

Design of New Inorganic Crystals with the Desired Composition Using Deep Learning

Seunghee Han, Jaewan Lee, Sehui Han, Seyed Mohamad Moosavi, Jihan Kim, and Changyoung Park*



Cite This: *J. Chem. Inf. Model.* 2023, 63, 5755–5763



Read Online

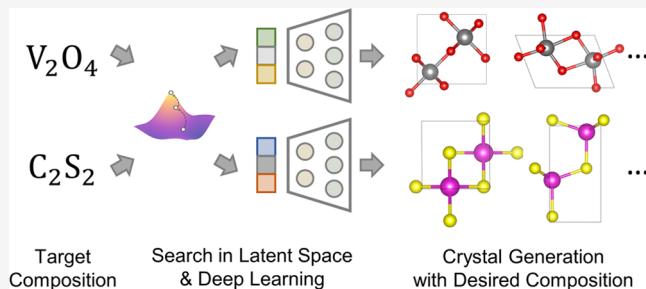
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: New solid-state materials have been discovered using various approaches from atom substitution in density functional theory (DFT) to generative models in machine learning. Recently, generative models have shown promising performance in finding new materials. Crystal generation with deep learning has been applied in various methods to discover new crystals. However, most generative models can only be applied to materials with specific elements or generate structures with random compositions. In this work, we developed a model that can generate crystals with desired compositions based on a crystal diffusion variational autoencoder. We generated crystal structures for 14 compositions of three types of materials in different applications. The generated structures were further stabilized using DFT calculations. We found the most stable structures in the existing database for all but one composition, even though eight compositions among them were not in the data set trained in a crystal diffusion variational autoencoder. This substantiates the prospect of the generation of an extensive range of compositions. Finally, 205 unique new crystal materials with energy above hull <100 meV/atom were generated. Moreover, we compared the average formation energy of the crystals generated from five compositions, two of which were hypothetical, with that of traditional methods like atom substitution and a generative model. The generated structures had lower formation energy than those of other models, except for one composition. These results demonstrate that our approach can be applied stably in various fields to design stable inorganic materials based on machine learning.



INTRODUCTION

The demand for exploring new inorganic crystals has been increasing due to their potential use in applications such as batteries and catalysts.¹ To discover new materials, scientists have traditionally employed a trial and error approach, repeatedly synthesizing and testing to discover new crystal candidates. However, experimental material research is restricted by time and cost considerations.^{2,3} An alternative method is to use computational simulations, such as density functional theory (DFT)⁴ calculations using the atom substitution method from scientific domain knowledge.⁵ With the aid of extensive computational work, many databases for inorganic crystals, such as inorganic crystal structure database (ICSD),⁶ open quantum materials database (OQMD),⁷ and materials project (MP),⁸ have been developed in the past decade, which has allowed screening of thousands of materials.⁹ Nevertheless, the potential combinations of atoms that can form crystals are vast, and most of the materials space has yet to be explored.¹⁰

Recently, the use of deep learning for material generation and design has been recognized as a potential avenue for discovering new materials.¹¹ Deep generative models, such as generative adversarial network (GAN)¹² and variational autoencoder (VAE),¹³ have been employed to generate inorganic crystal systems with compositions from specific

elements (e.g., perovskite and VO).^{3,14,15} In addition to these cases, deep learning models that generate universal crystal structures have also been developed beyond the limitations of specific materials. These models can generate any combination of atoms from the periodic table generally in a single generation. For example, Ren et al.¹⁶ developed Fourier transform crystal properties (FTCP) framework by introducing a variational autoencoder (VAE), and Xie et al.¹⁷ developed a crystal diffusion variational autoencoder (CDVAE). CDVAE was developed based on a diffusion model, which has been widely applied not only in the generation of fine images recently such as DALL-E-2,¹⁸ Imagen,¹⁹ and Stable Diffusion²⁰ but also generation of molecules²¹ and simulations on molecular docking.²² With the CDVAE, crystal structures can be searched from a latent space where the information, such as atom types (A), atom coordinates (X), and lattice parameters (L), are compressed in and generated through the diffusion

Received: June 20, 2023

Published: September 8, 2023



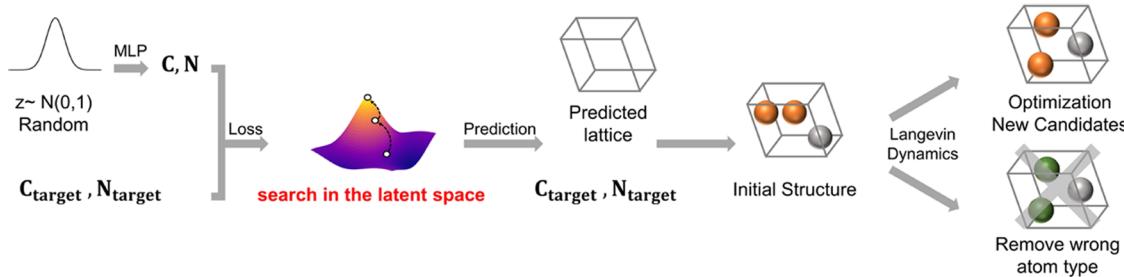


Figure 1. Overview of crystal generation with the desired composition. Each different colored sphere represents an atom.

process. Recent research utilizing this model has successfully achieved the generation of two-dimensional (2D) materials²³ and one-dimensional (1D) materials.²⁴

These models, however, have difficulty in obtaining the crystals with specific user-desired compositions as they generate crystals with random compositions based on target properties only. The generation and evaluation of materials with the desired compositions play a crucial role in the search for experimentally synthesizable materials. The search space could become overwhelmingly vast without such a condition, considering the numerous possible combinations of elements. There have been scientific papers dedicated to exploring synthesizable materials by imposing constraints on composition. Collins et al. experimentally discovered two crystal structure types by finding stable materials through computational methods for specific inorganic phase fields.²⁵ Gu et al. identified synthesizable compositions on perovskite through a graph neural network.²⁶

In this work, we could successfully find stable crystals with the desired compositions using our algorithm, which explores a latent space with the gradient descent method,^{17,27} and generates relaxed structures by the CDVAE model. Our approach for the generation was analyzed concerning versatility and stability. We generated various crystal systems of V_xO_y (catalyst), $Li_xMn_xO_z$ (battery), and Cd_xS_y (semiconductor), and compared the distribution and relationship with the existing database. We computed the formation energy through DFT calculation and found that our approach generated 205 nonoverlapping and stable crystal structures with promising prospects for material search. These results demonstrate the potential of deep learning generation models for discovering new inorganic crystals with the desired compositions.

