

# **A Universal Deep Learning Framework based on Graph Neural Network for Virtual Co-Crystal Screening**

Yuanyuan Jiang <sup>a</sup>, Jiali Guo <sup>a</sup>, Yijing Liu <sup>b</sup>, Yanzhi Guo <sup>a</sup>,

Menglong Li<sup>a</sup>, Xuemei Pu <sup>a,\*</sup>

<sup>a</sup> College of Chemistry, Sichuan University, Chengdu, 610064

<sup>b</sup> College of Computer Science, Sichuan University, Chengdu, 610064

\* Corresponding Author

Xuemei Pu ([xmpuscu@scu.edu.cn](mailto:xmpuscu@scu.edu.cn))



**Abstract:** Cocrystal has become an effective lever to achieve novel physiochemical properties for functional molecules, thus being widely applied in various fields. However, how to choose coformers of the cocrystal is still challenging on experiments. Although some computational strategies were proposed in order to accelerate cocrystal screening, they are limited in accuracy or universality. Thus, to address the challenges, we develop a novel graph neural network (GNN) based deep learning framework to predict formation of the cocrystal. In order to support the deep learning, a large and reliable data set is first collected and organized, which contains 6819 positive samples and 1052 negative samples. Instead of only using conventional molecular descriptors in traditional ML methods or molecular graph in conventional GNNs, a complementary strategy is proposed for the feature representation through combining the molecular graph and hand-selected molecular descriptors from *priori* knowledge. Based on the multilevel feature representation, a new GNN learning architecture is constructed to effectively embed the *priori* knowledge into the “end-to-end” learning on the molecular graph. In addition, multi-head attention mechanism is introduced into the GNN architecture to further optimize the feature space through weighted summation of atom vectors instead of simply summing. Consequently, the performance of our model achieves 98.86% accuracy for the independent test set, greatly surpassing some traditional machine learning models and classic GNN models. Furthermore, the out-of-distribution prediction on the energetic cocrystals is also high up to 97.11% accuracy, showing strong robustness and generalization. Overall, our model provides quick and accurate solution for the virtual cocrystal screening. Furthermore, the new cocrystal

dataset also provides a data support for subsequent ML-based cocrystal study. All the data and source codes are available at <https://github.com/Saoge123/ccgnet>.

**Keywords**

Cocrystal; virtual screening; deep learning; graph neural network

## 1. Introduction

Cocrystals (CCs) are defined as a kind of single-phase crystalline materials composed of two or more neutral molecules assembled by non-covalent forces in definite stoichiometric ratio, which are neither solvates nor simple salts <sup>1</sup>. Since the co-crystallization could offer opportunity to achieve new properties for functional molecules <sup>2</sup>, cocrystal engineering have been widely applied in medicine, chemistry and material fields. For example, cocrystals are used as means to address physicochemical, biopharmaceutical and mechanical properties and expand solid form diversity of the active pharmaceutical ingredient (API) <sup>3</sup>. For the organic functional materials, cocrystals have advanced optical functionalities, electrical functionalities, and innovative functionalities <sup>4</sup>. In addition, the cocrystal is also a potential lever to improve the performance of explosives in order to achieve low-sensitivity and high-energy <sup>5-8</sup>.

Despite of the fascinating promise, how to choose conformers of the cocrystal is a primary key since the co-crystallization only occurs between some certain molecules <sup>9-10</sup>. Experimental determination of new co-crystals generally involves in systematic screening with a large range of coformers, thus being costly in both time, effort and laboratory resources. To mitigate the challenge, various computational ways have been proposed to predict co-formers likely to form cocrystals, including molecular dynamics simulation <sup>11</sup>, structural analysis using experimental data from the Cambridge Structural Database (CSD) <sup>12</sup>, assessing the molecular complementarity using calculated molecular electrostatic potential surfaces (MEPS) <sup>13</sup>, COSMO-RS (Conductor like Screening Model for Real Solvents) based on calculation of mixing enthalpy in a

supercooled liquid phase<sup>14</sup> and coformer screening for pharmaceutical paracetamol and indomethacin based on cloud-computing crystal structure prediction (CSP) technology involving density functional theory (DFT-D) (about ~ 50 samples)<sup>15</sup>. These methods above roughly follow knowledge-based and physics-based paradigms, which could guide the selection of cofomers and prediction of molecular coassembly processes to some extent. However, they are limited in the generalities due to the diversity of noncovalent interactions and molecular chemical structures. Therefore, it is highly desired to develop more general strategy with lower computation cost.

In recent years, data-driven machine learning (ML) methods have become increasingly popular in chemical and material fields<sup>16</sup> since its optimization strategies are automatically improved by empirical data from statistical perspectives, which can provide smart navigation in nearly infinite chemical-space<sup>17</sup>. Several works already utilized the ML methods to conduct meaningful attempts in the cocrystal field. For example, artificial neural network (ANN) was used to predict melting point, density and lattice energy of cocrystals<sup>18</sup>. Recently, Jerome et al. applied support vector machines (SVMs) to predict cocrystal formation based on 657 samples represented by concatenating 195 descriptors of each coformer that are calculated by RDkit. The accuracy of the model reached 64.7% on the test set<sup>19</sup>. In 2019, Maciej et al. developed a cocrystal screening model based on MARSplines (Multivariate Adaptive Regression Splines) and dataset composed of 608 cocrystals and 104 eutectics<sup>20</sup>. They calculated 1444 descriptors for each co-former, and used the absolute value or square of the difference between the two co-former descriptors as the sample representation. Their

model achieved 86.38% balanced accuracy on external validation. Despite some advantages of the ML-based works with respect to the experimental determination, they still suffer from several limitations. First, sufficient data is the foundation of a well-trained model while hundreds of samples were applied in the prior works. The predictive model based on the small amount of dataset has a risk of overfit and its scaling/extrapolating to a new dataset is often questionable. Secondly, the sample representation is crucial for the performance of the machine learning. However, these previous works only conducted simple incorporation of molecular descriptors, not involving in feature selection/optimization that could better capture the dominating factors of co-crystallization. Thirdly, it may be limited for the traditional ML models with a small number of parameters to map the complex relationship between co-formers and co-crystallization. Consequently, the prediction accuracies are not satisfactory for the several ML-works. Hence, it remains to be a challenge to develop a highly efficient and universal computation-strategy for the cocrystal screening.

With rapidly accumulated data and booming of Graphic Processing Units (GPUs), deep learning (DL) have been far beyond conventional ML methods in many research domains, for example, computer vision<sup>21-22</sup>, natural language processing<sup>23-24</sup>, decision making<sup>25-26</sup>, chemistry<sup>27-28</sup> and materials<sup>29-31</sup>. In particular, graph neural networks (GNNs), a subset of DL, has been receiving more and more attentions due to the great expressive power of graphs<sup>32</sup>. In chemistry, organic molecules can naturally be represented as graphs, with atoms as nodes and covalent bonds as edges. The molecular graphs can maintain the invariance of rotation and translation, thereby avoiding the

conformational challenge of using 3D representation<sup>33</sup>. With the rise of graph neural networks (GNNs), end-to-end learning on the molecular graph has been enabled to model chemical properties, which replaces traditional feature engineering through learning feature from a graph-based molecular representation<sup>34-37</sup>. However, the application of DL methods in cocrystal field, including GNN, remains nonexistent.

