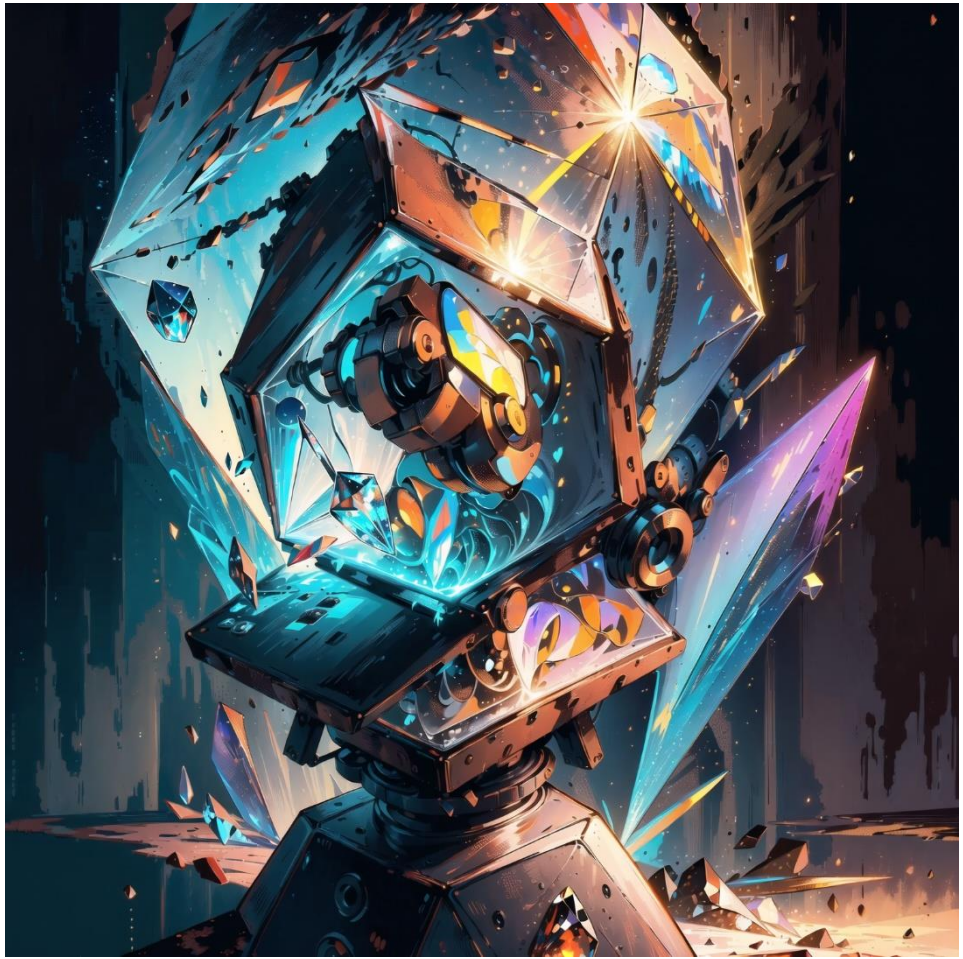
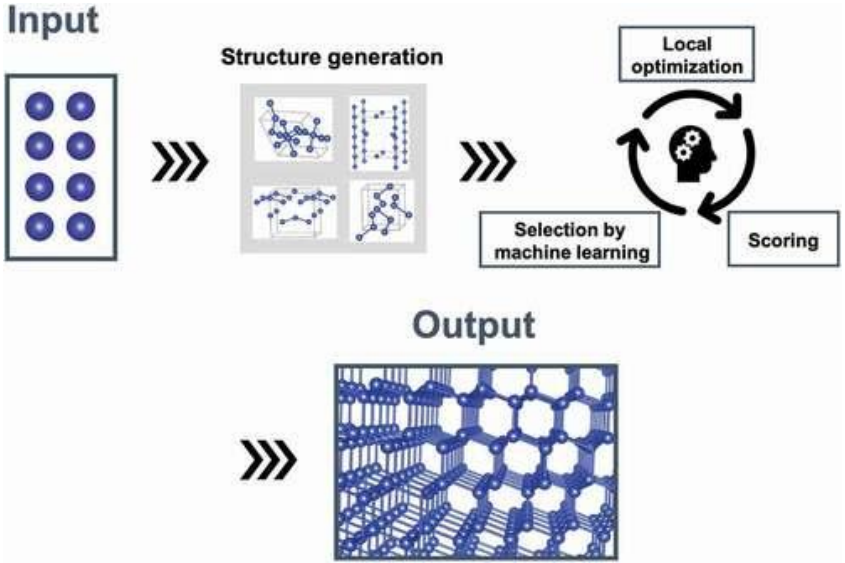
# 基于机器学习的晶格结构预测

