

## Supporting Information (SI)

### **High Precision Deep-Learning Model Combined with High-Throughput Screening to Discover Fused [5,5] Biheterocyclic Energetic Materials with Excellent Comprehensive Properties**

Youhai Liu<sup>a</sup>, Fusheng Yang<sup>\*a</sup>, Wenquan Zhang<sup>\*b</sup>, Honglei Xia<sup>b</sup>, Zhen Wu<sup>a</sup>, Zaoxiao Zhang<sup>a</sup>

<sup>a</sup>School of Chemical Engineering and Technology, Xi'an Jiaotong University, Xi'an 710049, China

<sup>b</sup>Research Center of Energetic Material Genome Science, Institute of Chemical Materials, China Academy of Engineering Physics (CAEP), Mianyang, 621900, P. R. China.

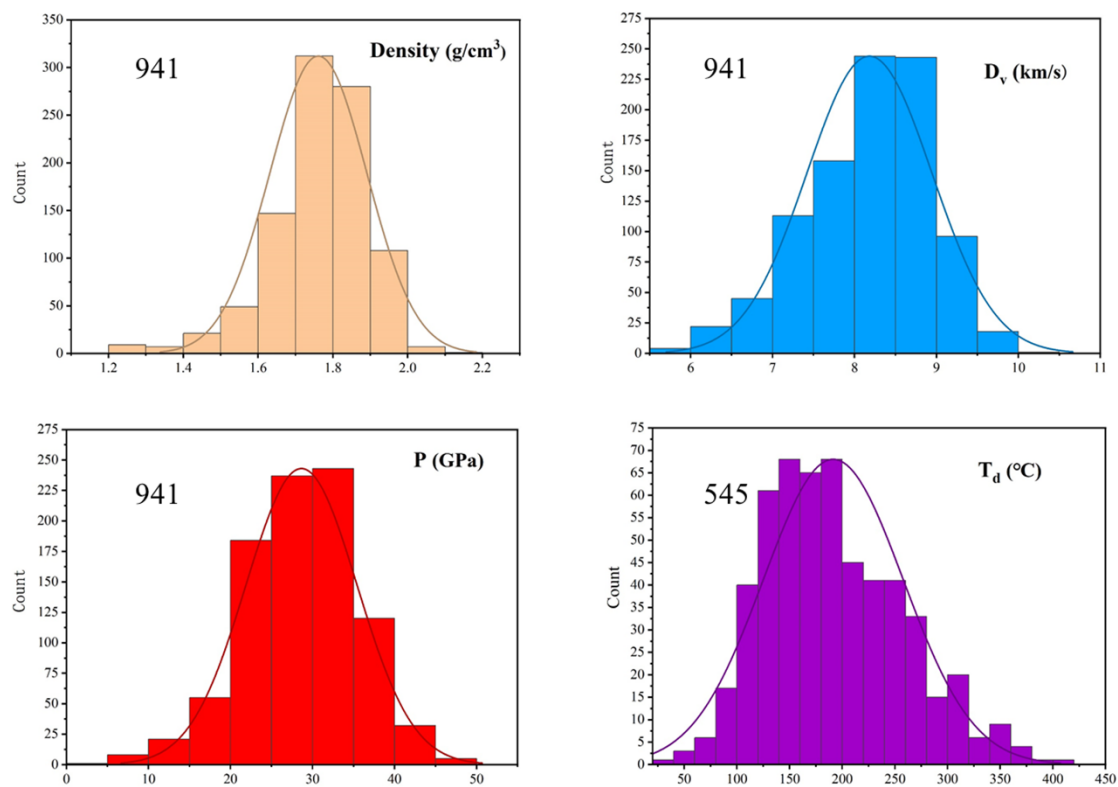
\*Corresponder Email: [yang.fs@mail.xjtu.edu.cn](mailto:yang.fs@mail.xjtu.edu.cn) (Fusheng Yang)

[zhangwq-cn@caep.cn](mailto:zhangwq-cn@caep.cn) (Wenquan Zhang)