Motivated by these challenges above, we develop here a novel GNN-based deep learning framework as a high-performance solution to the virtual cocrystal screening, named as Co-Crystal Graph Network (CCGNet). At first, to break through the data bottleneck, we collected and organized 7871 samples from Cambridge Structural Database (CSD) and literatures, which are much more than the dataset of the previous ML models and enough to support the training of the DL models. Secondly, instead of only using conventional molecular descriptors or the molecular graph, our feature representations follow a complementary or multilevel strategy that combined the molecular graph and some important hand-selected descriptors from priori knowledge, which may not well be reflected by the molecular graph. Thirdly, based on the multilevel features of cocrystals, we explore a new GNN learning architecture, which can facilitate learning about the multilevel entities (i.e., the atoms, the bonds and the molecular descriptors), relational reasoning and the embedding rules so that integrate the molecular descriptors as global state into the feature learning on molecular graph. In addition, we introduce multi-head attention mechanism in the CCGNet framework to replace simply summing the final node states in order to further optimize the feature space with respect to, also increasing the model interpretability. Consequently, our



CCGNet model achieves 98.86% accuracy on the independent test set, significantly outperforming than some control models. Furthermore, we utilized CCGNet to perform an out-of-distribution prediction for energetic cocrystals in order to evaluate its robustness and generalization. The out-of-distribution prediction is also a common problem for the ML application in chemistry, where the samples in the test set have different structure-features from those in the training set. For example, the energetic cocrystals have much more -NO<sub>2</sub> groups than the conventional organic-molecules. Similarly, our CCGNet model also achieves high up to 97.11% accuracy for them, exhibiting high generalization performance. To make our work as a useful tool for aiding the cocrystal screening, all source code and data can be accessed at <https://github.com/Saoge123/ccgnet>.

## 2. Results and Discussions

### 2.1 Data Collections

Data availability is the critical bottleneck that limits applications of DL in the cocrystal engineering. Hundreds of samples in the previous works<sup>19-20</sup> could not support the DL training, which would lead to an increase in the risk of overfitting. Thus, we reconstruct a large dataset containing 7871 samples, which are composed of 6819 positive samples and 1052 negative ones. As depicted by **Figure 1**, the positive samples under study come from Cambridge Structural Database (CSD)<sup>38</sup>, which contains more than 1 million crystal structures of small molecules and metal-organic molecular crystals resolved by X-ray and neutron diffraction experiments. As illustrated by **Figure 1**, the cocrystals are screened from CSD in terms of the following conditions:

- 1) only containing two chemically different polyatomic units;
- 2) having 3D structures and no disorder atoms to avoid low-quality structures;
- 3) not containing any of a set common solvents or small molecule<sup>9,39</sup>, which are liquid/gaseous at room temperature and listed in the Table S3;
- 4) only containing C, H, O, N, P, S, Cl, Br, I, F and Si elements, ruling out metal elements;
- 5) molecular weight of each component less than 700, considering the fact that most of organic cocrystals are generally small molecules;
- 6) being neutral components to exclude salts because most functional cocrystals are neutral/quasineutral<sup>40</sup>.
- 7) ruling out polymorphism to remove duplicate samples, considering that different crystal structures can be formed between the two same co-formers when the crystallization conditions change.

Consequently, 6819 positive samples are obtained (vide CSD ref-code in **Table S1**). The process of positive samples is performed by CCDC python API. For the negative samples, they are collected from 186 literatures (see **Table S2** for details). For example, eutectics reported are taken as the negative samples since they are lack of long-range-order<sup>41</sup>. In addition, those that are failed to co-crystallize in the cocrystal screening experiments are served as the negative samples. All coformer structures of the negative samples are downloaded from PubChem. Then, we use the PubChem Compound CID as the unique identification of each negative coformer to remove duplications. Taken together, we collected 1052 negative samples, as listed in **Table S2**. The 7871 samples

could support the DL training and also provides data resource for studying other properties of the cocrystals in the future.

## 2.2 Representation of Samples

As accepted, the sample representation is essential for the machine learning to fit the relationship between the molecular structure and the molecular property. Organic molecules can naturally be represented as graphs with atoms as nodes and covalent bonds as edges. Despite great expressive power of the molecular graph to represent the complex topological structure of organic molecules, they may be limited in reflecting molecular properties associated with 3D structures, for example, the size, shape, steric factors, which are considered to be important for adjusting the arrangement of cocrystals<sup>2</sup>. Thus, different from most of GNNs with samples characterized only by the molecular graph, we propose a multilevel feature representation by combining the molecular graph as atom and bond levels and twelve hand-selected molecular descriptors from priori knowledge as the global level, as depicted by **Figure 2(a)**. The multilevel representations could more completely capture the driving force of the cocrystallization, in turn laying the foundation for obtaining high-performance models.

**Table 1** lists atomic and covalent bond features for the molecular graphs in the work. **Table 2** shows the 12 selected molecular descriptors involving in the molecular shape, size, polarity, flexibility, and hydrogen bond tendency, which were revealed by previous related studies to be highly correlated for the cocrystallization<sup>12, 42-43</sup>. All these descriptors under study can be quickly calculated to facilitate high-throughput screening. As reflected by **Figure 2(a)**, we embed the covalent bond into an adjacency

tensor  $\mathcal{A}$  and each slice  $A_l$  is an adjacency matrix that represent one bond type, through which the other features besides covalent bonds also can be embedded into  $\mathcal{A}$  as extra slices, for example, intermolecular interactions. The features from the atomic level are transformed to the vertex matrix while the twelve molecular descriptors from each co-former as the global state are embedded into  $2*12$  matrix.

### 2.3 Construction of Co-Crystal Graph Network Model

Based on the multilevel feature representation, we construct a new GNN model named as CCGNet. Here, we formalize the CCGNet framework by introducing related concepts of Graph Nets (GNs)<sup>44</sup> and Message Passing Neural Networks (MPNNs)<sup>45</sup> paradigms. As shown in **Figure 2(b)**, CCGNet are mainly composed of two stages, i.e., message passing phase and readout phase. The message passing is the core of MPNNs and is defined by the message functions and vertex update functions, which propagate vertex embedding to neighbors and update its embedding. As depicted by **Figure 2(c)**, the message passing phase can be consist of N CCGBlocks, which are formalized by GN block. CCGBlock involves in two trainable functions that are  $\Phi_u$  and  $\Phi_v$ . Herein,  $\Phi_u$  is defined as a global state function and is constructed by a single-layer feedforward neural network, which computes hidden representation of the global state associated with the 12 hand-selected molecular descriptors.  $\Phi_v$ , a Graph-CNN layer<sup>46</sup>, is utilized to propagate and update information between nodes/atoms of the graph using adjacent tensor that represents the edges/bonds.  $\rho^{u \rightarrow v}$  is concatenation operation, which is used to embed hidden representation of  $u$  into atom vectors of each co-former.