RESULTS AND DISCUSSION

Search in Latent Space. A crystal generation model CDVAE, which was a variational autoencoder for periodic structures, was used to generate crystals with various targeted compositions. In the original generation of CDVAE, the initial material was formed by predicting the lattice parameter (L), composition (c), and number of atoms (n) from a latent space with a normal distribution, and the atom type and coordinates were adjusted and optimized through Langevin Dynamics.²⁸ For this reason, a diverse range of materials with random compositions was created at the end.

We successfully achieved the generation of inorganic crystals with the desired composition and a specific number of atoms by constraining and exploring the latent space in the process. The specific algorithm for searching the optimized latent space was followed as seen in Figure 1. The target material in the latent space z_{opt} was searched with the desired composition

(c_{target}) and the number of atoms (n_{target}) using gradient descent optimization algorithms²⁷ with Cross Entropy Loss.

The structures generated from z_{opt} were optimized by changing the atom types and positions within the predicted unit cell through the Langevin Dynamics²⁸ process of CDVAE, taking 7–8 seconds per structure. In this process, the atom types could change and therefore the structures with atom types different from the target were removed. The discussion about reconstruction with atom types can be found in Table S1. As a result, only structures with the desired composition were successfully obtained. We compared these structures with the existing database (e.g., MP database) using the Structure Matcher of Pymatgen.²⁹ In addition, the final candidates were filtered out by removing the structures that matched among the generated structures (removing duplicates).

Design of Materials with a Targeted Composition. We explored various material spaces: V_xO_y as a catalyst, $Li_xMn_xO_z$ as a battery material, and Cd_xS_y as a semiconductor material. The results of generation for each system by our algorithm are summarized in Table 1. The numbers were expressed in the chemical formula (e.g., V_2O_4) in this study as generation differs according to the number of atoms as well as composition. Our model can generate a variety of materials and their compositions, not just one composition or material. Some compositions generated by the model were identical to the ones in the MP database as can be seen in Table 1. In addition, the model was able to generate crystals with compositions, not in the training data set of CDVAE (e.g., Cd_xS_y , VO_2 , V_2O_4 , V_2O_5) and with the compositions (e.g., V_3O_4 , V_5O_6) in another database (e.g., iMatGen) (Table S2). In particular, some of the generated structures were matched with structures in the iMatGen database (Table 1). This represents the versatility and reliability of the model. However, the generation rate of the model gradually decreased as the number of atoms increased (Figure S1). This was due to the increased probability that the atom type would change according to the increased number of atoms. New candidates can be filtered out by removing duplicated structures within the ones generated to compute DFT optimization for unique structures. These candidates are shown in bold in Table 1. This was a promising result considering that the number of materials for each composition in the MP structure was less than 10 except for $Li_5Mn_3O_8$ (Table S2).

The DFT calculations were performed on these structures (the method for DFT calculation was detailed in the Methods section). After optimization with DFT, the stabilized materials and the materials from the training data set of CDVAE were visualized using principal component analysis (PCA),³⁰ which represents the latent space in two dimensions (Figure 2). For clear visualization, sampling was done for each composition at a ratio of 0.5. The latent space was not clearly distinguished for

Table 1. Generation Results from Our Algorithms^a

system	composition	numbers (<i>n</i>)	generation from latent space	generated target materials	matched	new candidates
VO system (MP)	VO ₂	3	1024	221	14 (3)	60
	V ₂ O ₄	6	5120	356	16 (1)	137
	V ₂ O ₅	7	10,240	377	0	150
	V ₄ O ₁₀	14	10,240	102	0	84
	V ₆ O ₁₀	16	10,240	99	1 (1)	86
VO (iMatGen)	V ₃ O ₄	7	5120	119	9 (8)	46
	V ₅ O ₆	11	5120	10	1 (1)	6
LiMnO system (MP)	Li ₂ Mn ₂ O ₄	8	5120	379	37 (3)	142
	Li ₄ Mn ₂ O ₆	12	5120	91	0	66
	LiMn ₄ O ₈	13	5120	56	2 (2)	45
	Li ₅ Mn ₃ O ₈	16	5120	28	1 (1)	25
CdS system (MP)	Cd ₂ S ₂	4	5120	380	9 (2)	89
	Cd ₂ S	3	5120	163	64 (1)	64
	Cd ₄ S ₈	12	5120	1	0	1

^aThe “Generated target materials” column means the number of structures in which wrong atom type structures were removed among the generated structures. The “Matched” column is the number of structures matched with the database. Parentheses indicate unique numbers in the database. The “New Candidates” column represents the number of structures filtered out by removing duplicates and is marked in bold.