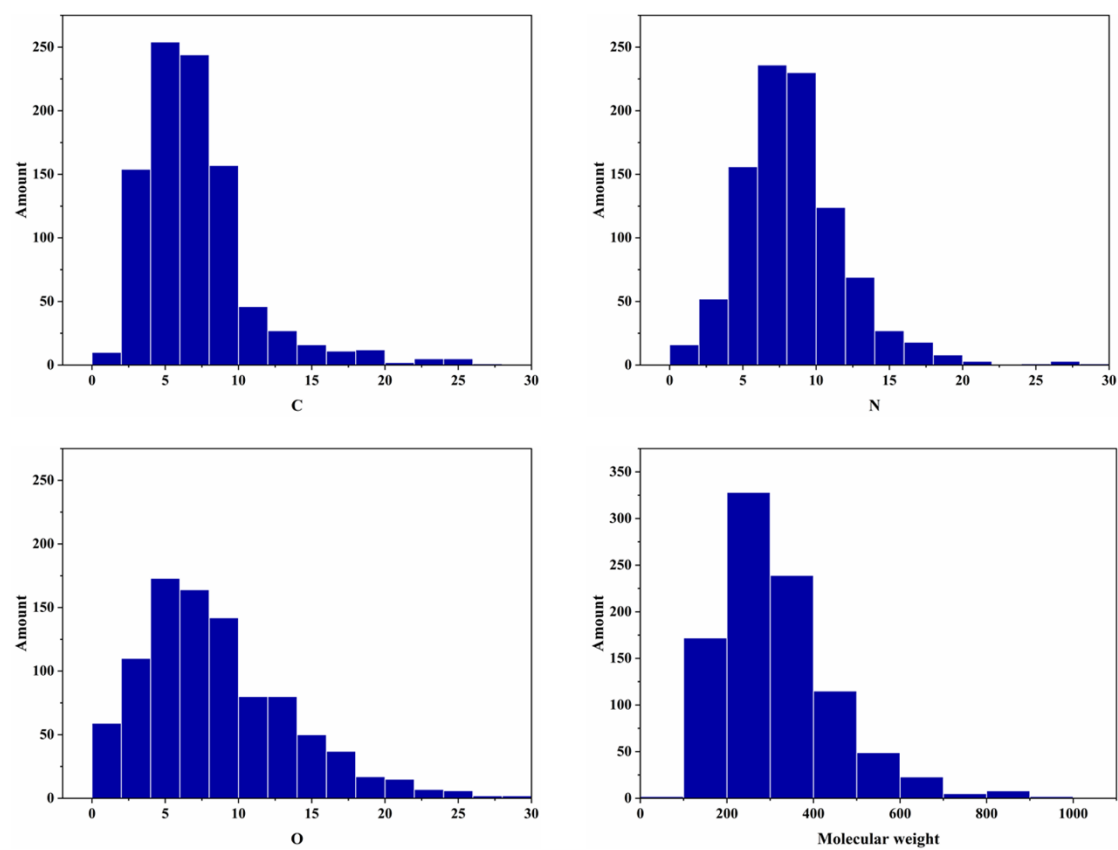
## Table of Contents

Data distribution.....	S3
Features distribution.....	S5
Prediction results for detonation pressure and decomposition temperature.....	S6
Regional distribution of molecular electrostatic potential of the ten compounds.....	S7
Electro-topological state (E-state) fingerprint.....	S9
Abbreviation and description for custom descriptors .....	S10
Atom and bond features in D-MPNN.....	S11
Computational detail .....	S12
The structures and SAScores of the screened 10 molecules.....	S16
Reference.....	S18

