

Znajdowanie stabilnych konformacji związków

Rafał Kajca, Mateusz Hurkała, Andrzej Sala

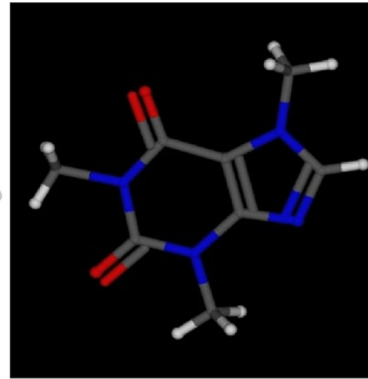
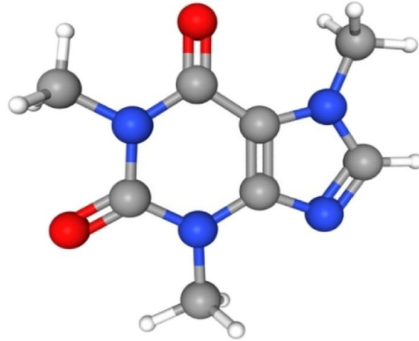
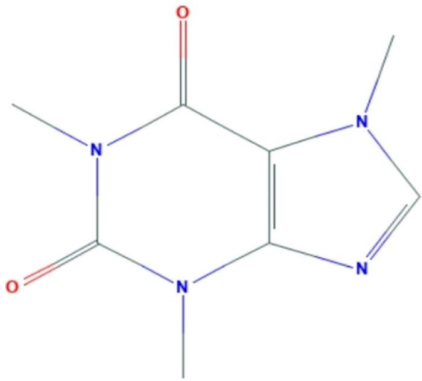
Przypomnienie Problemu

Większość związków chemicznych, jest w stanie przyjmować różne konformacje - możliwe ułożenia przestrzenne atomów w cząsteczce. Pomimo wielu możliwych konformacji, większość z nich przyjmuje głównie jedną, która odznacza się najniższą energią, a więc tym samym największą stabilnością.

Badania nad szukaniem optymalnych konformacji związków chemicznych są kluczowe dla zrozumienia ich właściwości fizykochemicznych, aktywności biologicznej i stabilności, co ma zastosowanie m.in. w projektowaniu leków.

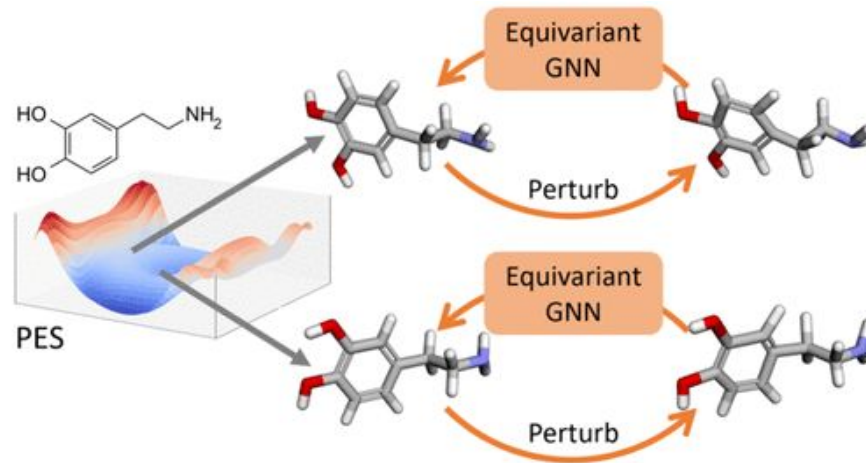
Nasze podejście

- Będziemy przewidywać modelem 'energię' cząstki, aby porównać ją z 'energiami' innych cząstek. Tak wyznaczymy najbardziej stabilną konformację.