In the readout phase, readout function calculates embedding for classification or

regression at the graph level. In order to weight the important variables contributed to the co-crystallization, we introduce global attention mechanism<sup>47</sup> for the readout function, through which CCGNet models could realize weighted summation of atom vectors instead of simply summing, as illustrated by **Figure 2(d)**. To stabilize the learning process of self-attention and further optimize hidden embedding, we construct multi-head attention framework, which parallelly calculates  $k$  independent attention coefficients of each atom to produce  $k$  independent embeddings and then concatenate them to a vector for whole sample representation. After Global Attention, we concatenate the hidden representation  $\mathbf{U}'$  of global state with the graph embedding to further enrich the information of embedding. Finally, two dense layers are applied to the final prediction for co-crystal formation, as highlighted in grey block in **Figure 2(a)**. The details regarding the node update function, global state function, concatenation operation and readout function (including the attention mechanism) are described in the Method section.

## **2.4 Performance of CCGNet and Comparison with Control Models Coupled with Different Feature Compositions**

To assess the performance of our CCGNet model and the multilevel feature representation, we conduct a comparison study between some other representative ML algorithms coupled with different feature representations. To carry out the comparison, we construct three widely used ML models (SVM, RF and DNN) incorporated with the 12 hand-selected molecular descriptors, and two classic GNNs following model architectures proposed by Felipe et al<sup>46</sup> (named as NCI1) and Gilmer et al<sup>45</sup> (named as

enn-s2s) only coupled with the molecular graph representation. The support vector machine (SVM) and random forest (RF) perform grid search to find best hyperparameters. **Figure S1** shows architecture of deep neural network (DNN) while **Figures S2-S3** show the model frameworks of the two GNN models. As mentioned above, our CCGNet model with 3 CCGBlocks uses the multilevel feature representation from the molecular graph and the 12 molecular descriptors.

The performance of each model to the independent test set are depicted in **Table 3** while the performance to the valid set of 10-fold cross validation is listed in **Table S4**. It can be seen that the performance of the three traditional MLs coupled with the 12 molecular descriptors is significantly higher than those of the two GNN models with only the molecular graph as feature representation, in particular for the prediction accuracy of the negative samples (NACC%). The result also indicates the 12 hand-selected molecular descriptors are closely associated with the cocrystal formation and indeed are important. However, our CCGNet model based on the multilevel feature representation remarkably improve the prediction balanced accuracy, high up to 98.86%, which is significantly outperform the three traditional ML models and the two classical GNN models. It is worthy to note that the gaps between NACC and PACC in the independent test set are in the range of 6%-8% for the SVM, RF and DNN models while ones are in the range of 14%-15% for the two classical GNNs, exhibiting unsatisfactory specificity. The significant gaps in the five models should suffer from the unbalanced dataset, where the positive samples are much more than the negative ones (6.5:1). However, the gap is significantly decreased to be 1.30% by our CCGNet

coupled with the multilevel feature representation, exhibiting high specificity even in unbalanced dataset. Herein, we also perform t-distributed stochastic neighbor embedding (t-SNE) analysis<sup>48</sup> to group hidden representation from output of the readout phase (*i.e.*, Concatenate layer). As evidenced by **Figure 3**, the positive samples and negative ones can be clearly separated in space for the independent test set, confirming that the feature embedding and the attention mechanism can capture the main driving force to form cocrystal.

In fact, we also test if the prediction accuracy can be further improved when adding more feature representations associated with the H-bond and  $\pi$ - $\pi$  interactions into the edge features, considering that the two interactions are considered to dominate the process of recognition and assembly for the co-crystallization<sup>4, 49-50</sup>. As shown in **Figure 4(a)**, we characterize possible hydrogen bonds (HB) between the two co-formers by the intermolecular connection between the hydrogen bond donor and acceptor while the possible  $\pi$ - $\pi$  interaction is represented by the intermolecular connection between all aromatic atoms. We add the two interactions as two types of new edge features to the adjacency tensor for inputting to the CCGNet model. The prediction performance to the test set is also shown in **Figure 4(b)**. It can be seen that only introducing the H-bond descriptor into the multilevel features above does not further improve the prediction performance. When further adding the  $\pi$ - $\pi$  interaction feature (vide CB+HB+ $\pi$ - $\pi$  in **Figure 4(b)**), the performance is instead decreased. The reason may be attributed to that the representation of possible  $\pi$ - $\pi$  interactions significantly increases the number of edges in the model input, leading to an increase

in the model complexity, in turn making the learning of the model more difficult.

## 2.5 Analysis of the feature integration framework

The model architecture is one of the most important factors that determine the performance of the deep learning. To effectively integrate the multilevel features from the molecular descriptors as the global state and the molecular graph as the atom and the bond states, we construct the feature integration framework in the message passing phase of CCGNet, as showed in **Figure 2**. In this work, the messaging passing phase of CCGNet is composed of three CCGBlocks. CCGBlock is a core unit of CCGNet, which integrates features from different levels and passes messages as well as updates node features. As highlighted by pink and cyan dotted boxes in **Figure 5**, the message passing phase contains GCN-like and MLP-like structures. The MLP-like structure carries out the feature transformation from the global states, and the GCN-like structure performs message passing from the molecular graph. We design  $\rho^{u \rightarrow v}$  operation in each layer, through which the molecular descriptors as the global state could be effectively embedded into the atomic feature of the molecular graph. To assess the impact of the feature integration framework in the message passing phase, we comparatively construct a CCGNet-simple model through removing the MLP-like structure and the  $\rho^{u \rightarrow v}$  operation from CCGNet so that the message passing phase is only composed of three layers of GCN, namely, only containing the GCN-like structure. Consequently, CCGNet-simple does not extract features from the molecular level in the message passing phase but simply concatenates the molecular descriptors into the graph embedding after the message passing stage, as shown in **Figure S4**. The performances



of the two models to the independent test set and the valid set are shown in **Table 4** and SI **Table S4**, respectively. It is clear that the performance of CCGNet-simple is significantly lower than CCGNet, clearly exhibiting the advantage of the feature integration framework constructed in the message passing stage of CCGNet.

## 2.6 Attention Visualization and Interpretation

The interpretability of the deep learning model has been a challenge. Herein, we introduce the attention mechanism to optimize the feature space on one side, and make the model interpretable on the other side. We could map attention weights to corresponding atoms in 2D structure, in turn visualizing some important patterns hidden under the data. Figure 4 representatively selected three cocrystals to exemplify this. More examples are shown in **Figures. S5-S9**. As shown in **Figure 6(a)**, the attention weights highlight on amino groups and cyano groups of BZTCNQ (Cambridge Structural Database refcode). The crystal structure of BZTCNQ shows that there are obvious intermolecular hydrogen bonds between the two types of groups, as highlighted by cyan dash lines. In addition, the coformers of BZTCNQ includes conjugated frameworks arranged in layers, exhibiting obvious  $\pi$ - $\pi$  interaction. The attention coefficients also focus on the benzene rings. The crystal structure of EBUDEM in **Figure 6(b)** shows that there are intermolecular hydrogen bonds, for example, -N4-H...N9 and -N10-H...N8, which are also captured by the attention weights. Similarly, O7, S4, N6, and O5 atoms exhibit obvious attention weights, which are associated with the intermolecular hydrogen bonds in the real cocrystal structure. In **Figure 6(c)**, the attention weights of UXATOE02 center on several aromatic N atoms, for example, N1A,

N1, N3, N3A, N5 and N5A. These atoms also act as potential hydrogen bond acceptors. The UATOE02 crystal structure shows that the N3, N5, N3A and N5A form intermolecular hydrogen bonds with trifluoroacetic acid. These observations indicate that our model coupled with the attention mechanism could capture important factors to cocrystal formation like the H-bonding and the  $\pi$ - $\pi$  interaction, thus do not needing additional representation regarding the two interactions. It also rationalizes the result above that the addition of the edge features associated with the H-bonding and  $\pi$ - $\pi$  interaction do not further improve the prediction performance, instead decreasing it to some extent due to increasing the redundancy of features and the complexity of prediction model.