**Fig. S1 Data distribution**



**(a)**

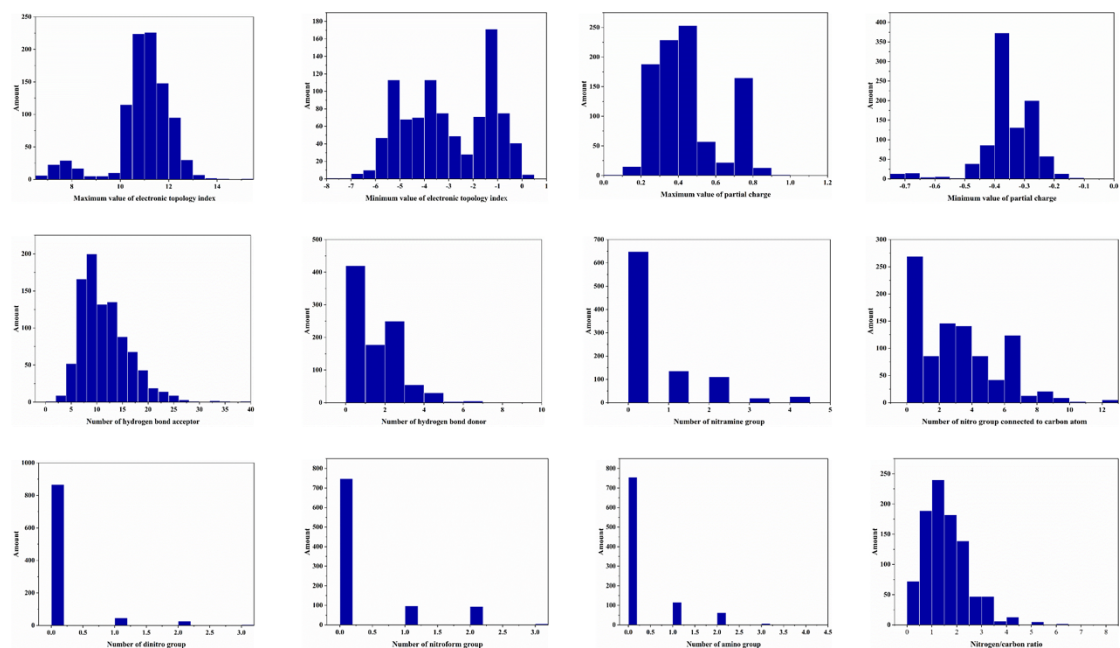


**(b)**

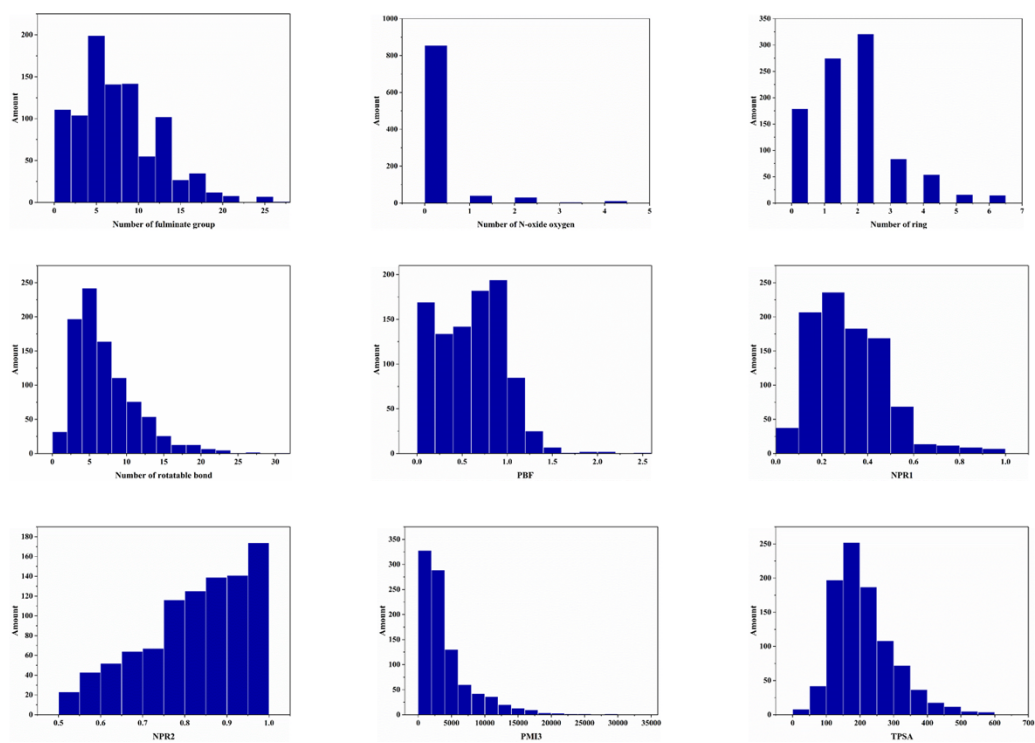
The data of density was mainly the crystal density obtained by single crystal X-ray diffraction. The detonation velocity and detonation pressure were derived from calculated from Explo5 software. For the decomposition temperature dataset, we used the initial decomposition temperature (temperature decomposition onset) as the data label, and only experimental data with heating rates between 5°C and 10°C are included. If different literature sources report different experimental temperatures for the same molecular structure, the higher decomposition temperature value is selected.

The distribution range of properties (density,  $D_v$ ,  $P$  and  $T_d$ ) for training regression models are summarized, it can be found that the distribution basically conforms to the "low high low" normal distribution curve. The distribution of data is reasonable and covers common distribution intervals of energetic materials.

Fig S2 Custom features distribution

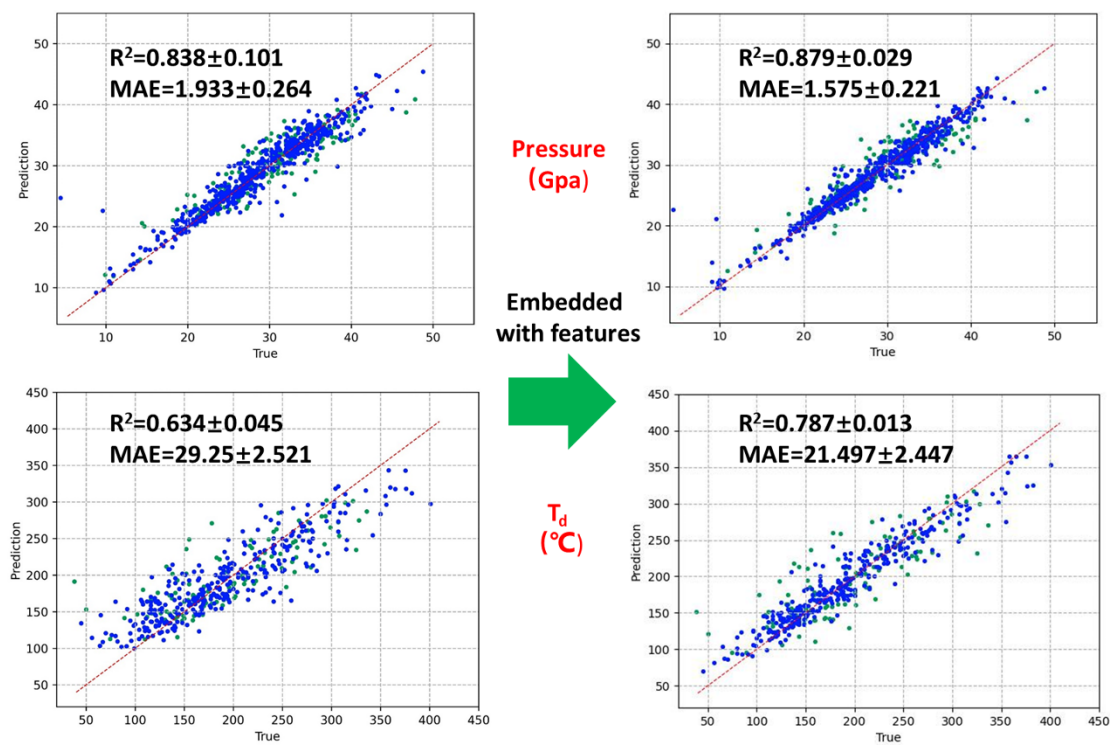


(a)