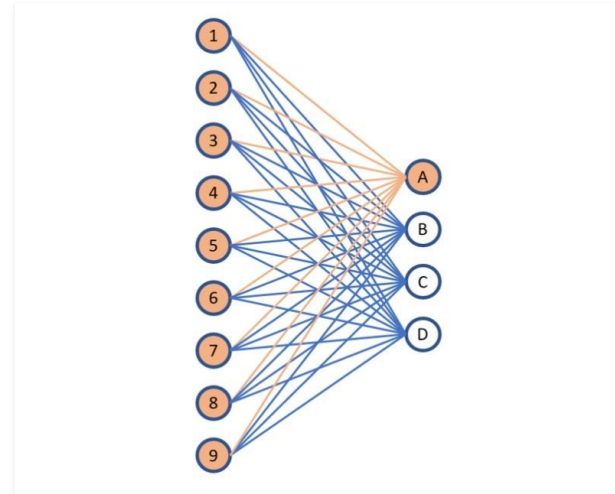


Model E(3)-GNN

- Wejście:
 - 16 skalarów (embedding atomu)
- Środek:
 - 16 skalarów
 - 16 wektorów
- Wyjście:
 - 1 skalar



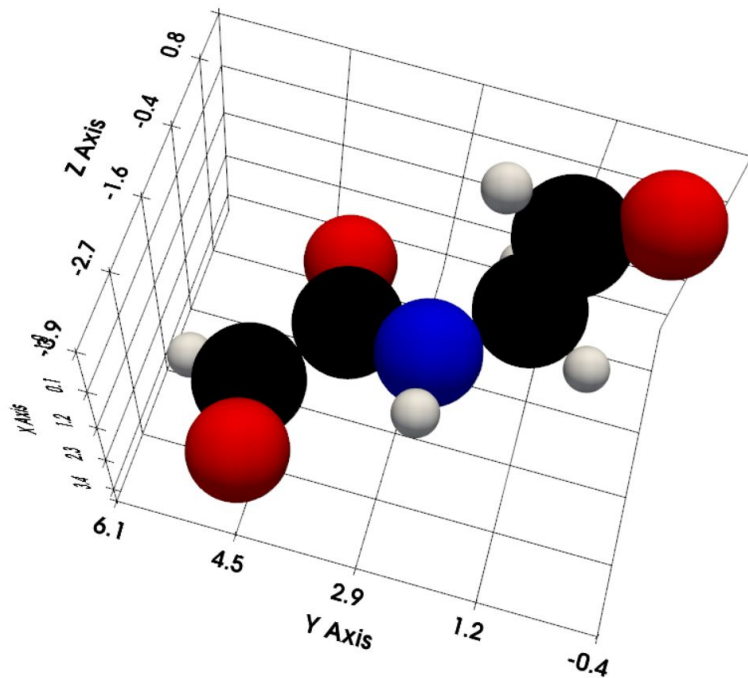
- Warstwa Fully-Connected In \rightarrow Hid
- Graf z krawędziami **max 5 Å**
- Message Passing
 - Warstwy Tensor Product Fully-Connected Hid \rightarrow Hid
 - Sigmoidalne bramki nieliniowe
- Warstwa Fully-Connected Hid \rightarrow Out
- Sumowanie po atomach



QM9

- Model 1 -> Trenowany wyłącznie na QM9
- MSE na zbiorze Testowym: **~20**

- pred: **-11857.59375 eV**
- target: **-11858.01953125 eV**



Energia

6	ZPVE	Zero point vibrational energy	eV
7	U_0	Internal energy at 0K	eV
8	U	Internal energy at 298.15K	eV
9	H	Enthalpy at 298.15K	eV
10	G	Free energy at 298.15K	eV
11	c_v	Heat capacity at 298.15K	$\frac{\text{cal}}{\text{mol K}}$
12	U_0^{ATOM}	Atomization energy at 0K	eV
13	U^{ATOM}	Atomization energy at 298.15K	eV
14	H^{ATOM}	Atomization enthalpy at 298.15K	eV
15	G^{ATOM}	Atomization free energy at 298.15K	eV

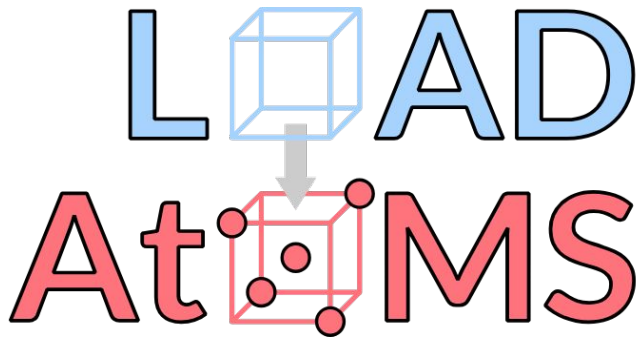
eV?

kcal/mol ?

hartree ?