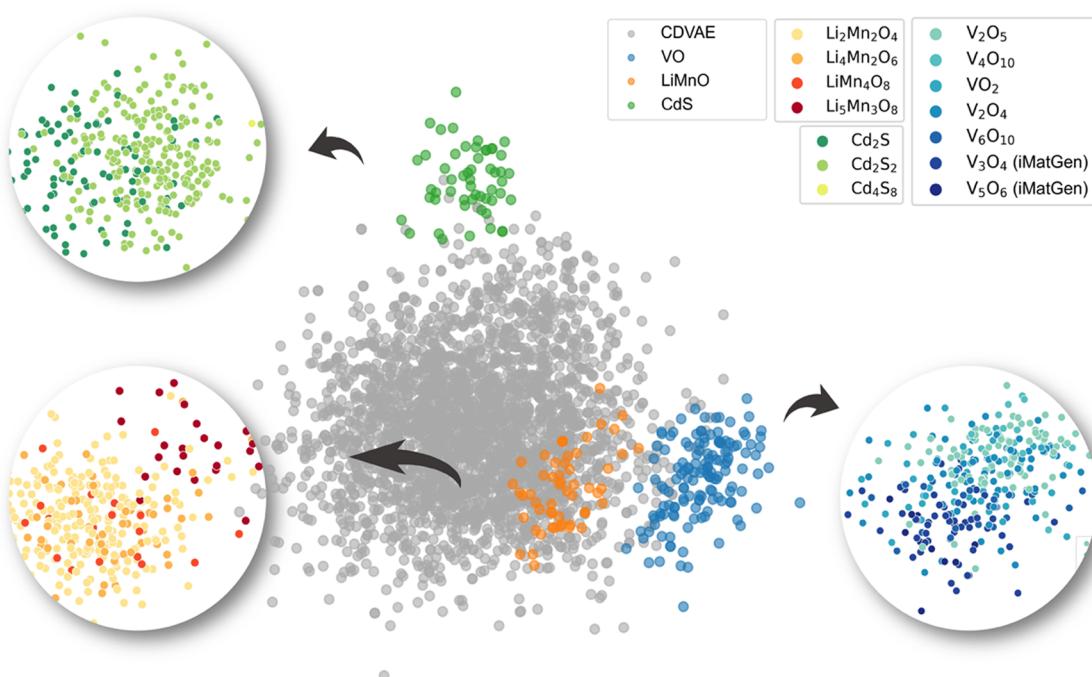


Figure 2. Visualization of latent space through principal component analysis (PCA). The materials of V_xO_y (blue), Li_xMn_yO_z (orange), Cd_xS_y (green), and materials in the training data set of CDVAE (gray) were encoded and expressed in latent space. The structures in each system were presented in circles with similar colors according to the composition.

each composition, but the results were noteworthy in the sense that each of the systems was clearly distinguished. The system of LiMnO exhibited complete overlap with the CDVAE data set in latent space. The system of VO slightly overlapped the CDVAE data set, and the system of CdS hardly overlapped with the CDVAE data set. It can be inferred that the LiMnO system was all included in the training data set of CDVAE, whereas some compositions of the VO and CdS systems were not. Our algorithm makes it possible to generate different and versatile new materials with new compositions, not in the CDVAE data set. This shows the potential to explore a wider material space, which shows a similar result to the previous paper.²³

Structure Stability Analysis through DFT Calculation.

The DFT calculation was conducted for each composition in each system. The results of each formation energy and energy above hull (E_{hull}) for 12 compositions can be found in Figures S2 and 3a–c, respectively. The formation energy results for V₃O₄ and V₅O₆ is covered in the next section for comparison with other models. The distribution graph for energy above hull (E_{hull}) for all generated structures is shown in Figure 3d. Out of the total 949 structures, 703 structures (approx. 74%) exhibited $E_{\text{hull}} < 200$ meV/atom. Furthermore, there were 379 structures with $E_{\text{hull}} < 100$ meV/atom and 213 structures with $E_{\text{hull}} < 50$ meV/atom. It can be inferred that many stable structures were generated well.

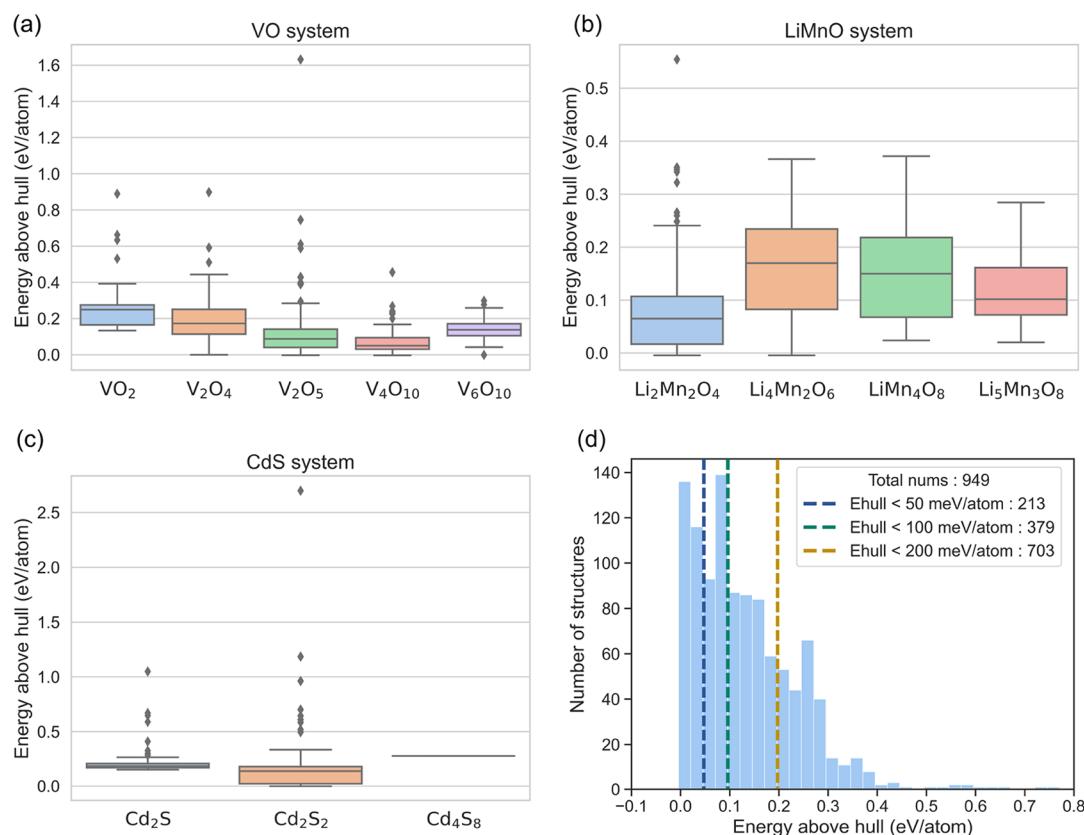


Figure 3. Plot of energy above hull for each structure through DFT calculation. Box plot of formation energy for the structures in (a) VO system, (b) LiMnO system, and (c) CdS system. (d) Histogram for the energy above hull (E_{hull}) of all generated structures in (a–c). Energies above 0.8 eV were excluded to facilitate clarity in viewing.

