater* 2020;6:104. DOI

62. Shi B, Zhou Y, Fang D, et al. Estimating the performance of a material in its service space via Bayesian active learning: a case study of the damping capacity of Mg alloys. *J Mater Inf* 2022;2:8. DOI

63. Landenberger KB, Matzger AJ. Cocrystals of 1,3,5,7-Tetranitro-1,3,5,7-tetrazacyclooctane (HMX). *Cryst Growth Des* 2012;12:3603–9. DOI

64. Millar DIA, Maynard-Casely HE, Allan DR, et al. Crystal engineering of energetic materials: co-crystals of CL-20. *Cryst Eng Comm* 2012;14:3742–9. DOI

65. Bidault X, Chaudhuri S. A flexible-molecule force field to model and study hexanitrohexaazaisowurtzitane (CL-20) – polymorphism under extreme conditions. *RSC Adv* 2019;9:39649–61. DOI

66. Zhang XQ, Chen XR, Kaliamurthi S, Selvaraj G, Ji GF, Wei DQ. Initial decomposition of the co-crystal of CL-20/TNT: sensitivity decrease under shock loading. *J Phys Chem C* 2018;122:24270–8. DOI

67. Zhang XQ, Yuan JN, Selvaraj G, Ji GF, Chen XR, Wei DQ. Towards the low-sensitive and high-energetic co-crystal explosive CL-20/TNT: from intermolecular interactions to structures and properties. *Phys Chem Chem Phys* 2018;20:17253–61. DOI

68. Bolotina NB, Hardie MJ, Speer RL Jr, Pinkerton AA. Energetic materials: variable-temperature crystal structures of γ- and ε-HNIW polymorphs. *J Appl Cryst* 2004;37:808–14. DOI

69. Cawkwell MJ, Zecevic M, Luscher DJ, Ramos KJ. Dependence of the elastic stiffness tensors of PETN, α-RDX, γ-RDX, ε-RDX, ε-CL-20, DAAF, FOX-7, and β-HMX on hydrostatic compression. *Prop Explos Pyrotech* 2022;47:e202100281. DOI

70. Deschamps JR, Frisch M, Parrish D. Thermal expansion of HMX. *J Chem Crystal* 2011;41:966–70. DOI

71. Eiland PF, Pepinsky R. The crystal structure of cyclotetramethylene tetranitramine. *Z Krist Cryst Mater* 1954;106:273–98. DOI

72. Cady HH, Larson AC, Cromer DT. The crystal structure of α-HMX and a refinement of the structure of β-HMX. *Acta Cryst* 1963;16:617–23. DOI

73. Oganov AR, Glass CW. Crystal structure prediction using *ab initio* evolutionary techniques: principles and applications. *J Chem Phys* 2006;124:244704. DOI

74. Oganov AR, Lyakhov AO, Valle M. How evolutionary crystal structure prediction works - and why. *Acc Chem Res* 2011;44:227–37. DOI

75. Gale JD, Raiteri P, van Duin ACT. A reactive force field for aqueous-calcium carbonate systems. *Phys Chem Chem Phys* 2011;13:16666–79. DOI

76. Soler JM, Artacho E, Gale JD, et al. The SIESTA method for *ab initio* order-N materials simulation. *J Phys Condens Matter* 2002;14:2745. DOI

77. Dion M, Rydberg H, Schröder E, Langreth DC, Lundqvist BI. Van der Waals density functional for general geometries. *Phys Rev Lett* 2004;92:246401. DOI

78. Kresse G, Furthmüller J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set. *Phys Rev B* 1996;54:11169–86. DOI

79. Gavezzotti A, Presti LL. Building blocks of crystal engineering: a large-database study of the intermolecular approach between C-H donor groups and O, N, Cl, or F acceptors in organic crystals. *Cryst Growth Des* 2016;16:2952–62. DOI

80. Guo F. ReaxFF-nn for lammps. Available from: https://github.com/fenggo/ReaxFF-nn_for_lammps. [Last accessed on 15 May 2024]