Inne datasety

- GEOM ?
 - drugs -> 12 GB plik tar
 - Open QDC loader -> Http Error 403
- MD17
 - Wartości energii zbyt różne od QM9

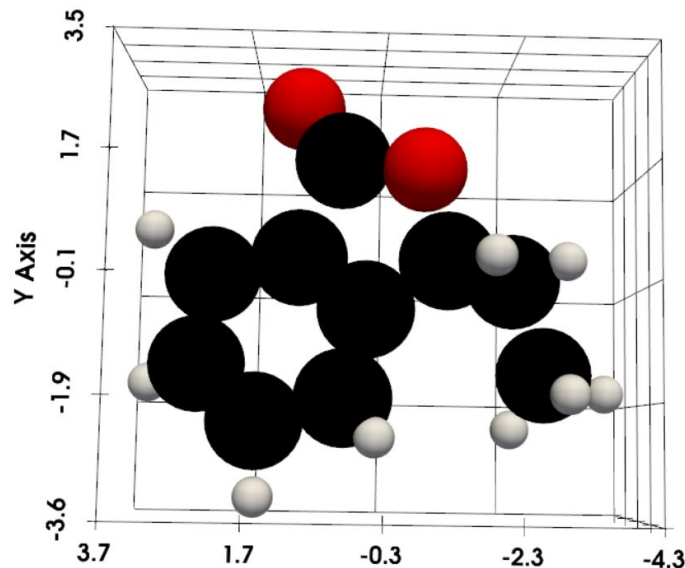


<https://jla-gardner.github.io/load-atoms/index.html>

ANI-1ccx

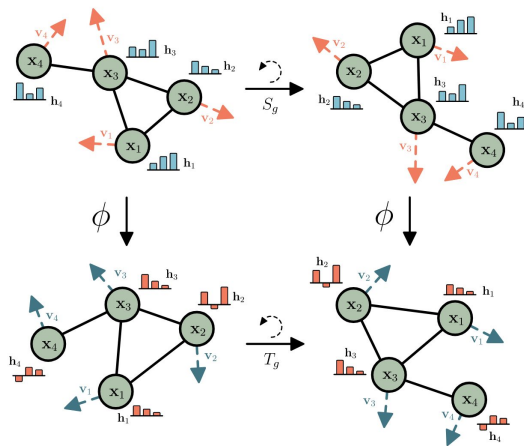
- Podzbiór ANI-1x
- Większe cząstki -> to co chcemy (20 ~ 30 atomów)
- Model 1:
 - pred: -13653.0 eV
 - target: -14604.6 eV

Nie jest idealnie, nasz model nie daje sobie rady z większymi cząstkami



Hipoteza 0

Model E3GNN można skutecznie zgeneralizować zarówno na małe, jak i duże cząstki, zachowując wysoką jakość predykcji niezależnie od rozmiaru analizowanego układu.



Model 2

Model 2 to Model 1 *fine tunowany* do datasetu ANI-1ccx

Chcieliśmy osiągnąć model sprawny do określenia energii dla małych, jak i dla dużych cząstek.

ANI-1ccx: (test MSE: **46497**)