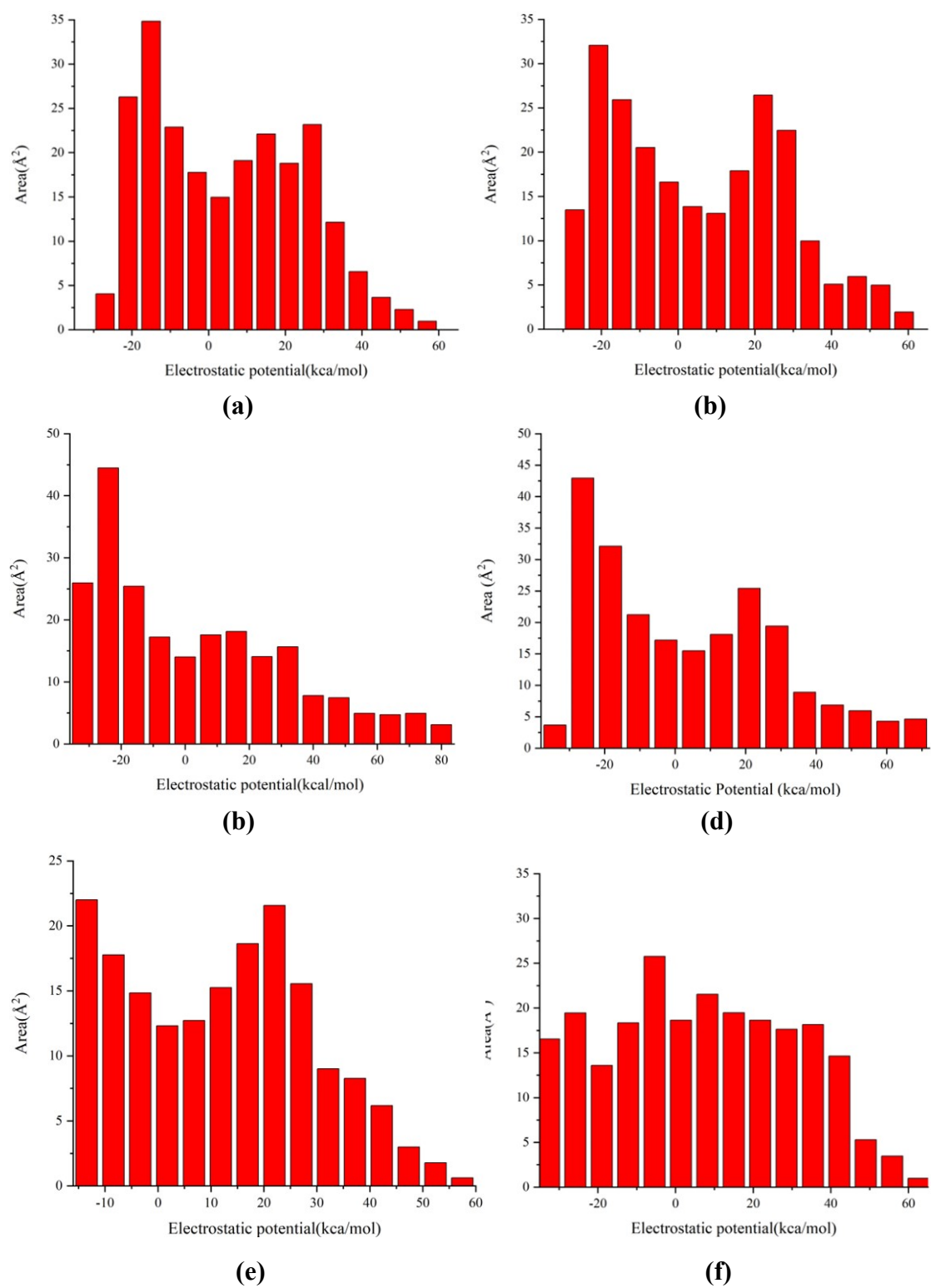


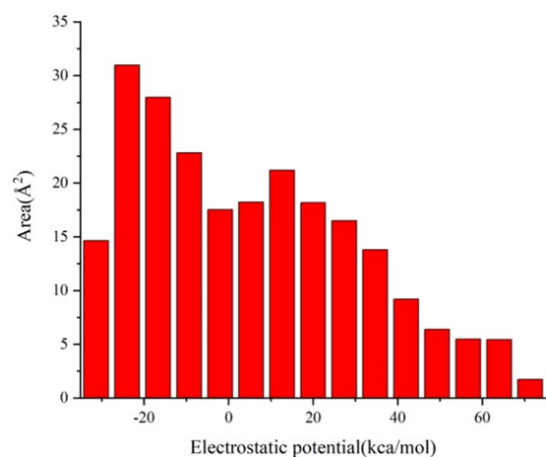
(b)

**Fig S3. Prediction results before and after the inclusion of embedded features for detonation pressure and decomposition temperature**

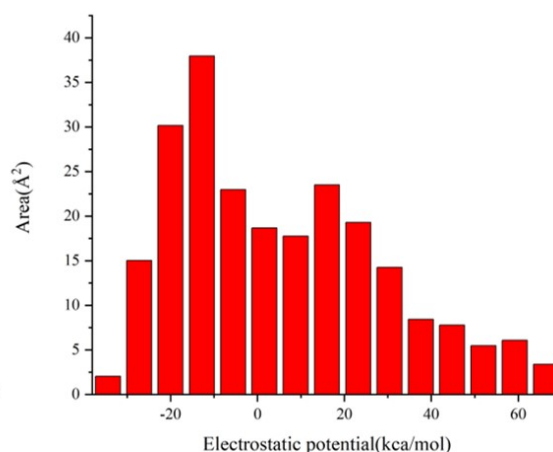


**Figure S4. Regional distribution of molecular electrostatic potential of the ten compounds (a)-(i) represents 805, 799, 1039, 1518, 1525, 879, 1439, 1521, 1441, 469, respectively**

