DeepBioisostere: Discovering Bioisosteres with Deep Learning for a Delicate Control of Multiple Molecular Properties

Hyeongwoo Kim, Seokhyun Moon, Wonho Zhung,
Jaechang Lim, and Woo Youn Kim*

ICL-Team

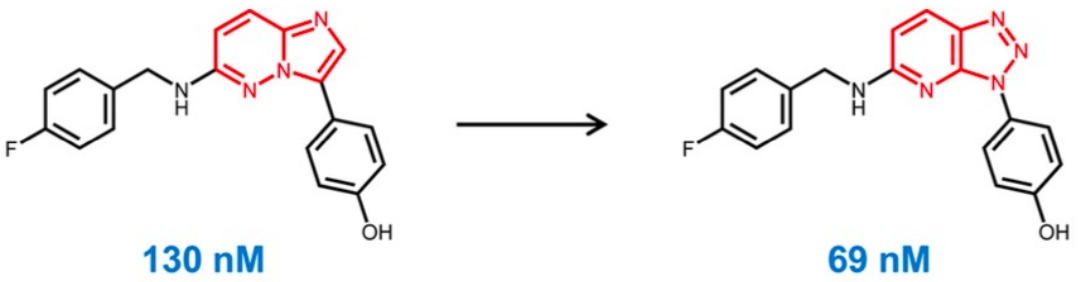
KAIST Chemistry

2024.04.26 (Fri)

Introduction

Molecule optimization

- Improves molecular properties for chemical / biological purposes.
- In **drug discovery**, 1) Binding affinity and 2) Pharmacokinetic properties (ADEMT) are the crucial properties to be improved.



Molecule optimization
with bioisosteric replacement

Frequently replaced moiety pairs are identified and named as **bioisosteres** of each other.

Bioisosteres can act as a toolbox for medicinal chemists to optimize lead compounds.

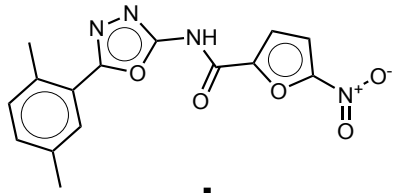
Introduction

Optimization of molecule via bioisosteric replacement

(1) Removal fragment selection (2) Inserting fragment selection — Two crucial steps

Optimization process

Original molecule

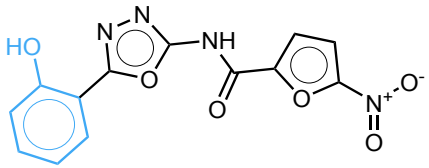


Optimization objective

Goal 1: Higher solubility

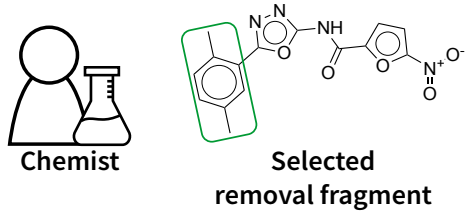
Goal 2: Retained mol. weight

Optimized molecule

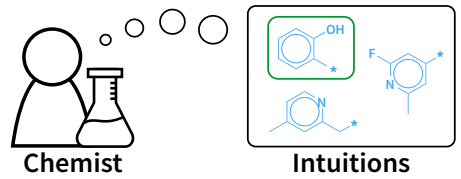


Manual approaches

Selection by **chemists**

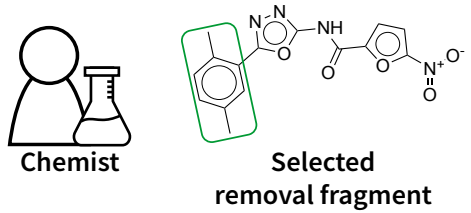


Proposed by **chemists** based on their intuitions

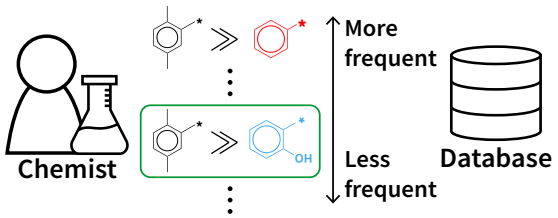


Database mining approaches

Selection by **chemists**

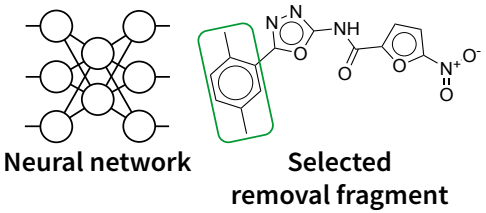


Selection by **chemists** based on a replacement database

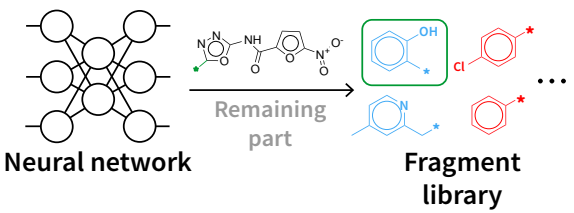


DeepBioisostere [1]