pred: **-24648.5 eV**, target: **-23847.3 eV**

pred: **-9068.7 eV**, target: **-9062.3 eV**

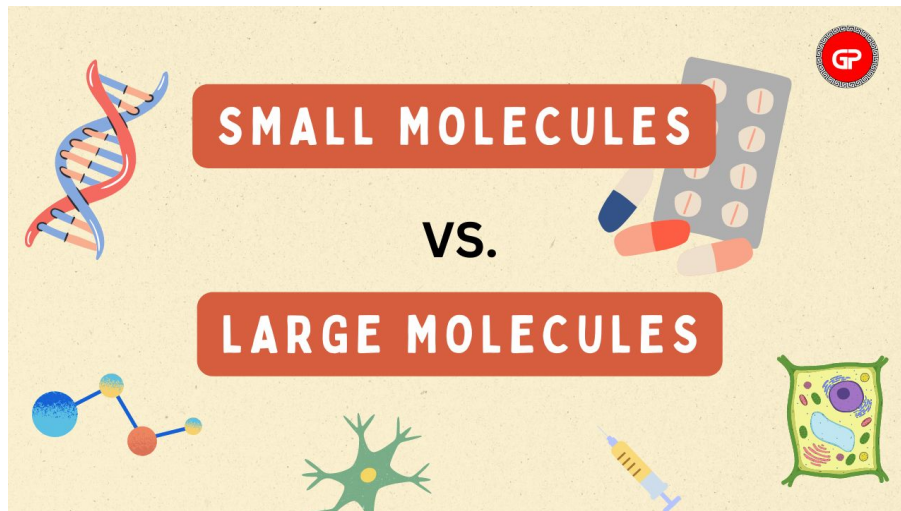
QM9: (test MSE: **190490**)

pred: **-9195.1 eV**, target: **-8660.2 eV**

pred: **-12188.4 eV**, target: **-11705.5 eV**

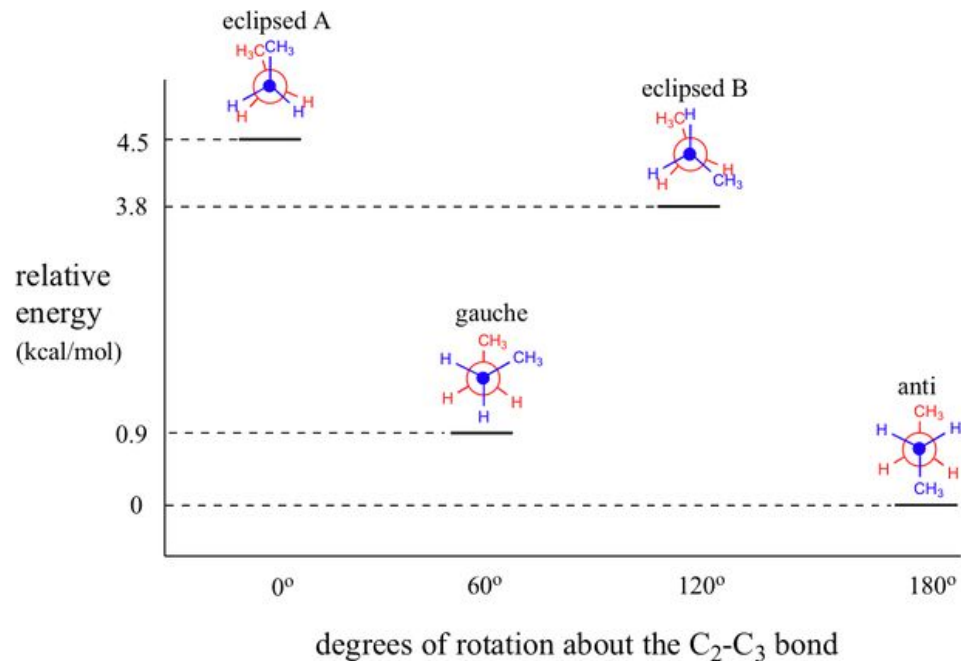
Wynik hipotezy

Obecnie nie możemy tego potwierdzić ze względu na brak odpowiednich danych treningowych obejmujących szeroki zakres rozmiarów oraz ograniczone zasoby obliczeniowe, które utrudniają optymalizację modelu na dużych układach. W efekcie model dobrze działa tylko na rozmiarach zbliżonych do danych treningowych, lecz wykazuje pewne (choć niewielkie) zdolności adaptacyjne.



Hipoteza 1

Model poprawnie znajduje konformację o najniższej energii.



Eksperyment 1

Próba rozpoznania optymalnej konformacji dla butanu:

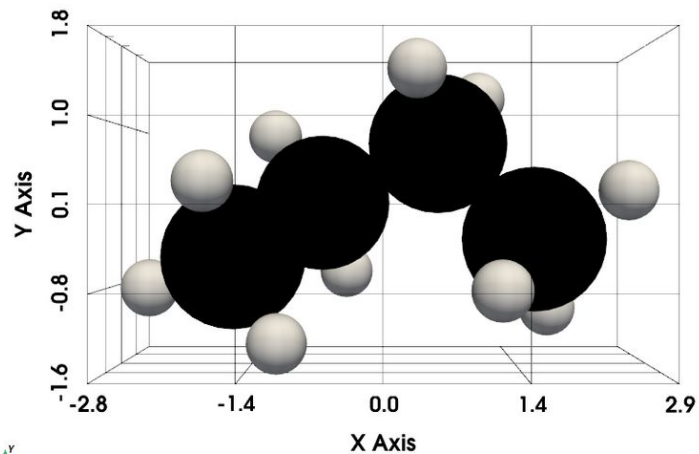
- Konformacja antyperiplanarna
- Konformacja synperiplanarna
- Konformacja synklinalna (Gauche)

Predykcje zwrócone przez pierwszy model (U_0 [eV]):

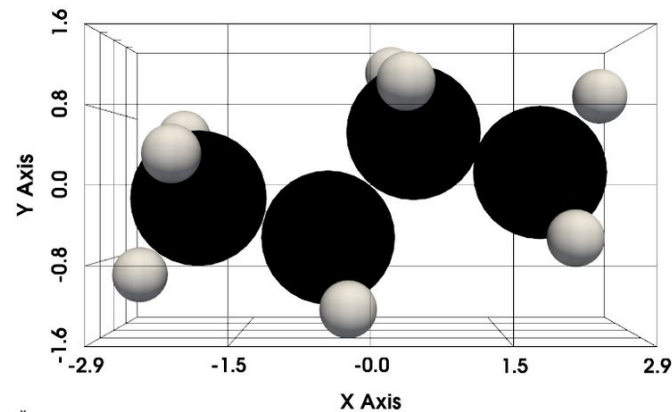
antyperiplanarna: **-32384 eV**, synperiplanarna: **-32005 eV**, synklinalna (Gauche): **-32025 eV**

Predykcje zwrócone przez drugi model (U_0 [eV]):

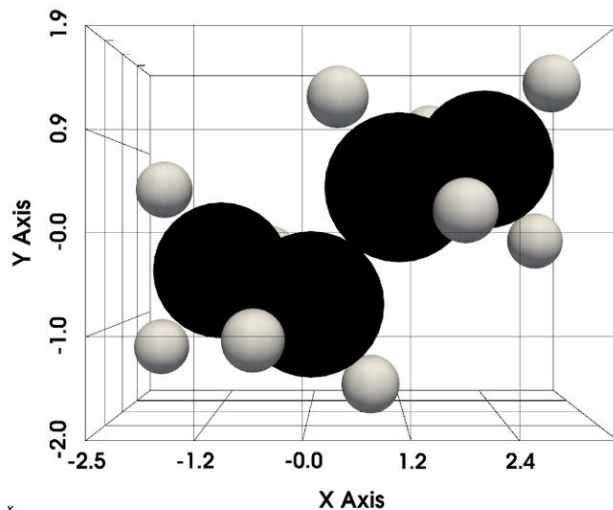
antyperiplanarna: **-15343 eV**, synperiplanarna: **-15183 eV**, synklinalna (Gauche): **-15149 eV**



konformacja synperiplanarna
 model 1: -32005 eV
 model 2: -15183 eV



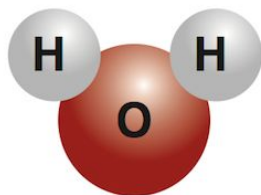
konformacja antyperiplanarna
 (optymalna)
 model 1: -32384 eV
 model 2: -15343 eV



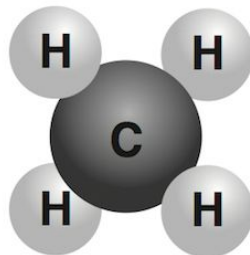
konformacja synklinalna (Gauche)
 model 1: -32025 eV
 model 2: -15149 eV

Hipoteza 2

Model działa także na związkach niepodobnych do treningowych.