## 2.7 Out-Of-Distribution Prediction on Energetic Cocrystals (ECCs)

In order to assess the generalization ability of our model, we also collected some energetic cocrystals as dataset of out-of-distribution prediction. Energetic materials like explosives, propellants and fireworks play an important role in military and civilian fields. Cocrystal engineering serves as a potential way to design and manufacture explosives with low-sensitivity and high-energy. Compared to conventional organic molecules like pharmaceutical compounds, the energetic compounds are rich in the nitro groups, which are solitary moieties and generally offers few predictable interactions<sup>5</sup>. Thus, different from the pharmaceutical cocrystals, the hydrogen bond interactions are often not dominant for the energetic cocrystals, but the  $\pi$ - $\pi$  or the closed shell interactions like  $O\cdots O$ ,  $O\cdots N$ , etc.<sup>10, 51-55</sup>. In addition to satisfy the co-crystal screening conditions above, the energetic cocrystals under study are selected in terms

of the following conditions: 1) only containing C, H, O, N; 2) containing -NO<sub>2</sub> or -NH<sub>2</sub> but no -OH, -COOH, -CO -, -CHO, -COOC- groups. Consequently, we collected a total of 152 energetic cocrystal as the out of distribution dataset, as listed in **Table S5**. The CCGNet model is used to predict the energetic cocrystals. Similarly, the prediction accuracy is high up to 97.11%, confirming high generalization of our model.

### 3. Conclusions

In the work, we developed a novel GNN-based deep learning framework to predict formation of the cocrystal. Three key contributions are devoted: 1) a large and reliable cocrystal dataset that provides a data support for subsequent ML-based cocrystal study; 2) A multilevel strategy of the cocrystal representation by combining the molecular graph and the important molecular descriptors from the priori knowledge, which can more completely capture the main driving force to the co-crystallization; 3) A feasible GNN architecture that effectively integrate the priori knowledge into the “end-to-end” feature learning on the molecular graph. In addition, attention mechanism to further optimize the feature space and offer the model interpretation. Consequently, the performance of our model achieves 98.86% accuracy, greatly surpassing some traditional machine learning models and classic graph neural networks. Furthermore, the out-of-distribution prediction on the energetic cocrystals is also high up to 97.11% accuracy, showing strong robustness and generalization. The results from different feature compositions indicate that embedding important prior knowledges can further improve the performance of the deep learning, although modern deep learning methods generally follow an “end-to-end” design philosophy to emphasize minimization of a

priori representation or “hand-engineering”. Overall, our model provides a quick and accurate solution for the cocrystal screening, and also provides a feasible strategy for the machine learning on multi-component systems. The data and all source codes are available at <https://github.com/Saoge123/ccgnet>.

## 4. Methods

### 4.1 Co-Crystal Graph Network (CCGNet)

#### 4.1.1 Node Update Function $\Phi_v$

Graph-CNN, a spatial-based graph convolution network from Felipe et al <sup>46</sup>, is used for the message passing and node update. The Graph-CNN relies on convolutional filter  $\mathbf{H}$  to propagate and update node features.  $\mathbf{H}$  is a  $N \times N \times C$  filter tensor, which is a stack of  $N \times N$  filter matrices indexed by the node feature they filter.  $N$  is node number and  $C$  is the number of node feature.  $\mathbf{H}^{(c)}$  is defined in terms of equation (1):

$$\mathbf{H}^{(c)} = \sum_{l=1}^L h_l^{(c)} \mathbf{A}_l \quad (1)$$

$\mathbf{A}_l$  is the  $l$ -th slice of adjacency tensor  $\mathcal{A}$  whose shape is  $N \times N \times L$ .  $h_l^{(c)}$  is a scalar corresponding to a given input feature and a given slice of  $\mathbf{A}_l$ .  $L$  is the number of edge feature. The operation that filter the node feature  $\mathbf{V}_{in}$  is defined by equation (2)

$$\mathbf{V}_{out} = \sum_{c=1}^C \mathbf{H}^{(c)} \mathbf{V}_{in}^{(c)} + b \quad (2)$$

where  $\mathbf{V}_{in}^{(c)} \in \mathbb{R}^{N \times 1}$  represents the  $c$ -th node feature that is the column of  $\mathbf{V}_{in}$ .  $b$  is a scalar and  $\mathbf{V}_{out} \in \mathbb{R}^{N \times 1}$  is the result of the operation that filter the node feature  $\mathbf{V}_{in}$ .

Here, multiple filters can be set by adding another dimension to  $\mathbf{H}$  and then it becomes a tensor  $\in \mathbb{R}^{N \times N \times C \times F}$ . As a result, the output  $\mathbf{V}_{out}$  (equations (3)-(4)) also

becomes a tensor  $\in \mathbb{R}^{N \times F}$ .

$$\mathbf{V}_{out}^{(f)} = \sum_{c=1}^C \mathbf{H}^{(c,f)} \mathbf{V}_{in}^{(c)} + \mathbf{b} \quad (3)$$

$$\mathbf{V}_{out} = \parallel_{f=1}^F \mathbf{V}_{out}^{(f)} \quad (4)$$

where  $\mathbf{V}_{out}^{(f)}$  is a column of  $\mathbf{V}_{out} \in \mathbb{R}^{N \times F}$  and  $\parallel$  is concatenation. For brevity, this operation is also written as equation (5)

$$\mathbf{V}_{out} = \text{GConv}(\mathbf{V}_{in}, F) + \mathbf{b} \quad (5)$$

Finally, to consider self-loop of nodes and activation function, the convolutional operation can be described as equation (6)

$$\mathbf{V}_{out} = \sigma(\mathbf{I}\mathbf{V}_{in}\mathbf{W}_0 + \text{GConv}(\mathbf{V}_{in}, F) + \mathbf{b}) \quad (6)$$

$\sigma$  is activation function (ReLU<sup>56</sup> in this work).  $\mathbf{I}$  is a diagonal matrix that represents self-loop of nodes. Here  $\mathbf{W}_0$  is trainable weight and  $\mathbf{b} \in \mathbb{R}^F$  is bias.

#### 4.1.2 Global State Function $\Phi_u$

A single-layer feedforward neural network is used as Global State Function to perform nonlinear transformation for the global attribute of molecules. It is defined by equation (7):

$$\mathbf{u}_{out} = \sigma(\mathbf{u}\mathbf{W} + \mathbf{b}) \quad (7)$$

where  $\mathbf{u}$  is the global attribute of a molecule;  $\sigma$  is activation function (ReLU<sup>56</sup> in this work).  $\mathbf{W}$  and  $\mathbf{b}$  are trainable weight and bias, respectively.