Additionally, we conducted recalculations of the structures with the lowest energy for each composition in the MP database to compare with our results, which can be found in Tables S3–S5. The lowest formation energies of the MP database were found for all five compositions in the VO system. In particular, the V_2O_5 structure with a lower formation energy than the MP data set was found in the case of the same composition and the same number of atoms. Unfortunately, V_2O_4 structures had three unstable structures, while V_2O_5 structures had one unstable structure. In Figure 3a, these three structures were excluded. All structures of the LiMnO system were stable, including the lowest formation energy of the MP database. In addition, we found a $\text{Li}_4\text{Mn}_2\text{O}_6$ structure with a lower formation energy than the MP database in the same condition. The CdS system had 2 errors for Cd_2S , and the composition of Cd_2S_2 encompassed the structure with the lowest formation energy of the MP database. Our model generated a more stable structure of Cd_2S than the most stable structure in the MP database. However, only one structure was obtained for Cd_4S_8 after Langevin Dynamics.

While not all of the compositions had structures with $E_{\text{hull}} = 0$, it is noteworthy that the lowest formation energies were observed in the MP database for all compositions, except for Cd_4S_8 (Figure S2d). Notably, for the three compositions (V_2O_5 , $\text{Li}_4\text{Mn}_2\text{O}_6$, and Cd_2S), we identified structures with lower formation energy than the structure found in MP, as depicted in Figure 4. The V_2O_5 structure in the MP database was layered, but the V_2O_5 structure generated was not. This is due to a coordination difference in vanadium (V). The coordination number of V in V_2O_5 (MP) is 4, while that in

V_2O_5 (gen) it is 5. While Mn is uniformly distributed in $\text{Li}_4\text{Mn}_2\text{O}_6$ of MP, Mn and Li were separately distributed in the generated structure. In both V_2O_5 and $\text{Li}_4\text{Mn}_2\text{O}_6$, the space group number of the MP structure was 12, and the space group number of the generated structure was 1. The composition of Cd_2S also had a difference in coordination number and space group. The coordination number of S was 4 in the MP database, but 6 in our generated structure. Additionally, the space group number of the MP structure was 156 and the space group number of the generated structure was 8.

The process of selecting the final structures was carried out. The structures that matched the MP database were removed from the candidates, and then the structures that matched each other were removed to get unique structures. Furthermore, the structures with energy above the hull greater than 100 meV/atom were eliminated. Finally, a total of 205 structures with $E_{\text{hull}} < 100$ meV/atom and 104 structures with $E_{\text{hull}} < 50$ meV/atom were screened, and the exact number for each structure is shown in Table 2. The structures in Figure 4 are included in the final structure. The structures of VO_2 , Cd_2S , and Cd_4S_8 were removed because the energy above the hull was greater than 100 meV/atom. The most stable structures for each composition are shown in Figure S3.

Stability Comparison with Other Models. Next, we compared our algorithms with other existing methods concerning the formation energy. In Figure 5, each graph was drawn using a normalized histogram to describe the distribution of the formation energy. First, the distribution for atom substitution was analyzed for a specific composition in each system, compared with that of our work (Figure 5a–c).

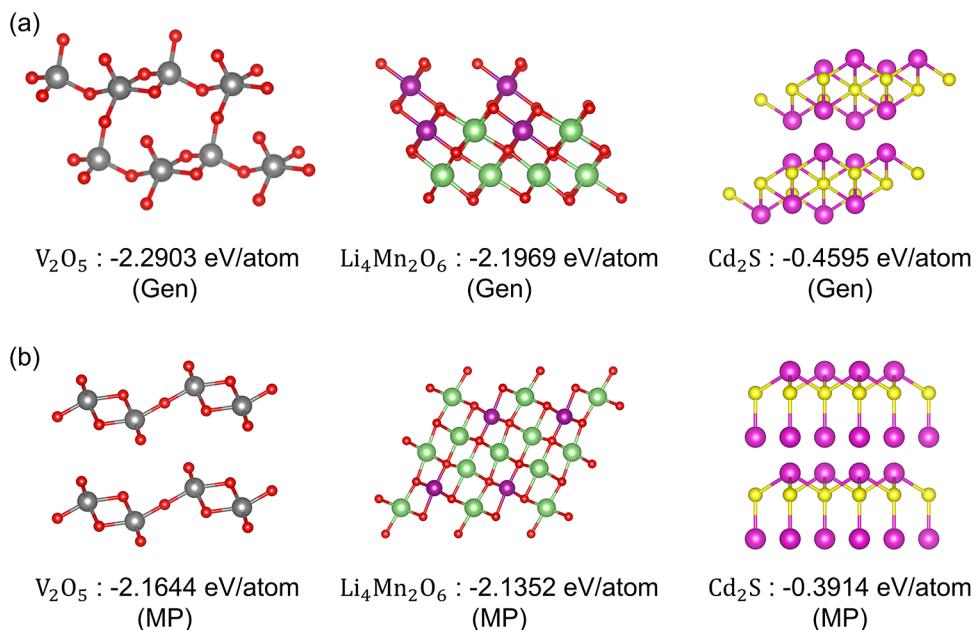


Figure 4. Crystal structures with a formation energy of 0.01 eV lower than those in MP structures. Structures were described with increased unit cell size for clear visualization (red: O; gray: V; purple: Mn; green: Li; yellow: S; and pink: Cd).