Water
 H_2O



Methane
 CH_4

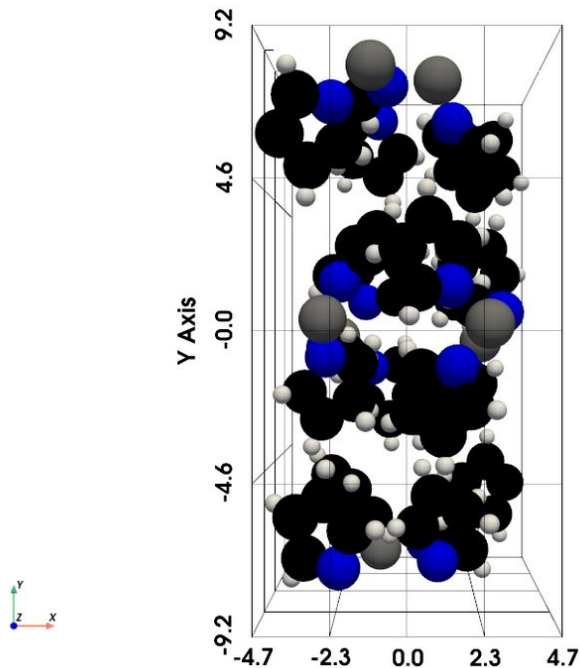
Eksperyment 2

- **500** plików cif oryginalnych z bazy krystalograficznej
- Zaszumianie oryginalnych plików o max wybrane wartości [Anstrems]
- Sprawdzamy czy rozpoznamy oryginalny plik

10 Å, 5 Å, 2.5 Å, 1 Å, 0.5 Å, 0.25 Å

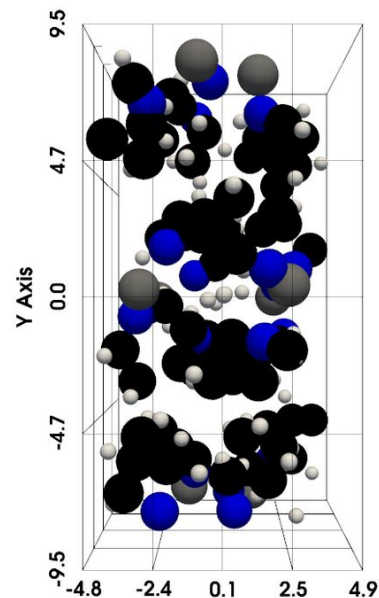
Pierwszy model: **93 %, 61 %, 45 %, 46 %, 50 %, 50 %**

Drugi model: **100 %, 96%, 68%, 55 %, 52 %, 50 %**



Struktura
optymalna

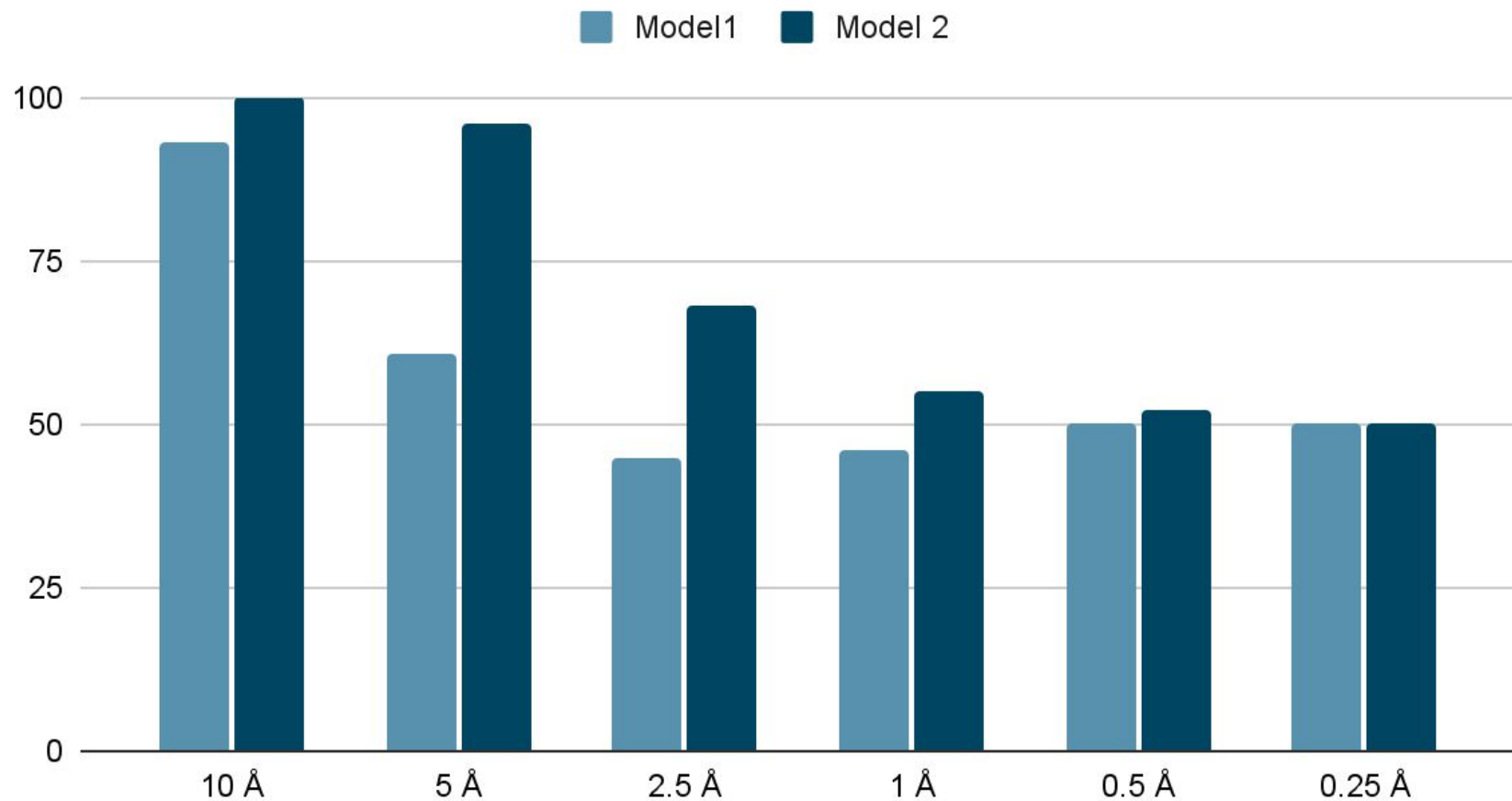
Original file predicted value:
-234930.265625 eV