#### 4.1.3 Concatenation operation $\rho^{u \rightarrow v}$

In CCGBlock,  $\rho^{u \rightarrow v}$  concatenates the global state (i.e., the 12 molecular descriptors) of each co-former and the node embeddings together. Cocystal input (CCGraph) can be expressed as equation (8):

$$CCGraph = \{U(u_1, u_2), A(A_1, A_2), V(V_1, V_2)\} \quad (8)$$

where the subscript refers to each co-former.  $V_1$  and  $V_2$  can be expressed as equation (9-10):

$$V_1 = (v_1^1, v_1^2, \dots, v_1^i, \dots, v_1^n) \quad (9)$$

$$V_2 = (v_2^1, v_2^2, \dots, v_2^j, \dots, v_2^m) \quad (10)$$

where the subscript refers to each co-former and the superscript denotes each atom. We perform the concatenation for every atom in terms of equations (11-12):

$$v_1^{i'} = v_1^i \oplus u_1' \quad (11)$$

$$v_2^{j'} = v_2^j \oplus u_2' \quad (12)$$

where  $\oplus$  denotes concatenation operation.

#### 4.1.4 Readout Function

Herein, we use multi-head global attention as the readout function. Following the way of human thinking, the attention mechanism uses limited attention resources to quickly screen out high-value information from a large amount of information, which has achieved remarkable performance in different tasks, for example, natural language processing<sup>57</sup>, image classification<sup>58</sup> and speech recognition<sup>59</sup>. Thus, we introduce the attention mechanism in the readout function to optimize the feature space derived from the message passing phase. Through highlighting atoms by the attention weights, we can explore how model learns the chemical structures and make the model interpretable.

Global attention calculates the attention coefficient of each node based on node features. Then the feature at the graph level is obtained by summing the product of attention coefficient and corresponding node feature, as described by equation (13)-

(14):

$$\mathbf{a} = \text{softmax}(\varphi(\mathbf{X}_{in})) \quad (13)$$

$$\mathbf{X}_{graph} = \sum_{i=1}^N a_i \mathbf{x}_i \quad (14)$$

where  $\varphi$  denote neural network (MLP in this work),  $\mathbf{a} \in \mathbb{R}^N$  is N-dimensional vector composed by attention coefficient of each node.  $\mathbf{x}_i$  represents the feature of node  $i$ , which is a row of node features  $\mathbf{X}_{in}$ .

Herein, we construct the multi-head attention into the global attention, which computes  $K$  attention coefficients of each node in parallel, yielding an attention matrix  $\boldsymbol{\alpha} \in \mathbb{R}^{N \times K}$  (vide equation (15)). Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions<sup>60</sup>.

$$\boldsymbol{\alpha} = \text{softmax}(\phi(\mathbf{X}_{in})) \quad (15)$$

where  $\phi$  denotes neural network (MLP in this work). Similar to the global attention, we calculate graph level embedding  $K$  times. As expressed by equations (16)-(17), these embeddings are concatenated to produce the final graph embedding  $\mathbf{X}_{graph}^{cat}$  that is a  $K \times C$  dimensional vector.

$$\mathbf{X}_{graph}^j = \sum_{i=1}^N \alpha_{i,j} \mathbf{x}_i \quad (16)$$

$$\mathbf{X}_{graph}^{cat} = \big\|_{j=1}^K \mathbf{X}_{graph}^j \quad (17)$$

where  $\mathbf{X}_{graph}^j$  is graph embedding calculated by using the  $j$ -th version of attention coefficients that is the  $j$ -th column of  $\boldsymbol{\alpha}$ .  $K$  is the head number and  $\alpha_{i,j}$  is an element of  $\boldsymbol{\alpha}$ .

## 4.2 Training

We randomly split 7871 samples into 10% as the independent test set, and the remaining 90% for 10-fold cross-validation. To avoid the interference of the training data, each model uses the same data partitioning for each fold. Due to the unbalance distribution of the negative and positive samples (6.5:1 ratio in the work), in terms of equations (18)-(20), we use Balanced Accuracy (BACC), Negative Accuracy (NACC) and Positive Accuracy (PACC) as the main metrics for the prediction performance of the model.

$$PACC = Sensitivity = \frac{TP}{TP + FN} \quad (18)$$

$$NACC = Specificity = \frac{TN}{FP + TN} \quad (19)$$

$$BACC = \frac{PACC + NACC}{2} \quad (20)$$

where TP is True Positive; FP is False Positive; TN is True Negative; FN is False Negative. All models are trained with Adam<sup>61</sup> optimizer with 256 samples per mini-batch.

### 4.3 Model Implementation

The CCGNet models are constructed under the opensource machine learning framework of TensorFlow<sup>62</sup>. We train the models on Nvidia RTX 2080ti GPU. SVM and RF are built by Scikit-learn<sup>63</sup>, and grid-search to find best hyper-parameters. See supporting information for details regarding constructions of DNN, NCI1 and enn-s2s. The representation of samples is implemented by RDkit, OpenBabel and CCDC Python API.

### 4.4 Data and Code Availability.

To ensure reproducibility of the results, the models and the data of this work can be



acquired at <https://github.com/Saoge123/ccgnet>.

## **Supporting Information**

Refcodes of positive samples and CIDs of negative samples collected in this work (Tables S1 and S2). Solvents involved in collecting cocrystal positive samples from Cambridge Structural Database (Table S3). Construction of several control models: DNN, NCI1, enn-s2s and CCGNet-simple (Figures S1-S4). More examples for the attention visualization (Figures S5-S9). Performances of various models with different feature compositions for the valid set of 10-fold cross validation (Table S4). Refcodes of energetic cocrystals collected from CSD for the out-of-distribution prediction (Table S4).

## **Author Information**

### **Corresponding Author**

\*Mailing address: College of Chemistry, Sichuan University, Chengdu, China. E-mail: [xmpuscu@scu.edu.cn](mailto:xmpuscu@scu.edu.cn). Phone: +86-(0)28-8541-2290.

### **Author Contributions**

X.P designed the research. Y.J performed the research. J.G and Y.L contributed to the model construction and data analysis. Y.G and M.L contributed to the interpretation of results. Y.J. and X.P. wrote the manuscript. All authors reviewed the manuscript.

## **Funding Information**

This project is supported by NSAF (Grand No. U1730127) and Key Laboratory Foundation (Grand No. 6142603190305).

## **Notes**

The authors declare no competing financial interest.