Table 2. Number of Structures Computed by DFT and New Materials Finally Filtered Out^a

		DFT calculation	new materials $E_{\text{hull}} < 100$ meV/atom	new materials $E_{\text{hull}} < 50$ meV/atom
VO	VO_2	60	0	0
	V_2O_4	137	25	12
	V_2O_5	150	32	11
	V_4O_{10}	84	55	31
	V_6O_{10}	86	19	4
LiMnO	$Li_2Mn_2O_4$	142	29	19
	$Li_4Mn_2O_6$	66	9	2
	$LiMn_4O_8$	45	15	8
	$Li_5Mn_3O_8$	25	8	4
CdS	Cd_2S_2	89	13	13
	Cd_2S	64	0	0
	Cd_4S_8	1	0	0
sum			205	104

^a(The structures that matched both the MP database and each other, as well as those with energy above hull greater than 100 meV/atom were removed.) The new materials consist of unique structures that are not in the MP data set.

In the LiMnO system, atom substitution was carried out by replacing 150 $X_2Y_2Z_4$ structures with $Li_2Mn_2O_4$. The DFT calculation was performed under the same conditions for each substituted structure. As a result, 8 out of 150 structures had errors, and only the remaining 142 structures were relaxed. In the same way, 477 V_2O_4 and 232 Cd_2S_2 were computed, resulting in 440 optimized structures (37 errors) and 228 optimized structures (4 errors), respectively. The structures we generated were plotted for 142 $Li_2Mn_2O_4$, 137 V_2O_4 , and 89 Cd_2S_2 .

Next, the formation energy was compared with another model (iMatGen¹⁴) other than atom substitution (Figure 5d,e). For V_3O_4 , the DFT calculation was performed for 57 structures with atom substitution, 112 structures of iMatGen, and 119 structures of our model. In atom substitution, 9 structures were excluded due to errors. For V_5O_6 , 11 structures with atom substitution, 42 structures from iMatGen, and 10 structures from our model were optimized.

For all of the compositions, the structures generated with our model had lower formation energy on averages compared to the structures generated with iMatGen or with atom substitution except for V_3O_4 . For V_3O_4 , iMatGen produced the structures with a lower average formation energy. The structures from atom substitution have relatively high formation energies on average for all compositions or cause errors during structure optimization. It can be deduced that experimental intuition or domain-related knowledge would be essential for atom substitution.

However, structures with the lowest energy in all compositions were generated with atom substitution except for V_5O_6 , which was generated in our work. This might arise because most of the structures from atom substitution were optimized from the structures with symmetry, whereas the structures of this work and iMatGen were optimized from the structures without any assurances of symmetry. The energy difference between the structures with the lowest formation energy from our model, the iMatGen model, and atom

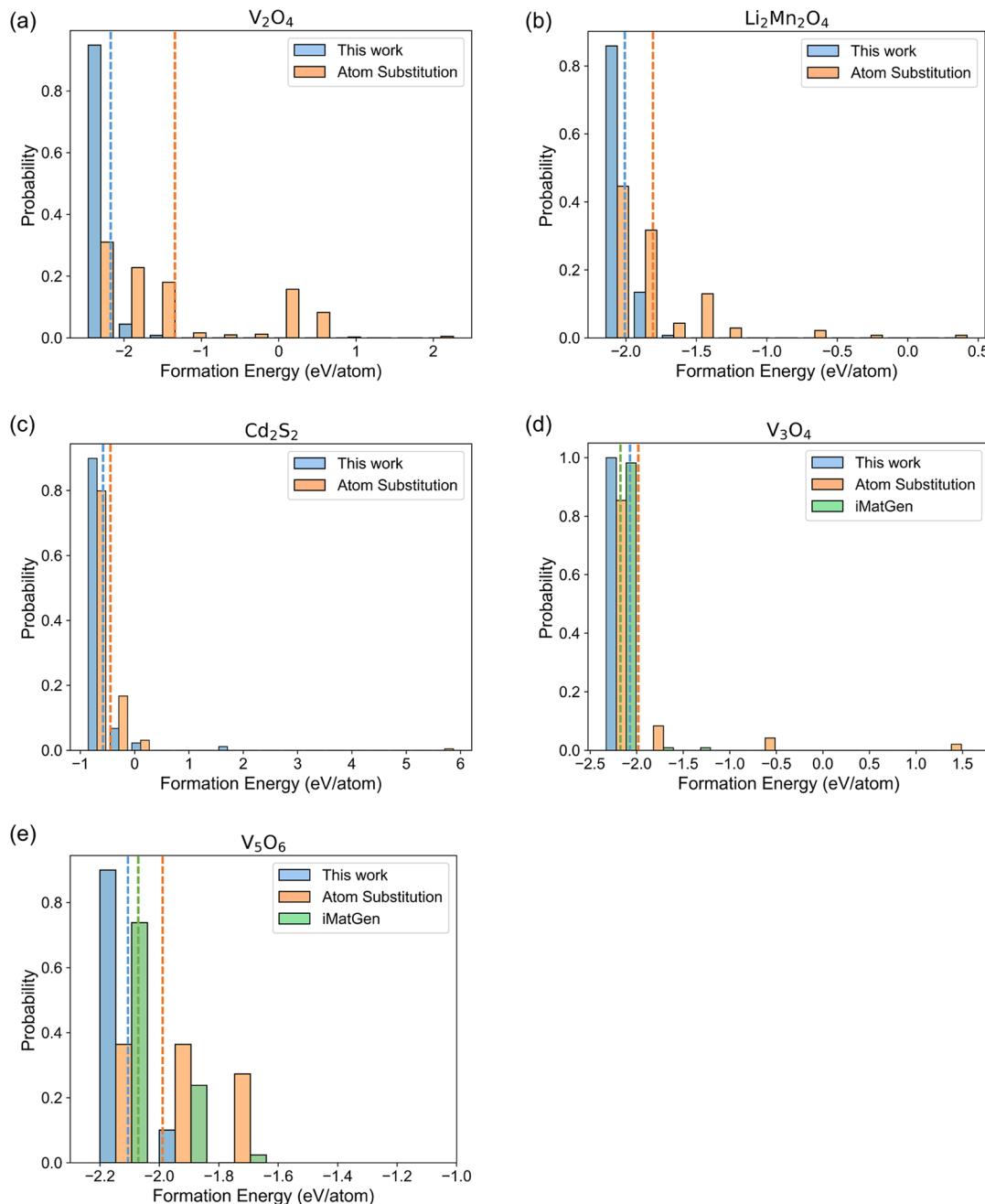


Figure 5. Normalized histogram of the formation energy distribution. The dotted lines denote the mean values. (a–c) Comparison of the formation energy of the structures with the existing compositions between this work (orange) and atom substitution (blue). (d, e) Comparison of the formation energy of the structures with hypothetical compositions in this work (orange), atom substitution (blue), and iMatGen (green).

substitution was minimal, $Li_2Mn_2O_4$, Cd_2S_2 , and $V_5O_6 < 10$ meV/atom, V_2O_4 and $V_3O_4 < 20$ meV/atom.