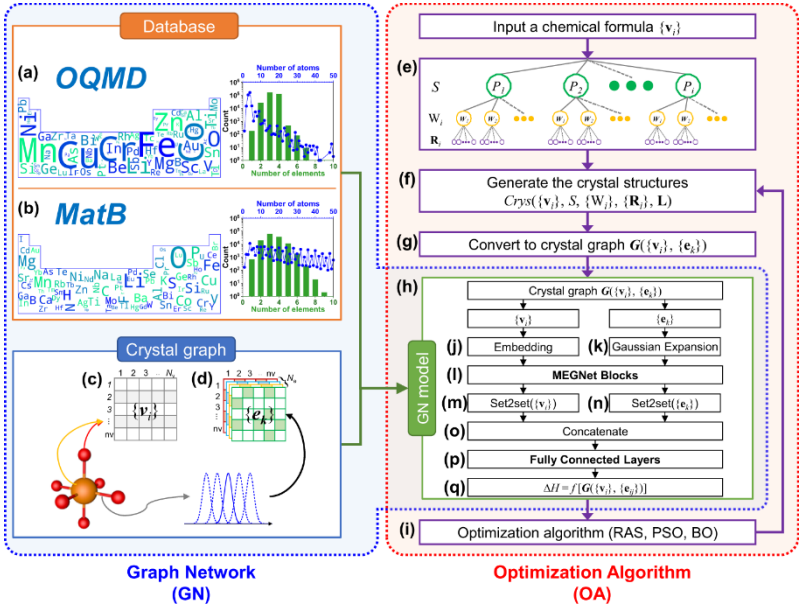
Ziyuan Cheng and Yuhang Zhang



**A picture made by stable diffusion with prompt about “crystal prediction with machine learning”**

# Common practice of machine learning to predict crystal structures




Science and Technology of Advanced Materials: Methods, 1:1, 87-97 (2021)

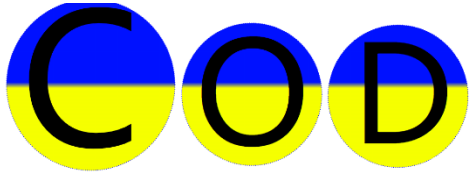
*Nature Communications* **volume 13**, Article number: 1492 (2022)

# Test 1 Direct prediction

Obtaining data

Database A

Download Cif from




**Open-access collection of crystal structures of organic, inorganic, metal-organic compounds and minerals, excluding biopolymers.**

*g data and software from CrystalEye, developed by Nick Day at the department of Chemistry, the University of Cambridge under supervision of Peter Murray-Rus*

All data on this site have been placed in the public domain by the contributors.

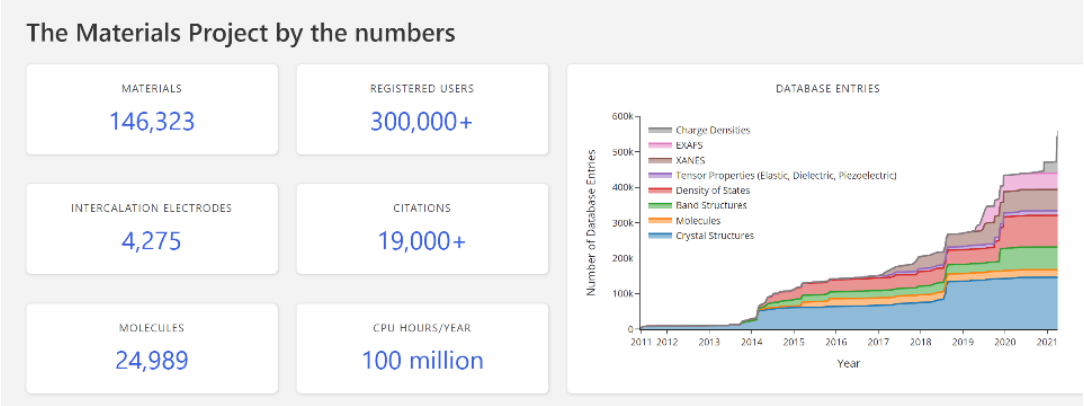
Currently there are **501477** entries in the COD.
Latest deposited structure: 7062445 on **2023-04-27** at **00:26:44 UTC**