## Reference

1. Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetty, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Thaper, R. K.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J., Polymorphs, Salts, and Cocrystals: What's in a Name? *Crystal Growth & Design* **2012**, *12* (5), 2147-2152.
2. Park, S. K.; Kim, J. H.; Park, S. Y., Organic 2D Optoelectronic Crystals: Charge Transport, Emerging Functions, and Their Design Perspective. *Advanced Materials* **2018**, *30* (42).
3. Cerreia Vioglio, P.; Chierotti, M. R.; Gobetto, R., Pharmaceutical aspects of salt and cocrystal forms of APIs and characterization challenges. *Advanced Drug Delivery Reviews* **2017**, *117*, 86-110.
4. Sun, L.; Wang, Y.; Yang, F.; Zhang, X.; Hu, W., Cocrystal Engineering: A Collaborative Strategy toward Functional Materials. *Advanced Materials* **2019**, *31* (39).
5. Bolton, O.; Matzger, A. J., Improved Stability and Smart-Material Functionality Realized in an Energetic Cocrystal. *Angewandte Chemie-International Edition* **2011**, *50* (38), 8960-8963.
6. Bolton, O.; Simke, L. R.; Pagoria, P. F.; Matzger, A. J., High Power Explosive with Good Sensitivity: A 2:1 Cocrystal of CL-20:HMX. *Crystal Growth & Design* **2012**, *12* (9), 4311-4314.
7. Landenberger, K. B.; Matzger, A. J., Cocrystal Engineering of a Prototype Energetic Material Supramolecular Chemistry of 2,4,6-Trinitrotoluene. *Crystal Growth & Design* **2010**, *10* (12), 5341-5347.
8. Zhang, C.; Jiao, F.; Li, H., Crystal Engineering for Creating Low Sensitivity and Highly Energetic Materials. *Crystal Growth & Design* **2018**, *18* (10), 5713-5726.
9. Taylor, C. R.; Day, G. M., Evaluating the Energetic Driving Force for Cocrystal Formation. *Crystal Growth & Design* **2018**, *18* (2), 892-904.
10. Kent, R. V.; Wiscons, R. A.; Sharon, P.; Grinstein, D.; Frimer, A. A.; Matzger, A. J., Cocrystal Engineering of a High Nitrogen Energetic Material. *Crystal Growth & Design* **2018**, *18* (1), 219-224.
11. Barua, H.; Gunnam, A.; Yadav, B.; Nangia, A.; Shastri, N. R., An ab initio molecular dynamics method for cocrystal prediction: validation of the approach. *Crystengcomm* **2019**, *21* (47), 7233-7248.
12. Fábián, L., Cambridge Structural Database Analysis of Molecular Complementarity in Cocrystals. *Crystal Growth & Design* **2009**, *9* (3), 1436-1443.
13. Musumeci, D.; Hunter, C. A.; Prohens, R.; Scuderi, S.; McCabe, J. F., Virtual

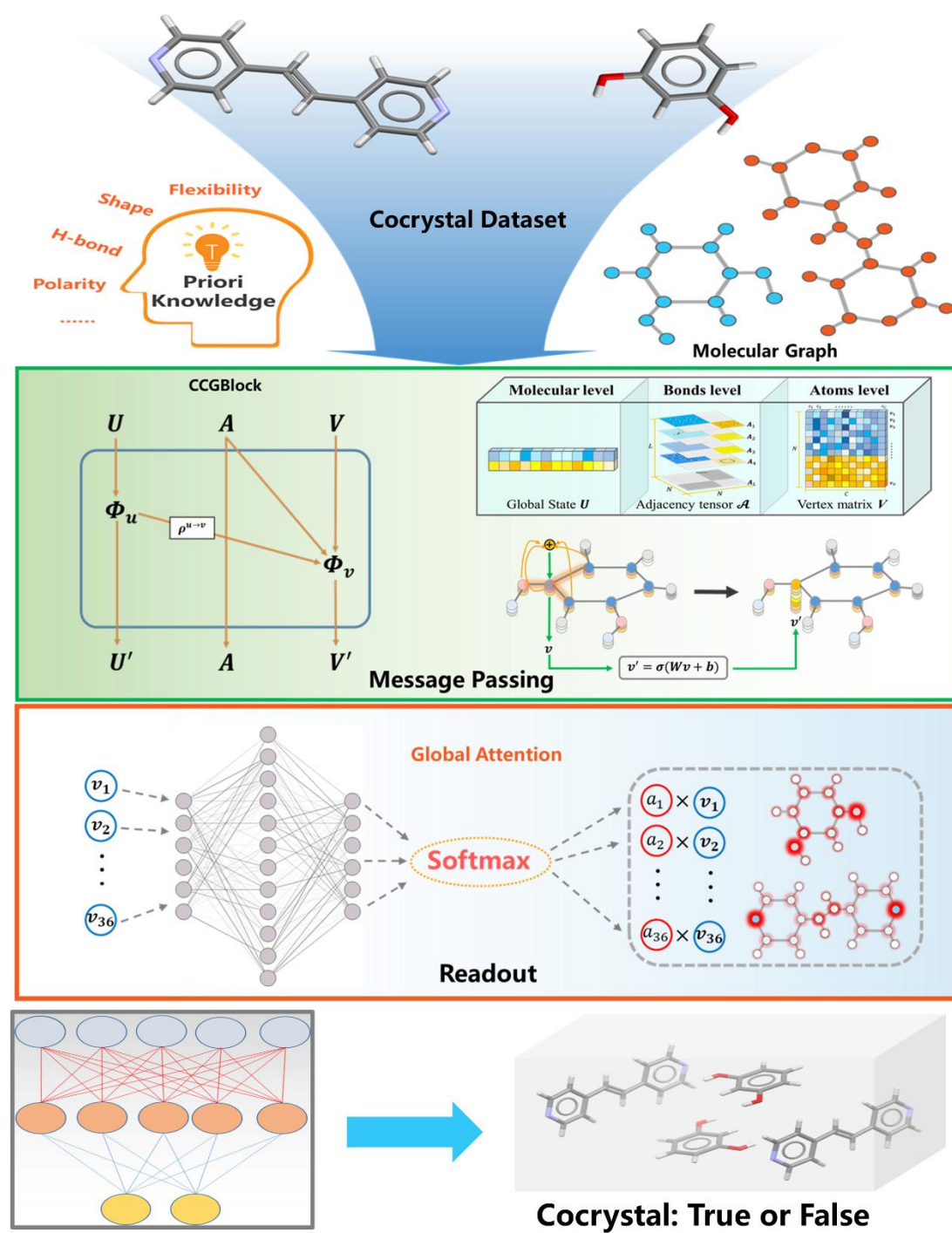
- cocrystal screening. *Chemical Science* **2011**, 2 (5), 883-890.
14. Klamt, A., The COSMO and COSMO-RS solvation models. *Wiley Interdisciplinary Reviews Computational Molecular Science* **2011**, 1 (5), 699-709.
15. Sun, G.; Jin, Y.; Li, S.; Yang, Z.; Shi, B.; Chang, C.; Abramov, Y. A., Virtual Coformer Screening by Crystal Structure Predictions: Crucial Role of Crystallinity in Pharmaceutical Cocrystallization. *The Journal of Physical Chemistry Letters* **2020**, 11 (20), 8832-8838.
16. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., Machine learning for molecular and materials science. *Nature* **2018**, 559 (7715), 547-555.
17. Jordan, M. I.; Mitchell, T. M., Machine learning: Trends, perspectives, and prospects. *Science* **2015**, 349 (6245), 255.
18. Rama Krishna, G.; Ukrainczyk, M.; Zeglinski, J.; Rasmuson, Å. C., Prediction of Solid State Properties of Cocrystals Using Artificial Neural Network Modeling. *Crystal Growth & Design* **2018**, 18 (1), 133-144.
19. Wicker, J. G. P.; Crowley, L. M.; Robshaw, O.; Little, E. J.; Stokes, S. P.; Cooper, R. I.; Lawrence, S. E., Will they co-crystallize? *CrystEngComm* **2017**, 19 (36), 5336-5340.
20. Przybyłek, M.; Jeliński, T.; Słabuszewska, J.; Ziółkowska, D.; Mroczyńska, K.; Cysewski, P., Application of Multivariate Adaptive Regression Splines (MARSplines) Methodology for Screening of Dicarboxylic Acid Cocrystal Using 1D and 2D Molecular Descriptors. *Crystal Growth & Design* **2019**, 19 (7), 3876-3887.
21. Krizhevsky, A.; Sutskever, I.; Hinton, G. E., ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, Curran Associates Inc.: Lake Tahoe, Nevada, 2012; pp 1097-1105.
22. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J.; Ieee, Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 Ieee Conference on Computer Vision and Pattern Recognition*, 2014; pp 580-587.
23. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y., Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. *Computer Science* **2014**.
24. Brown, T. B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A., Language Models are Few-Shot Learners. **2020**.
25. Silver, D.; Huang, A.; Maddison, C. J.; Guez, A.; Sifre, L.; van den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; Dieleman, S.; Grewe, D.; Nham, J.; Kalchbrenner, N.; Sutskever, I.; Lillicrap, T.; Leach, M.; Kavukcuoglu, K.; Graepel, T.; Hassabis, D., Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, 529 (7587), 484-489.
26. Vinyals, O.; Babuschkin, I.; Czarnecki, W. M.; Mathieu, M.; Dudzik, A.; Chung, J.; Choi, D. H.; Powell, R.; Ewalds, T.; Georgiev, P.; Oh, J.; Horgan, D.; Kroiss, M.; Danihelka, I.; Huang, A.; Sifre, L.; Cai, T.; Agapiou, J. P.; Jaderberg, M.; Vezhnevets, A. S.; Leblond, R.; Pohlen, T.; Dalibard, V.; Budden, D.; Sulsky, Y.; Molloy, J.; Paine, T. L.; Gulcehre, C.; Wang, Z.; Pfaff, T.; Wu, Y.; Ring, R.; Yogatama, D.; Wunsch, D.;