Struktura o losowym
przesunięciu atomów w zakresie
od **-0,5** do **0,5** Å względem
struktury optymalnej

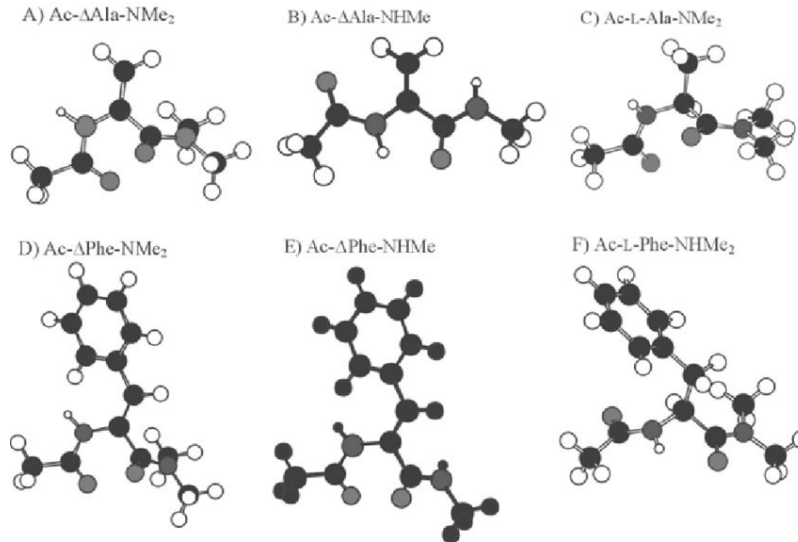
Modified file predicted value:
-225240.296875 eV

Accuracy



Hipoteza 3

Model działa w przypadku porównań bardzo podobnych konformacji



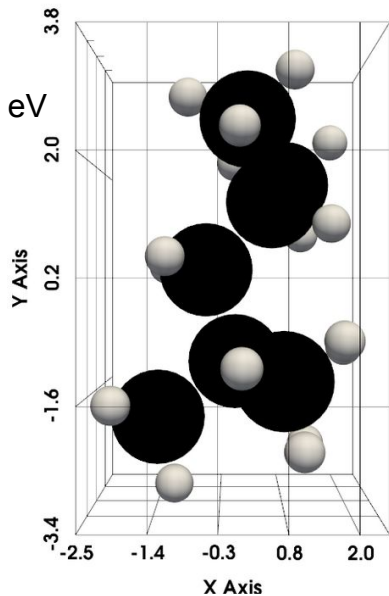
Eksperyment 3

- Oryginalna konformacja
- **5** zaszumionych konformacji (pozycje atomów zaszumione o maksymalnie **0.1 Angstremy**)
- QM9 dataset