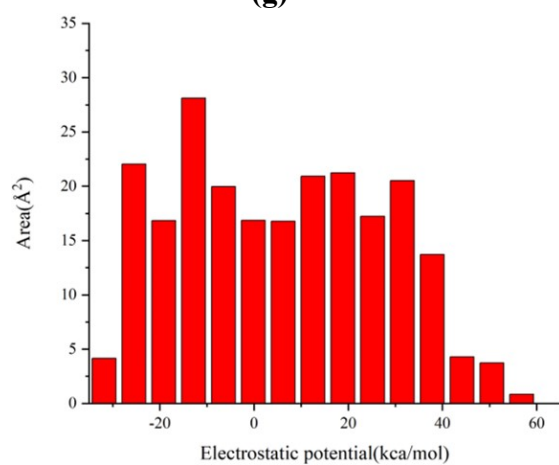




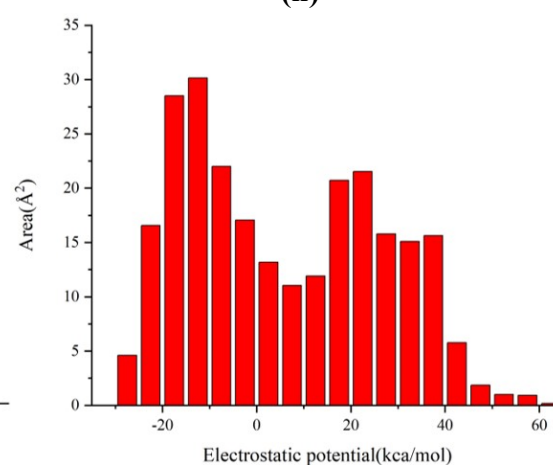
**(g)**



**(h)**



**(i)**



**(j)**

It is obvious from Figure S3 that the directions of the histograms of the molecular surface ESP of these compounds are approximately the same, showing an inverted “U” shape. The integral area of the negative region is larger than that of the positive region, which indicates that all the ten compounds show low mechanical sensitivity <sup>1</sup>.



**Table S1. E-state fingerprint**

Index	Atomic type	Index	Atomic type	Index	Atomic type
1	-Li	28	=N-	55	-GeH <sub>3</sub>
2	-Be-	29	aNa	56	-GeH <sub>2</sub> -
3	>Be< [-2]	30	>N-	57	>GeH-
4	-BH-	31	-N<<	58	>Ge<
5	>B-	32	aaNs	59	-AsH>
6	>B< [-1]	33	>N< [+1]	60	-AsH-
7	-CH <sub>3</sub>	34	-OH	61	>As
8	=CH <sub>2</sub>	35	=O	62	->AS=
9	-CH <sub>2</sub> -	36	-O-	63	->As<
10	≡CH	37	aOa	64	-SeH
11	=CH-	38	-F-	65	=Se
12	aCHa	39	-SiH <sub>3</sub>	66	-Se-
13	>CH-	40	-SiH <sub>2</sub> -	67	aSea
14	=C=	41	>SiH-	68	>Se=
15	≡C-	42	>Si<	69	≥Se=
16	=C<	43	-PH <sub>2</sub>	70	-Br
17	aCa-	44	-PH-	71	-SnH <sub>3</sub>
18	aaCa	45	>P-	72	-SnH <sub>2</sub> -
19	>C<	46	->p=	73	>SnH-
20	-NH <sub>3</sub> [+1]	47	->p<	74	>Sn<
21	-NH <sub>2</sub>	48	-SH	75	-I
22	-NH <sub>2</sub> - [+1]	49	=S	76	-PbH <sub>3</sub>
23	=NH	50	-S	77	-PbH <sub>2</sub> -
24	-NH-	51	aSa	78	>PbH-
25	aNHa	52	>S=	79	>Pb<
26	≡N	53	≥S≤		
27	>NH- [+1]	54	-Cl		

In this work, E-state fingerprint only involved carbon (C), hydrogen (H), oxygen (O), nitrogen (N) elements (index 7-37).

**Table S2. Abbreviation and description for custom descriptors**

Index	Abbreviation	Description	Index	Abbreviation	Description
1	nHA	Number of hydrogen bond acceptor	20	nC	Number of carbon atom
2	nHD	Number of hydrogen bond donor	21	nN	Number of nitrogen atom
3	TPSA	Topological polar surface area	22	nO	Number of oxygen atom
4	MinPartialCharge	Minimum value of partial charge	23	nH	Number of hydrogen atom
5	MaxPartialCharge	Maximum value of partial charge	24	OB	Oxygen balance
6	MaxEstateIndex	Maximum value of electronic topology index	25	Molecular weight	Molecular weight
7	MinEstateIndex	Minimum value of electronic topology index	26	nN/nC	Nitrogen/carbon ratio
8	nNH <sub>2</sub>	Number of amino group	27	PBF	Plane of best fit
9	nNNO <sub>2</sub>	Number of nitramine group	28	NPR1	Normalized principal moments ratios 1
10	nC(NO <sub>2</sub> ) <sub>3</sub>	Number of nitroform group	29	NPR2	Normalized principal moments ratios 2
11	nC(NO <sub>2</sub> ) <sub>2</sub>	Number of dinitro group	30	PMI3	Principal moments of inertia 3
12	nC(NO <sub>2</sub> )	Number of nitro group connected to carbon atom			
13	nNO <sub>2</sub>	Number of nitro fragment			
14	n(C=N-O)	Number of fulminate group			
15	n(N-O-C)	Number of N-oxide oxygen			
16	nAHC	Number of aromatic heterocycle			
17	nACC	Number of aromatic carbocycle			
18	nR	Number of ring			
19	nRbond	Number of rotatable bond			