- McKinney, K.; Smith, O.; Schaul, T.; Lillicrap, T.; Kavukcuoglu, K.; Hassabis, D.; Apps, C.; Silver, D., Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* **2019**, *575* (7782), 350- 354.
27. Zheng, S.; Li, Y.; Chen, S.; Xu, J.; Yang, Y., Predicting drug-protein interaction using quasi-visual question answering system (Feb, 10.1038/s42256-020-0152-y, 2020). *Nature Machine Intelligence* **2020**, *2* (9), 551-551.
28. Segler, M. H. S.; Preuss, M.; Waller, M. P., Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **2018**, *555* (7698), 604-610.
29. Wang, D.; Wei, S.; Yuan, A.; Tian, F.; Cao, K.; Zhao, Q.; Zhang, Y.; Zhou, C.; Song, X.; Xue, D.; Yang, S., Machine Learning Magnetic Parameters from Spin Configurations. *Advanced Science* **2020**, *7* (16).
30. Ma, W.; Cheng, F.; Xu, Y.; Wen, Q.; Liu, Y., Probabilistic Representation and Inverse Design of Metamaterials Based on a Deep Generative Model with Semi-Supervised Learning Strategy. *Advanced Materials* **2019**, *31* (35).
31. Wang, H. S.; Ji, Y. J.; Li, Y. Y., Simulation and design of energy materials accelerated by machine learning. *Wiley Interdiscip. Rev.-Comput. Mol. Sci.* **2020**, *10* (1), 18.
32. Zhou, J.; Cui, G.; Zhang, Z.; Yang, C.; Liu, Z.; Wang, L.; Li, C.; Sun, M., Graph neural networks: A review of methods and applications. *arXiv preprint arXiv:1812.08434* **2018**.
33. Mater, A. C.; Coote, M. L., Deep Learning in Chemistry. *Journal of Chemical Information and Modeling* **2019**, *59* (6), 2545-2559.
34. Grambow, C. A.; Pattanaik, L.; Green, W. H., Deep Learning of Activation Energies. *Journal of Physical Chemistry Letters* **2020**, *11* (8), 2992-2997.
35. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P., Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chemistry of Materials* **2019**, *31* (9), 3564-3572.
36. Roszak, R.; Beker, W.; Molga, K.; Grzybowski, B. A., Rapid and Accurate Prediction of pK(a) Values of C-H Acids Using Graph Convolutional Neural Networks. *Journal of the American Chemical Society* **2019**, *141* (43), 17142-17149.
37. St. John, P. C.; Guan, Y.; Kim, Y.; Kim, S.; Paton, R. S., Prediction of organic homolytic bond dissociation enthalpies at near chemical accuracy with sub-second computational cost. *Nature Communications* **2020**, *11* (1), 2328.
38. Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C., The Cambridge Structural Database. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **2016**, *72* (Pt 2), 171-179.
39. Grothe, E.; Meekes, H.; Vlieg, E.; ter Horst, J. H.; de Gelder, R., Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System. *Crystal Growth & Design* **2016**, *16* (6), 3237-3243.
40. Sun, L. J.; Zhu, W. G.; Yang, F. X.; Li, B. L.; Ren, X. C.; Zhang, X. T.; Hu, W. P., Molecular cocrystals: design, charge-transfer and optoelectronic functionality. *Physical Chemistry Chemical Physics* **2018**, *20* (9), 6009-6023.
41. Stoler, E.; Warner, J. C., Non-covalent derivatives: cocrystals and eutectics. *Molecules* **2015**, *20* (8), 14833-14848.
42. Safont-Sempere, M. M.; Fernández, G.; Würthner, F., Self-Sorting Phenomena in