大小: 21.2 GB 共 501354 个文件和 6978 个文件夹 压缩率 24.4%

Database B

Obtain data with API from

With pymatgen



Data_ The lowest energy configuration of Combinations of the first 85 elements of the periodic table of elements

| | formula | elements | nsites | element_symbols | coordinates | energy | spacegroup | spacegroup_num |
|---|---|---|---|---|---|---|---|---|
| **0** | Ag11Hg9 | Ag, Hg | 20 | ['Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag'... | [[ 1.609567 -0.92928553 33.90805742]\n [ 1.6... | -677.652749 | P3m1 | 156 |
| **1** | Ag2Br3 | Ag, Br | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'Br', 'Br', 'Br', 'Br'... | [[-3.46990705e+00 -2.00392334e+00 1.22402099e... | -168.691087 | R-3c | 167 |
| **2** | Ag2Cl3 | Ag, Cl | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'Cl', 'Cl', 'Cl', 'Cl'... | [[-3.30941698e+00 -1.91047971e+00 1.71721380e... | -127.332050 | R-3c | 167 |
| **3** | Ag2F | Ag, F | 3 | ['Ag', 'Ag', 'F'] | [[-1.83755070e-06 1.72922338e+00 3.97018472e... | -48.352742 | P-3m1 | 164 |
| **4** | Ag2F3 | Ag, F | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'F', 'F', 'F', 'F', '... | [[-2.87395102e+00 -1.65949299e+00 2.26233906e... | -116.183864 | R-3c | 167 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **10623** | ZrZn16 | Zr, Zn | 34 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'... | [[-2.17422601 6.06378581 8.71633518]\n [ 6.4... | -333.573781 | Cmcm | 63 |
| **10624** | ZrZn2 | Zr, Zn | 6 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn'] | [[ 5.25615317e+00 3.71666061e+00 9.10392570e... | -83.354040 | Fd-3m | 227 |
| **10625** | ZrZn22 | Zr, Zn | 46 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'... | [[ 4.26511414e+00 3.01589110e+00 7.38739309e... | -440.455327 | Fd-3m | 227 |
| **10626** | ZrZn3 | Zr, Zn | 8 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'] | [[ 2.24703936e-06 3.37600716e+00 3.24801562e... | -101.493900 | P6_3/mmc | 194 |
| **10627** | NaN | Na, N | 8 | ['Na', 'Na', 'Na', 'Na', 'N', 'N', 'N', 'N'] | [[3.93926 4.366545 2.46981583]\n [3.93926... | -50.234293 | Fddd | 70 |

10628 rows × 8 columns

Data_

| | formula | elements | nsites | element_symbols | coordinates | energy | spacegroup | spacegroup_num |
|---|---|---|---|---|---|---|---|---|
| **0** | Ag11Hg9 | Ag, Hg | 20 | ['Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag', 'Ag'... | [[ 1.609567 -0.92928553 33.90805742]\n [ 1.6... | -677.652749 | P3m1 | 156 |
| **1** | Ag2Br3 | Ag, Br | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'Br', 'Br', 'Br', 'Br'... | [[-3.46990705e+00 -2.00392334e+00 1.22402099e... | -168.691087 | R-3c | 167 |
| **2** | Ag2Cl3 | Ag, Cl | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'Cl', 'Cl', 'Cl', 'Cl'... | [[-3.30941698e+00 -1.91047971e+00 1.71721380e... | -127.332050 | R-3c | 167 |
| **3** | Ag2F | Ag, F | 3 | ['Ag', 'Ag', 'F'] | [[-1.83755070e-06 1.72922338e+00 3.97018472e... | -48.352742 | P-3m1 | 164 |
| **4** | Ag2F3 | Ag, F | 10 | ['Ag', 'Ag', 'Ag', 'Ag', 'F', 'F', 'F', 'F', '... | [[-2.87395102e+00 -1.65949299e+00 2.26233906e... | -116.183864 | R-3c | 167 |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... |
| **10623** | ZrZn16 | Zr, Zn | 34 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'... | [[-2.17422601 6.06378581 8.71633518]\n [ 6.4... | -333.573781 | Cmcm | 63 |
| **10624** | ZrZn2 | Zr, Zn | 6 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn'] | [[ 5.25615317e+00 3.71666061e+00 9.10392570e... | -83.354040 | Fd-3m | 227 |
| **10625** | ZrZn22 | Zr, Zn | 46 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'... | [[ 4.26511414e+00 3.01589110e+00 7.38739309e... | -440.455327 | Fd-3m | 227 |
| **10626** | ZrZn3 | Zr, Zn | 8 | ['Zr', 'Zr', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn', 'Zn'] | [[ 2.24703936e-06 3.37600716e+00 3.24801562e... | -101.493900 | P6_3/mmc | 194 |
| **10627** | NaN | Na, N | 8 | ['Na', 'Na', 'Na', 'Na', 'N', 'N', 'N', 'N'] | [[3.93926 4.366545 2.46981583]\n [3.93926... | -50.234293 | Fddd | 70 |