Index 1-7 represent molecular electronic states; index 8-15 represent the number of typical energy groups; index 16-26 represent molecular composition and index 27-30 represent three-dimensional shape of molecules.

**Table S3. Atom features<sup>a</sup>**

feature	description	size
atom type	type of atom (ex. C, N, O), by atomic number	100
# bonds	number of bonds the atom is involved in	6
Formal charge	integer electronic charge assigned to atom	5
chirality	unspecified, tetrahedral CW/CCW, or other	4
# Hs	number of bonded hydrogen atoms	5
hybridization	sp, sp <sup>2</sup> , sp <sup>3</sup> , sp <sup>3</sup> d, or sp <sup>3</sup> d <sup>2</sup>	5
aromaticity	whether this atom is part of an aromatic system	1
Atomic mass	mass of the atom, divided by 100	1

<sup>a</sup>All features are one-hot encodings except for atomic mass, which is a real number scaled to be on the same order of magnitude.

**Table S4. Bond features<sup>a</sup>**

feature	description	size
bond type	single, double, triple, or aromatic	4
conjugated	whether the bond is conjugated	1
in ring	whether the bond is part of a ring	1
stereo	none, any, E/Z or cis/trans	6

<sup>a</sup>All features are one-hot encodings

## Computational detail

We employ density functional theory (DFT) to investigate the key properties of the selected fused [5,5] bicyclic heterocycle energetic materials. The computations were implemented in the Gaussian 16 (A.03) <sup>2</sup> program and analyzed by wave function analysis software Multiwfn 3.8 (dev) <sup>3</sup>. Geometric structures of these compounds were optimized at the level of Becke three parameter, Lee-Yan-Parr (B3LYP) function with 6-311+G (d, p) basis set <sup>4</sup>. All the optimized structures of minimal local energy on the potential surface without imaginary frequency were ensured. Atomization energies were calculated by the CBS-4M <sup>5</sup>.

## Heat of formation

The predictions of heat of formation (HOF) adopt the hybrid DFT-B3LYP methods with 6-311+G\*\* basis set via designed isodesmic reactions. The isodesmic reaction processes, i.e., the number of each kind of formal bond is conserved, are used with application of the bond separation reaction (BSR) rules. In order to keep the number of bonds constant to reduce computational errors, the molecule is broken down into a set of two small molecules containing the same component bonds. The isodesmic reactions used to derive the HOF of the selected molecules are in Scheme. The change of enthalpy for the reactions at 298 K can be expressed as:

$$\Delta H_{298} = \sum \Delta_f H_p - \sum \Delta_f H_R \quad (1)$$

Where  $\Delta_f H_p$  and  $\Delta_f H_R$  are the HOF of reactants and products at 298 K, respectively, and  $\Delta H_{298}$  can also be calculated using the following equation:

$$\Delta H_{298} = \Delta E_{298} + \Delta(PV) = \Delta E_0 + \Delta ZPE + \Delta H_T + \Delta nRT \quad (2)$$

Where  $\Delta E_0$  is the change in total energy between the products and the reactants at 0 K;  $\Delta ZPE$  is the difference between the zero-point energies (ZPE) of the products and the reactants at 0 K;  $\Delta H_T$  is thermal correction from 0 to 298 K. For the isodesmic reaction,  $\Delta n = 0$ , so  $\Delta nRT = 0$ . On the left side of Eq. (1), apart from target compound, all the others can be obtained by calculated through atomization reactions at the CBS-4M level, if is not available from the experiments.

According to Hess' law of constant heat, the enthalpy of solid phase formation can be determined by the enthalpy of gas-phase formation and the enthalpy of sublimation <sup>6</sup>

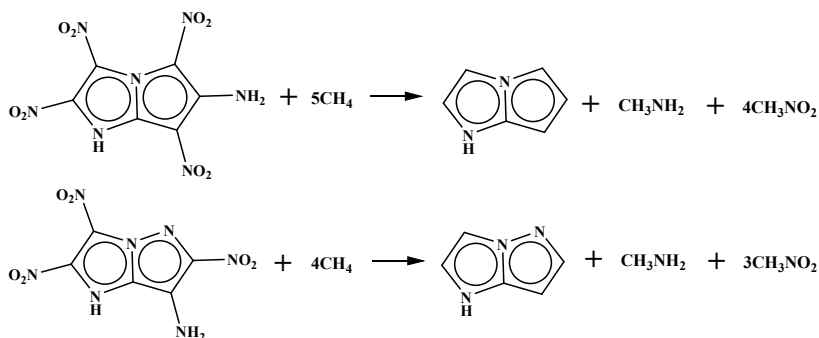
$$\Delta H_{f,solid} = \Delta H_{f,gas} - \Delta H_{sub} \quad (3)$$