Nevertheless, our algorithm's proficiency lies in consistently identifying more stable structures on average, and it already discovered the most stable structures in the MP database for these compositions, thus proving to be a valuable tool in materials exploration and research.

The time taken for DFT optimization for each structure can be found in Figure S4. The average time required for DFT optimization was slightly longer for our model than atom substitution and iMatGen except for the V_2O_4 composition. This is because most of the structures resulting from atom substitution were optimized from symmetric structures. In addition, since the training for IMatGen was conducted on a

limited material (VO), it is possible that initial structures close to more stable structures were created.

CONCLUSIONS

In this work, we successfully generated crystals with the desired composition using deep learning and analyzed them through comparison with the existing methods. Our approach enabled the production of various crystals with different compositions in VO, $LiMnO$, and Cds, which were matched with the existing database without the DFT calculation. For precise analysis, we performed DFT calculations after duplicated structures were removed. Finally, 205 unique structures with energy above the hull of less than 100 meV/atom were

obtained. In addition, it was confirmed that the average formation energy of generated structures was low when compared with atom-substituted and other generated structures. Our approach can produce stable crystals with different compositions. Unfortunately, generating crystals with a large number of atoms was challenging, but we plan to supplement this in future research. We explored only known compositions for comparison with the existing database; however, considering the performance, we expect to search for new crystals with unknown compositions. We believe that our approach helps discover new crystal candidates.

METHODS

Search in Latent Space for the Generation with the Desired Composition. The search for the latent space to generate crystals with the desired composition (c) and the desired number of atoms (n) was performed 5000 times by using the existing trained CDVAE model. Initial conditions for crystals were randomly generated from normal distribution as a shape ($d_{\text{latent}}, d_{\text{hidden}}$), where d_{latent} and d_{hidden} are 256. The composition (c) and the number of atoms (n) were predicted from the latent space, and the total loss was calculated as the average of Cross Entropy Losses for the composition and the number of atoms, which was followed:

$$\begin{aligned}\mathcal{L}_{\text{total}} &= \frac{1}{2}(\mathcal{L}_N + \mathcal{L}_C) = \frac{1}{2}(\mathcal{L}(p_N, q_N) + \mathcal{L}(p_C, q_C)) \\ &= \frac{1}{2} \left(- \sum_{i=1}^n p_{N_i} \log q_{N_i} - \sum_{i=1}^n p_{C_i} \log q_{C_i} \right)\end{aligned}$$

where \mathcal{L}_N is the loss for the number of atoms, \mathcal{L}_C is the loss for the composition, p_{N_i} is the true distribution of number of atoms, p_{C_i} is the true distribution of composition, q_{N_i} is the estimated probability distribution of number of atoms, and q_{C_i} is the estimated probability distribution of compositions.

The latent space was optimized in the direction of reducing these losses. During the exploration of the latent space, the relaxed structures were obtained every 1250 times through Langevin Dynamics²⁸ (i.e., four Langevin Dynamics processes). A total of 1024 ($=d_{\text{latent}} \times 4 = 256 \times 4$) crystal structures were obtained. This process was detailed in **Algorithm 1** and repeated from 1 to 10 times. The structures that matched the existing database (e.g., MP database) were determined using Structure Matcher of Pymatgen²⁹ ($\text{ltol} = 0.2$, $\text{stol} = 0.3$, $\text{angle_tol} = 5$). Duplicated structures were also filtered using the tool and criteria, which means removing duplicated structures among the created structures.

```

1. Input: Target Composition ( $p_C$ ), Target Number ( $p_N$ ), Latent space from a normal distribution ( $Z$ ), The number of Crystals ( $C$ ), The number of gradients steps ( $M$ ), The number of Langevin Dynamics Steps ( $D$ )
2. Output: New Crystals with new coordinates ( $X_{\text{New}}$ ), new atom types ( $A_{\text{New}}$ ), and Lattice Parameter ( $L$ )
3. Predict compositions ( $q_C$ ), number of atoms ( $q_N$ ), and Lattice Parameter( $L$ ) from  $Z$ .
4. for  $i \leftarrow 1$  to  $M$  do
     $\mathcal{L}_{\text{total}} = \frac{1}{2}(\mathcal{L}(p_N, q_N) + \mathcal{L}(p_C, q_C))$ 
     $Z \leftarrow Z - \alpha \nabla \mathcal{L}_{\text{total}}$ 
    if  $i = 0$  then
        continue
    if  $i = M // (C - 1)$  or  $i = M - 1$  then
        for  $j \leftarrow 1$  to  $D$  do
             $X, A \leftarrow X, A : \text{Langevin Dynamics}$ 
     $X_{\text{New}}, A_{\text{New}} \leftarrow X, A$ 
Get Crystal Structures ( $X_{\text{New}}, A_{\text{New}}, L$ )

```

Algorithm 1 The process of generating crystals with desired composition and the number of atoms.

DFT Calculation. The DFT calculation was implemented by the VASP program.³¹ We used the Perdew–Burke–Ernzerhof (PBE) functional and the projector augmented wave-PBE (PAW-PBE) pseudopotentials.³² The crystal structures including atomic positions and cell parameters were relaxed using the conjugate gradient descent method. The convergence criteria for energy was 1.0×10^{-5} eV, and grid spacing was 0.5 \AA^{-1} for considering computational costs. The cutoff energy for the plane-wave-basis set was set at 520 eV.