- Complex Supramolecular Systems. *Chemical Reviews* **2011**, *111* (9), 5784-5814.
43. Wu, A.; Isaacs, L., Self-Sorting: The Exception or the Rule? *Journal of the American Chemical Society* **2003**, *125* (16), 4831-4835.
44. Battaglia, P. W.; Hamrick, J. B.; Bapst, V.; Sanchez-Gonzalez, A.; Zambaldi, V.; Malinowski, M.; Tacchetti, A.; Raposo, D.; Santoro, A.; Faulkner, R., Relational inductive biases, deep learning, and graph networks. *arXiv preprint arXiv:1806.01261* **2018**.
45. Gilmer, J.; Schoenholz, S. S.; Riley, P. F.; Vinyals, O.; Dahl, G. E., Neural message passing for quantum chemistry. *arXiv preprint arXiv:1704.01212* **2017**.
46. Such, F. P.; Sah, S.; Dominguez, M. A.; Pillai, S.; Zhang, C.; Michael, A.; Cahill, N. D.; Ptucha, R., Robust Spatial Filtering With Graph Convolutional Neural Networks. *IEEE J. Sel. Top. Signal Process.* **2017**, *11* (6), 884-896.
47. Li, Y.; Tarlow, D.; Brockschmidt, M.; Zemel, R., Gated graph sequence neural networks. *arXiv*, 2015. *arXiv preprint arXiv:1511.05493* **2015**.
48. Maaten, L. v. d.; Hinton, G., Visualizing data using t-SNE. *Journal of machine learning research* **2008**, *9* (Nov), 2579-2605.
49. Gavezzotti, A.; Colombo, V.; Lo Presti, L., Facts and factors in the formation and stability of binary crystals. *Crystal Growth & Design* **2016**, *16* (10), 6095-6104.
50. Prins, L. J.; Reinhoudt, D. N.; Timmerman, P., Noncovalent synthesis using hydrogen bonding. *Angewandte Chemie International Edition* **2001**, *40* (13), 2382-2426.
51. Landenberger, K. B.; Matzger, A. J., Cocrystals of 1, 3, 5, 7-Tetranitro-1, 3, 5, 7-tetrazacyclooctane (HMX). *Crystal growth & design* **2012**, *12* (7), 3603-3609.
52. Ma, Q.; Jiang, T.; Chi, Y.; Chen, Y.; Wang, J.; Huang, J.; Nie, F., A novel multi-nitrogen 2, 4, 6, 8, 10, 12-hexanitrohexaazaisowurtzitane-based energetic co-crystal with 1-methyl-3, 4, 5-trinitropyrazole as a donor: experimental and theoretical investigations of intermolecular interactions. *New Journal of Chemistry* **2017**, *41* (10), 4165-4172.
53. Bolton, O.; Simke, L. R.; Pagoria, P. F.; Matzger, A. J., High power explosive with good sensitivity: a 2: 1 cocrystal of CL-20: HMX. *Crystal Growth & Design* **2012**, *12* (9), 4311-4314.
54. Liu, G.; Li, H.; Gou, R.; Zhang, C., Packing Structures of CL-20-Based Cocrystals. *Crystal Growth & Design* **2018**, *18* (11), 7065-7078.
55. Bu, R.; Xiong, Y.; Zhang, C., pi-pi Stacking Contributing to the Low or Reduced Impact Sensitivity of Energetic Materials. *Crystal Growth & Design* **2020**, *20* (5), 2824-2841.
56. Nair, V.; Hinton, G. E. In *Rectified linear units improve restricted boltzmann machines*, ICML, 2010.
57. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I., Attention Is All You Need. In *Advances in Neural Information Processing Systems 30*, Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; Garnett, R., Eds. Neural Information Processing Systems (Nips): La Jolla, 2017; Vol. 30.
58. Hu, J.; Shen, L.; Albanie, S.; Sun, G.; Wu, E., Squeeze-and-Excitation Networks.

- IEEE Transactions on Pattern Analysis and Machine Intelligence* **2017**, PP (99).
59. Chorowski, J.; Bahdanau, D.; Cho, K.; Bengio, Y., End-to-end Continuous Speech Recognition using Attention-based Recurrent NN: First Results. *Eprint Arxiv* **2014**.
60. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I., Attention Is All You Need. *arXiv* **2017**.
61. Kingma, D. P.; Ba, J. A., A method for stochastic optimization. arXiv 2014. *arXiv preprint arXiv:1412.6980* **2019**, 434.
62. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devin, M.; Ghemawat, S.; Irving, G.; Isard, M.; Kudlur, M.; Levenberg, J.; Monga, R.; Moore, S.; Murray, D. G.; Steiner, B.; Tucker, P.; Vasudevan, V.; Warden, P.; Wicke, M.; Yu, Y.; Zheng, X.; Assoc, U., *TensorFlow: A system for large-scale machine learning*. 2016; p 265-283.
63. Swami, A.; Jain, R., Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **2013**, 12 (10), 2825-2830.

## Graphical Abstract



## TABLES

**Table 1.** Atomic and Bond Attributes Used in CCGNet

Feature	Description
<b>Atom</b>	
Atom type	Cl, N, P, Br, B, S, I, F, C, O, H (one-hot)
Hybridization	SP2, SP3, SP, S (one-hot)
Chirality	None, R, S (binary)
is_chiral	True or False (binary)
is_spiro	True or False (binary)
is_cyclic	True or False (binary)
is_aromatic	True or False (binary)
is_acceptor	True or False (binary)
is_donor	True or False (binary)
Explicitvalence	integer
implicitvalence	integer
Formal charge	integer
Degree	integer
Total H number	integer
Vdw radius	float
atomic_number	integer
<b>Bond</b>	
bond type	Single, Double, Triple, Aromatic (one-hot)



**Table 2.** Molecular Descriptors used as Global States in CCGNet

Molecular Descriptor	Description
S	short axis of an enclosing box (float)
S_L	S / long axis of an enclosing box (float)
S_M	S / medium axis of an enclosing box (float)
M_L	medium axis of an enclosing box / long axis of an enclosing box (float)
Globularity	surface of a sphere with the same volume as the molecule / area (float)
FrTPSA	TPSA / SASA (float)
Fr_NO	$(n_N + n_O) / n_{\text{heavy}}$ (float)
Fr_AromaticAtoms	$n_{\text{AromaticAtom}} / n_{\text{heavy}}$ (float)
HBA	the number of H-bond acceptor (integer)
HBD	the number of H-bond donor (integer)
RBN	the number of rotatable bond (integer)
Dipole_Moment	Dipole moment (float)

**Table 3.** Performances of Various models coupled with different feature compositions for the independent test set <sup>a</sup>.

<b>Model</b>	<b>PACC (%)</b>	<b>NACC (%)</b>	<b>BACC (%)</b>
<b>SVM</b>	98.98( $\pm 0.27$ )	90.27( $\pm 0.84$ )	94.62( $\pm 0.42$ )
<b>RF</b>	<b>99.99(<math>\pm 0.04</math>)</b>	93.04 ( $\pm 0.96$ )	96.51 ( $\pm 0.47$ )
<b>DNN</b>	99.15( $\pm 0.35$ )	91.07( $\pm 2.29$ )	95.11( $\pm 1.06$ )
<b>NCI1</b>	98.65( $\pm 0.30$ )	83.39( $\pm 3.02$ )	91.02( $\pm 1.57$ )
<b>enn-s2s</b>	98.08( $\pm 0.45$ )	85.45( $\pm 2.85$ )	91.77( $\pm 1.50$ )
<b>CCGNet</b>	99.51( $\pm 0.40$ )	<b>98.21(<math>\pm 0.98</math>)</b>	<b>98.86(<math>\pm 0.53</math>)</b>

<sup>a</sup> PACC is the prediction accuracy for the positive samples while NACC denotes the accuracy for the negative samples. BACC represents the balanced accuracy involving in the positive and negative samples.

**Table 4.** A comparison of performance between CCGNet-simple and CCGNet frameworks.

<b>Model</b>	<b>PACC(%)</b>	<b>NACC(%)</b>	<b>BACC(%)</b>
<b>CCGNet-simple</b>	99.16 ( $\pm 0.48$ )	94.55( $\pm 3.65$ )	96.86( $\pm 1.69$ )
<b>CCGNet</b>	<b>99.51(<math>\pm 0.10</math>)</b>	<b>98.21(<math>\pm 0.98</math>)</b>	<b>98.86(<math>\pm 0.53</math>)</b>