The enthalpy of sublimation can be calculated using the following equation <sup>7</sup>:

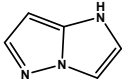
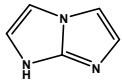
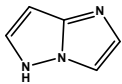
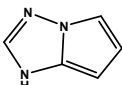
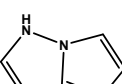
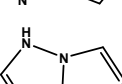
$$\Delta H_{sub} = a(A)^2 + b(\sigma_{tot}^2 \nu)^{0.5} + c \quad (4)$$

Where  $A$  is the molecular surface area with level of 0.001 electron/bohr<sup>3</sup>,  $\sigma_{tot}^2$  is represented as the total mean square error of the electrostatic potential on the molecular surface,  $\nu$  is the balance between positive and negative electrostatic potentials on the molecular surface, and  $a$ ,  $b$ ,  $c$  are fitting parameters (values are  $2.670 \times 10^{-4}$  kcal/mol/Å<sup>4</sup>, 1.650 kcal/mol, 2.966 kcal/mol, respectively).

**Scheme** Isodesmic reactions for 469 and others nine compounds have similar structures



**Table S5.** Ab initio computational values of small molecules used in isodesmic reactions

Compound	$E_0^a$	ZPE <sup>b</sup>	$H_T^c$	HOF <sup>d</sup>
CH <sub>4</sub>	-40.5339263	112.26	10.04	-74.6
CH <sub>3</sub> NH <sub>2</sub>	-95.8938402	160.78	11.64	-22.5
CH <sub>3</sub> NO <sub>2</sub>	-245.09	124.93	11.6	-80.8
	-357.840	260.366	17.599	301.197
	-357.861	260.875	17.415	244.951
	-357.825	260.670	16.948	341.703
	-357.836	259.819	17.507	296.460
	-357.821	259.562	17.152	349.818
	-341.772	290.273	17.512	313.220

<sup>a</sup>Total energy calculated by B3LYP/6-311+G\*\*method (a.u);

<sup>b</sup>Zero-point correction (kJ/mol);

<sup>c</sup>Thermal correction to enthalpy (kJ/mol);

<sup>d</sup>Heat of formation (kJ/mol).

## Density

Density can be obtained by using the following equation proposed by Politzer et al. <sup>8</sup>

$$\rho = \alpha \left( \frac{M}{V(0.001)} \right) + \beta (\nu \sigma_{tot}^2) + \gamma \quad (5)$$

where  $V$  (0.001) is the volume (cm<sup>3</sup>/mol) of a molecule on an isoelectronic density surface of 0.001e/bohr<sup>3</sup>,  $M$  is the molecular molar mass (g/mol),  $\alpha$ ,  $\beta$ , and  $\gamma$  are coefficients, whose values are 0.9183, 0.0028 and 0.0443, respectively.

### Detonation velocity and detonation pressure

Detonation velocity ( $D_v$ ) and detonation pressure ( $P$ ) are predicted by semi-empirical Kamlet–Jacobs equations <sup>9</sup>:

$$D_v = 1.01(N\bar{M}^{0.5}Q^{0.5})^{0.5}(1+1.30\rho) \quad (6)$$

$$P = 1.558\rho^2 N\bar{M}^{0.5}Q^{0.5} \quad (7)$$

Where  $N$  is the number of moles of gas product per gram of explosive detonation (mol/g),  $\bar{M}$  is the average molecular weight of the gas product (g/mol),  $Q$  is the detonation chemical energy (cal/g) per gram of explosive, i.e. the heat of detonation,  $\rho$  is the theoretical density obtained through Eq. (5). Calculation of  $Q$ ,  $N$ , and  $\bar{M}$  for the energetic compound C<sub>a</sub>H<sub>b</sub>O<sub>c</sub>N<sub>d</sub> is shown in Table S6.

**Table S6.** Calculation methods of  $Q$ ,  $N$ , and  $\bar{M}$  for the energetic compound C<sub>a</sub>H<sub>b</sub>O<sub>c</sub>N<sub>d</sub>

Parameter	Explosive compositions		
	$c \geq 2a + \frac{b}{2}$	$\frac{b}{2} \leq c \leq 2a + \frac{b}{2}$	$c < 2a + b$
$Q \times 10^{-3}$	$\frac{28.9b + 94.05a + 0.239\Delta H_{298k}}{M}$	$\frac{28.9b + 94.05(\frac{c}{2} - \frac{b}{4}) + 0.239\Delta H_{298k}}{M}$	$\frac{57.8c + 0.239\Delta H_{298k}}{M}$
$N$	$\frac{b + 2c + 2d}{4M}$	$\frac{b + 2c + 2d}{4M}$	$\frac{b + d}{2M}$
$\bar{M}$	$\frac{4M}{b + 2c + 2d}$	$\frac{56d + 88c - 8b}{b + 2c + 2d}$	$\frac{2b + 28d + 32c}{b + d}$

\*  $M$ ,  $\Delta H_{298k}$  is the molecular mass (g/mol) and the enthalpy of formation (KJ/mol) of C<sub>a</sub>H<sub>b</sub>O<sub>c</sub>N<sub>d</sub>, respectively.