Wyniki:

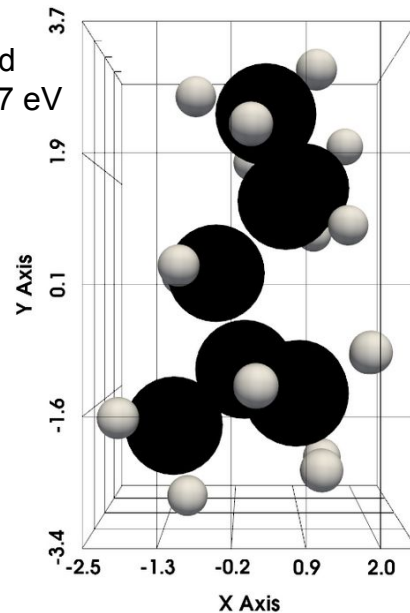
Udało się uzyskać **42.53%** poprawności w rozpoznaniu oryginalnej konformacji

Original file predicted
value: -32786.367188 eV



Struktura
optymalna

Modified file predicted
value: -32786.031247 eV

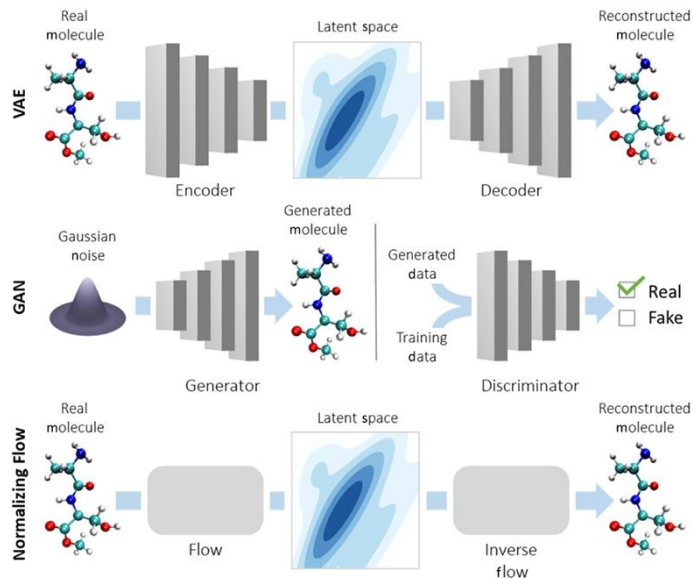


Struktura o losowym
przesunięciu atomów w
zakresie od -0,1 do 0,1 Å
względem struktury
optymalnej

42,53% poprawności modelu

Hipoteza 4

Czy tak nauczony model E3GNN można wykorzystać do modelu dyfuzyjnego zdolnego do generowania optymalnych konformacji związku (np. na podstawie SMILES)



Hipoteza 4 została nie rozstrzygnięta.

Okazuje się że modele dyfuzyjne (któż by się tego spodziewał 😲) wymagają dużych zasobów obliczeniowych)

Trenowaliśmy model **12 h** na Google Collabie (T4) bez żadnych skutków.



Bibliografia

- https://chem.libretexts.org/Courses/Athabasca_University/Chemistry_350%3A_Organic_Chemistry_I/03%3A_Organic_Compounds-_Alkanes_and_Their_Stereochemistry/3.07%3A_Conformations_of_Other_Alkanes
- https://projects.volkamerlab.org/teachopencadd/talktorials/T036_e3_equivariant_gnn.html
- <https://arxiv.org/pdf/2102.09844>
- https://www.sciencedirect.com/science/article/pii/S2001037024002228?ref=pdf_download&fr=RR-2&rr=92fb83491d7ac3bd
- <https://arxiv.org/pdf/2203.17003>
- <https://jla-gardner.github.io/load-atoms/index.html>
- <https://e3nn.org/>