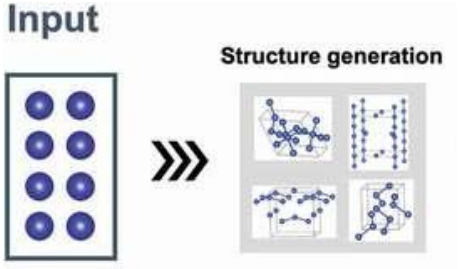
10628 rows × 8 columns

An MLP network

Accuracy: 0.0973659454374412

Accuracy: 0.14487300094073377

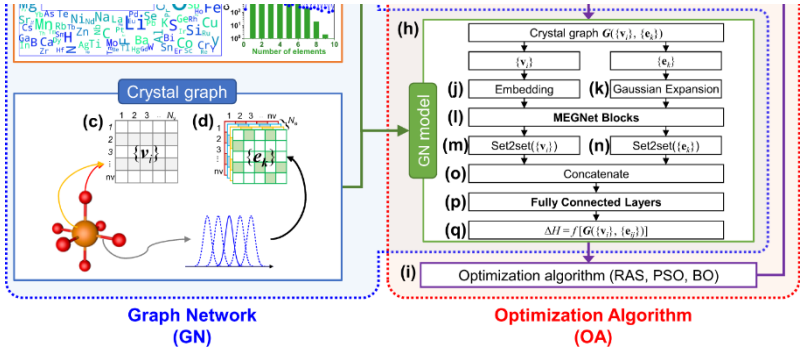
# From elements to structures




Machine learning here with

Generative models ?

# From structures to energy or other ways to pick one




GNN and …

# Just like …

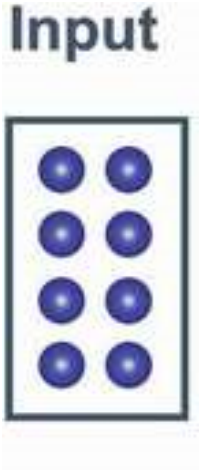
Generate well-defined structures and Then pick one

Seed: 3040976647

Seed: 1301843965

Seed: 1441793906

Seed: 3939324262

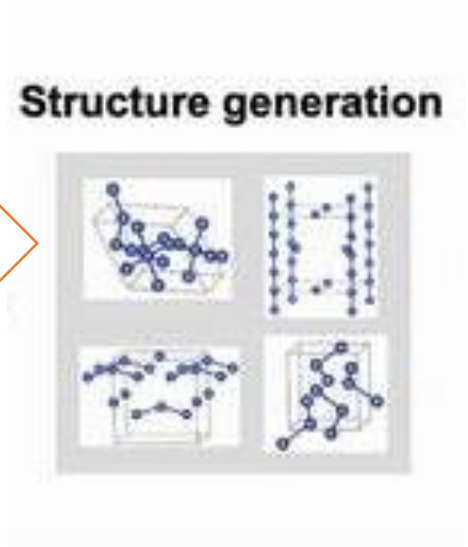



Seqs

(strings, etc)

Seqs

(strings, etc)




Apply transformers and

could Using some pre-trained models

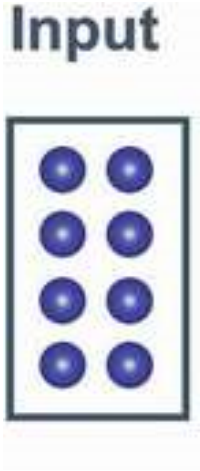
like: gpt-2 , T5, etc.