### Bond dissociation energy (BDE)

BDE can be used to measure the strength of chemical bonds. The dissociation energy of the initiating bond can be used to measure the thermal stability of energetic compounds <sup>10</sup>. At 298 K and 1 atm, the enthalpy changes of the chemical reaction  $A-B(g) \rightarrow A\cdot(g) + B\cdot(g)$  can be defined as the bond dissociation energy of  $A-B$ . At 0 K, the BDE of compound  $A-B$  is calculated as following:

$$BDE_0(A-B) = E_0(A\cdot) + E_0(B\cdot) - E_0(A-B) \quad (8)$$

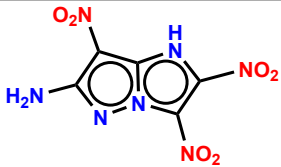
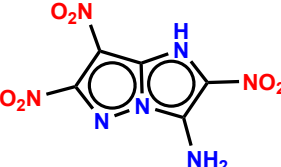
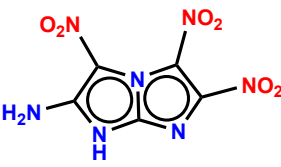
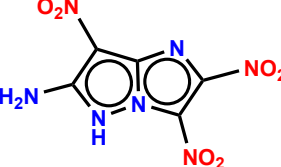
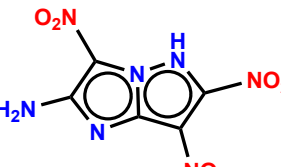
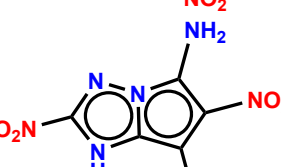
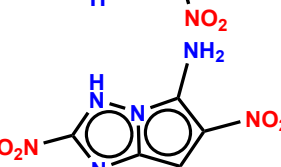
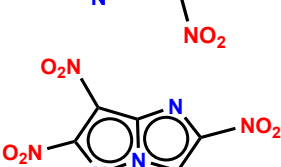
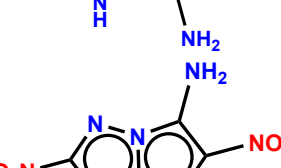
The  $A-B$  bond dissociation energy can be obtained by correcting the difference between the zero-point-corrected energy:

$$BDE(A-B) = BDE_0(A-B) + \Delta ZPE \quad (9)$$

## Crystal structure prediction

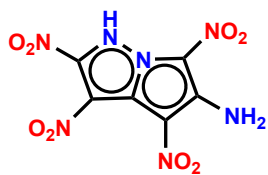
In this work, we use the DMol3 module in the Materials Studio (MS) program <sup>11</sup> to predict crystal structure. The process is as follows: after the structure of the selected compound was optimized in the Gaussian 16 program, its electrostatic potential charge (ESP) can be calculated with the widely recognized GGA-PBE function. Then, the ESP charge is embedded in the atom and the Dreiding force field suitable for predicting crystal structure <sup>12</sup>, is applied in the Polymer Predictor (PP) module of MS. Due to the fact that about 90% of organic crystal structures from the Cambridge Structural Database are belong to the space groups:  $P2_1/c$ ,  $P-1$ ,  $PNA2_1$ ,  $P2_1$ ,  $Cc$ ,  $C2/C$ ,  $P2_12_12_1$ , and  $PBCA$  <sup>13</sup>, our PP calculation is only limited to this 8 space groups. Next, the structures predicted through the eight spatial groups are arranged according to their lowest energy. Finally, the crystal structure with the lowest energy was selected to further understand the stability of the molecule with a two-dimensional fingerprint based on Hirshfeld surface <sup>3</sup>.

**Table S7.** The Index and SAScores of the screened 10 molecules corresponding of its ID number

Index	Molecule structure	SAScore
A(805)		3.718
B(799)		3.709
C(1039)		3.721
D(1518)		3.658
E(1525)		3.743
F(879)		3.708
G(1439)		3.696
H(1521)		3.779
I(1441)		3.774



J(469)



3.746

---

## Reference

- 1 D. Fischer, T. M. Klapötke and J. Stierstorfer, *Eur. J. Inorg. Chem.*, 2014, **2014**, 5808–5811.
- 2 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson and D. J. Fox., Gaussian 16, Revision A.03; Gaussian, Inc., Wallingford CT, 2016.
- 3 T. Lu and F. Chen, *J. Comput. Chem.*, 2012, **33**, 580–592.
- 4 P. C. Hariharan and J. A. Pople, *Theor. Chim. Acta.*, 1973, **28**, 231–222.
- 5 D. Fischer, T. M. Klapötke and J. Stierstorfer, *Angew. Chemie Int. Ed.*, 2015, **54**, 10299–10302.
- 6 P. W. Atkins, *Physical chemistry, 2nd edn. Oxford University Press: Oxford*, 1982.
- 7 P. Politzer, Y. Ma, P. Lane and M. C. Concha, *Int. J. Quantum Chem.*, 2005, **105**, 341–347.
- 8 P. Politzer, J. Martinez, J. S. Murray, M. C. Concha and A. Toro-Labbé, *Mol. Phys.*, 2009, **107**, 2095–2101.
- 9 M. J. Kamlet and S. J. Jacobs, *J. Chem. Phys.*, 1968, **48**, 23–35.
- 10 P. Politzer and J. S. Murray, *J. Mol. Struct.*, 1996, **376**, 419–424.
- 11 A. Inc., *Mater. Stud. 6.0 V (Accelrys Inc San Diego, CA, 2010).*, 2010.
- 12 S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
- 13 Y. Chang, Z. Lu, L. An and J. Zhang, *J. Phys. Chem. C*, 2012, **116**, 1195–1199.