For the structures containing transition metal (e.g., VO system, LiMnO system), spin-polarized GGA + U calculation was performed.³³ The parameters for the U value of transition metals and the initial magnetic moment of each atom were taken from the Materials Project database. The computed energy was uncorrected energy, which should be corrected for calculating the accurate formation energy. We considered the GGA/GGA + U mixing correction energy for transition metal atoms and anion correction energy.³⁴ The correction energies for the V atom and Mn atom were -1.70 and -1.67 eV, respectively. The anion correction energies for oxide and S were -0.69 and -0.5 eV, respectively.⁸ These values were also taken from the Materials Project database. The corrected energy can be obtained by the following equation, where n_i is the number of atoms i and E_i is the correction energy of atom i .

$$E_{\text{corrected}} = E_{\text{uncorrected}} - \sum_i n_i E_i$$

The formation energy can be calculated by subtracting the bulk energy from the compound energy. The equation was followed.

$$E_f = E_{X_x Y_y \dots Z_z} - \sum_i x_i E_{X_i}$$

$E_{X_x Y_y \dots Z_z}$ is the corrected energy of the compound (e.g., VO, LiMnO, CdS), E_{X_i} is the bulk energy of atom X , and x_i is the ratio of X in the compound. The bulk energy calculation for each atom and the ground state energy calculation for each compound were implemented based on the above algorithm and criteria. The details and results can be shown in **Tables S3–S5**.

ASSOCIATED CONTENT

Data Availability Statement

Final structures can be found on GitHub: https://github.com/seunghhs/Crystal_Generation/tree/master/final_csv. The codes are available at https://github.com/seunghhs/Crystal_Generation.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jcim.3c00935>.

Reconstruction ratio of the CDVAE model; database and the training data set; the relation between the number of atoms and the ratio generated correctly; the graphs of formation energy for each structure through DFT calculation; the details of VASP calculation; the crystal structure images with the lowest formation energy; and the graphs for DFT calculation time ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

Changyoung Park – LG AI Research, ISC, Seoul 07796, Republic of Korea; orcid.org/0009-0008-1350-2057; Email: changyoung.park@lgresearch.ai

Authors

Seunghhee Han – Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea; orcid.org/0000-0001-8696-6823

Jaewan Lee – LG AI Research, ISC, Seoul 07796, Republic of Korea; orcid.org/0009-0008-1517-8086

Sehui Han – LG AI Research, ISC, Seoul 07796, Republic of Korea

Seyed Mohamad Moosavi – Department of Chemical Engineering & Applied Chemistry, University of Toronto, Ontario M5S 3E5, Canada; orcid.org/0000-0002-0357-5729

Jihan Kim – Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, Republic of Korea; orcid.org/0000-0002-3844-8789

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jcim.3c00935>

Author Contributions

S.H. and C.P. conceived the idea, implemented the work, and wrote the manuscript with inputs from all the authors. All authors contributed to discussions informing the research.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF) under Project Number 2021M3A7C208974513.

REFERENCES

- (1) (a) Cao, C.; Li, Z.-B.; Wang, X.-L.; Zhao, X.-B.; Han, W.-Q. Recent advances in inorganic solid electrolytes for lithium batteries. *Front. Energy Res.* **2014**, *2*, No. 25. (b) Brammer, L. Developments in inorganic crystal engineering. *Chem. Soc. Rev.* **2004**, *33*, 476–489.

(2) Hautier, G.; Jain, A.; Ong, S. P. From the computer to the laboratory: materials discovery and design using first-principles calculations. *J. Mater. Sci.* **2012**, *47*, 7317–7340.

(3) Court, C. J.; Yildirim, B.; Jain, A.; Cole, J. M. 3-D inorganic crystal structure generation and property prediction via representation learning. *J. Chem. Inf. Model.* **2020**, *60*, 4518–4535.

(4) Kohn, W. Nobel Lecture: Electronic structure of matter—wave functions and density functionals. *Rev. Mod. Phys.* **1999**, *71*, 1253–1266.

(5) (a) Gao, Y.; Wang, X.; Ma, J.; Wang, Z.; Chen, L. Selecting substituent elements for Li-rich Mn-based cathode materials by density functional theory (DFT) calculations. *Chem. Mater.* **2015**, *27*, 3456–3461. (b) Chen, H.; Ma, N.; Li, J.; Wang, Y.; She, C.; Zhang, Y.; Li, X.; Liu, J.; Feng, X.; Zhou, S. Effect of atomic iron on hydriding reaction of magnesium: Atomic-substitution and atomic-adsorption cases from a density functional theory study. *Appl. Surf. Sci.* **2020**, *504*, No. 144489. (c) Liu, X.; Zhou, J.; Zhou, Y.; Wei, Q.; Wei, Y.; Zhang, P. Substitution of sulfur atoms on Ni-Mo-S by ammonia—a DFT study. *Catal. Today* **2020**, *353*, 17–25.

(6) Belsky, A.; Hellenbrandt, M.; Karen, V. L.; Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 364–369.

(7) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: the open quantum materials database (OQMD). *JOM* **2013**, *65*, 1501–1509.

(8) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, No. 011002.

(9) (a) Avdeev, M.; Sale, M.; Adams, S.; Rao, R. P. Screening of the alkali-metal ion containing materials from the Inorganic Crystal Structure Database (ICSD) for high ionic conductivity pathways using the bond valence method. *Solid State Ionics* **2012**, *225*, 43–46. (b) Jain, D.; Chaupe, S.; Khullar, P.; Srinivasan, S. G.; Rai, B. Bulk and surface DFT investigations of inorganic halide perovskites screened using machine learning and materials property databases. *Phys. Chem. Chem. Phys.* **2019**, *21*, 19423–19436. (c) Antoniuk, E. R.; Schindler, P.; Schroeder, W. A.; Dunham, B.; Pianetta, P.; Vecchione, T.; Reed, E. J. Novel Ultrabright and Air-Stable Photocathodes Discovered from Machine Learning and Density Functional Theory Driven Screening. *Adv. Mater.* **2021**, *33*, No. 2104081. (d) Kim, M.; Singh, S. P.; Lee, J.-W.; Izawa, T.; Kim, D.; Yun, B.; Yoon, C.; Park, W. B.; Sohn, K.-S. Identification of a narrow band red light-emitting phosphor using computational screening of ICSD: Its synthesis and optical characterization. *J. Alloys Compd.* **2019**, *774*, 338–346.