Information of Structures

na(3.93926, 4.366545, 2.46981583)_na(3.93926, 4.366545, 7.46388417)_na(1.96963, 2.1832725, 0.01360917)_na(1.96963, 2.1832725, 4.95324083)_n(3.93926, 4.96663928, 4.96685)_n(3.93926, 3.76645072, 4.96685)_n(5.90889, 5.94972322, 7.450275)_n(5.90889, 7.14991178, 7.450275)_sg70

na_r166.00_e[ne]3s1_p0.5_n_r71.00_e[he]
2s2 2p3_p0.5

Information of Atoms

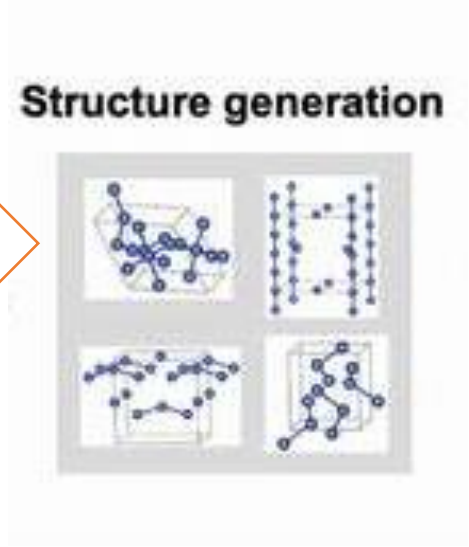



Seqs

(strings, etc)

Seqs

(strings, etc)




Apply transformers and

could Using some pre-trained models

like: gpt-2 , T5, etc.

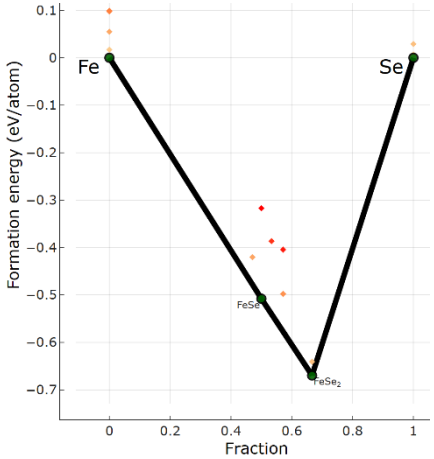
Information of Structures

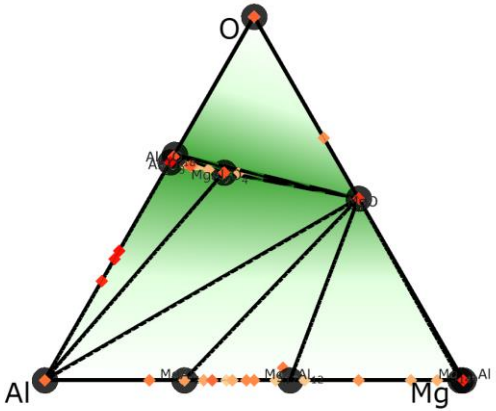
na(3.93926, 4.366545, 2.46981583)_na(3.93926, 4.366545, 7.46388417)_na(1.96963, 2.1832725, 0.01360917)_na(1.96963, 2.1832725, 4.95324083)_n(3.93926, 4.96663928, 4.96685)_n(3.93926, 3.76645072, 4.96685)_n(5.90889, 5.94972322, 7.450275)_n(5.90889, 7.14991178, 7.450275)_sg70

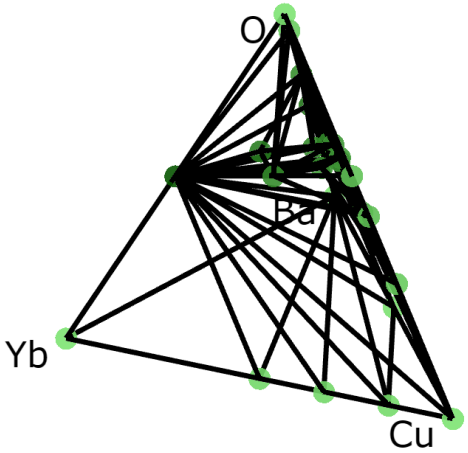
na_r166.00_e[ne]3s1_p0.5_n_r71.00_e[he]
2s2 2p3_p0.5

Information of Atoms

# Further




If one has a good binary or ternary structure generating model and then trains a small amount of structured data such as quadratic, quintuplet, etc., can one perform well with a wider variety of elements?

# Thanks

Special thanks to chatgpt