(10) Davies, D. W.; Butler, K. T.; Jackson, A. J.; Morris, A.; Frost, J. M.; Skelton, J. M.; Walsh, A. Computational screening of all stoichiometric inorganic materials. *Chem* **2016**, *1*, 617–627.

(11) (a) Noh, J.; Gu, G. H.; Kim, S.; Jung, Y. Machine-enabled inverse design of inorganic solid materials: promises and challenges. *Chem. Sci.* **2020**, *11*, 4871–4881. (b) Kim, B.; Lee, S.; Kim, J. Inverse design of porous materials using artificial neural networks. *Sci. Adv.* **2020**, *6*, No. eaax9324.

(12) Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks (2014), arXiv preprint arXiv:1406.2661. arXiv.org e-Print archive, 2014. <https://doi.org/10.48550/arXiv.1406.2661>.

(13) Kingma, D. P.; Welling, M. Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114. arXiv.org e-Print archive, 2013. <https://doi.org/10.48550/arXiv.1312.6114>.

(14) Noh, J.; Kim, J.; Stein, H. S.; Sanchez-Lengeling, B.; Gregoire, J. M.; Aspuru-Guzik, A.; Jung, Y. Inverse design of solid-state materials via a continuous representation. *Matter* **2019**, *1*, 1370–1384.

(15) Kim, S.; Noh, J.; Gu, G. H.; Aspuru-Guzik, A.; Jung, Y. Generative adversarial networks for crystal structure prediction. *ACS Cent. Sci.* **2020**, *6*, 1412–1420.

- (16) Ren, Z.; Tian, S. I. P.; Noh, J.; Oviedo, F.; Xing, G.; Li, J.; Liang, Q.; Zhu, R.; Aberle, A. G.; Sun, S.; et al. An invertible crystallographic representation for general inverse design of inorganic crystals with targeted properties. *Matter* **2022**, *5*, 314–335.
- (17) Xie, T.; Fu, X.; Ganea, O.-E.; Barzilay, R.; Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation, arXiv preprint arXiv:2110.06197. arXiv.org e-Print archive, 2021. <https://doi.org/10.48550/arXiv.2110.06197>.
- (18) Ramesh, A.; Dhariwal, P.; Nichol, A.; Chu, C.; Chen, M. Hierarchical text-conditional image generation with clip latents, arXiv preprint arXiv:2204.06125. arXiv.org e-Print archive, 2022. <https://arxiv.org/abs/2204.06125>.
- (19) Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E. L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T. Photorealistic text-to-image diffusion models with deep language understanding. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 36479–36494.
- (20) Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; Ommer, B. In *High-Resolution Image Synthesis with Latent Diffusion Models*, IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); IEEE, 2022; pp 10684–10695.
- (21) (a) Xu, M.; Yu, L.; Song, Y.; Shi, C.; Ermon, S.; Tang, J. Geodiff: A geometric diffusion model for molecular conformation generation, arXiv preprint arXiv:2203.02923. arXiv.org e-Print archive, 2022. <https://doi.org/10.48550/arXiv.2203.02923>. (b) Hoogeboom, E.; Satorras, V. G.; Vignac, C.; Welling, M. Equivariant diffusion for molecule generation in 3d. *Int. Conf. Mach. Learn.* **2022**, 8867–8887.
- (22) Corso, G.; Stärk, H.; Jing, B.; Barzilay, R.; Jaakkola, T. Diffdock: Diffusion steps, twists, and turns for molecular docking, arXiv preprint arXiv:2210.01776. arXiv.org e-Print archive, 2022. <https://doi.org/10.48550/arXiv.2210.01776>.
- (23) Lyngby, P.; Thygesen, K. S. Data-driven discovery of 2D materials by deep generative models. *npj Comput. Mater.* **2022**, *8*, No. 232.
- (24) Moustafa, H.; Lyngby, P. M.; Mortensen, J. J.; Thygesen, K. S.; Jacobsen, K. W. Hundreds of new, stable, one-dimensional materials from a generative machine learning model. *Phys. Rev. Mater.* **2023**, *7*, No. 014007.
- (25) Collins, C.; Dyer, M.; Pitcher, M.; Whitehead, G.; Zanella, M.; Mandal, P.; Claridge, J.; Darling, G.; Rosseinsky, M. Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature* **2017**, *546*, 280–284.
- (26) Gu, G. H.; Jang, J.; Noh, J.; Walsh, A.; Jung, Y. Perovskite synthesizability using graph neural networks. *npj Comput. Mater.* **2022**, *8*, No. 71.
- (27) Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.
- (28) Song, Y.; Ermon, S. Generative modeling by estimating gradients of the data distribution. *Proc. Int. Conf. Neural Inf. Process. Syst.* **2019**, *32*, 11918–11930.
- (29) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.
- (30) Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *London, Edinburgh, Dublin Philos. Mag. J. Sci.* **1901**, *2*, 559–572.
- (31) Kresse, G.; Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, No. 11169.
- (32) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **1999**, *59*, No. 1758.
- (33) Liechtenstein, A. I.; Anisimov, V. I.; Zaanen, J. Density-functional theory and strong interactions: Orbital ordering in Mott-Hubbard insulators. *Phys. Rev. B* **1995**, *52*, No. R5467.
- (34) Jain, A.; Hautier, G.; Ong, S. P.; Moore, C. J.; Fischer, C. C.; Persson, K. A.; Ceder, G. Formation enthalpies by mixing GGA and GGA+ U calculations. *Phys. Rev. B* **2011**, *84*, No. 045115.