Selection by an **AI**



Selection by an **AI** considering compatability with the remaining part

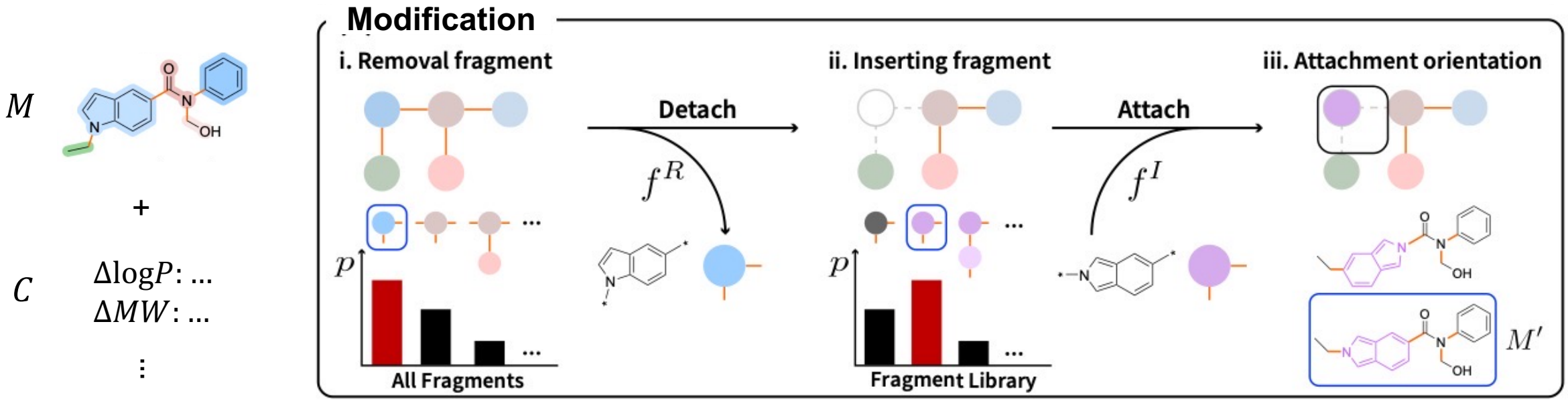


[1] Kim, Hyeongwoo, et al. "DeepBioisostere: Discovering Bioisosteres with Deep Learning for a Fine Control of Multiple Molecular Properties." arXiv preprint arXiv:2403.02706 (2024).

Method

Architecture of DeepBioisostere and training

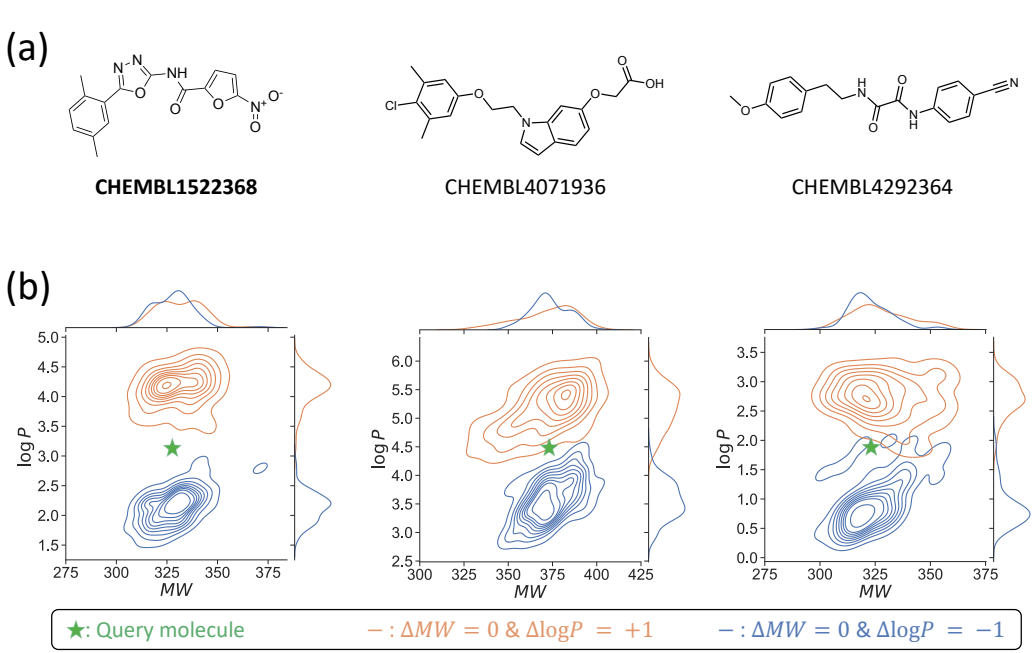
- 20M molecular pairs enumerated from ChEMBL[2], composing of bio-active molecules.
- For each molecular pair, the DeepBioisostere model learns three steps involving the selection of (1) **removal fragment**, (2) **inserting fragment**, and (3) **attachment orientation**.

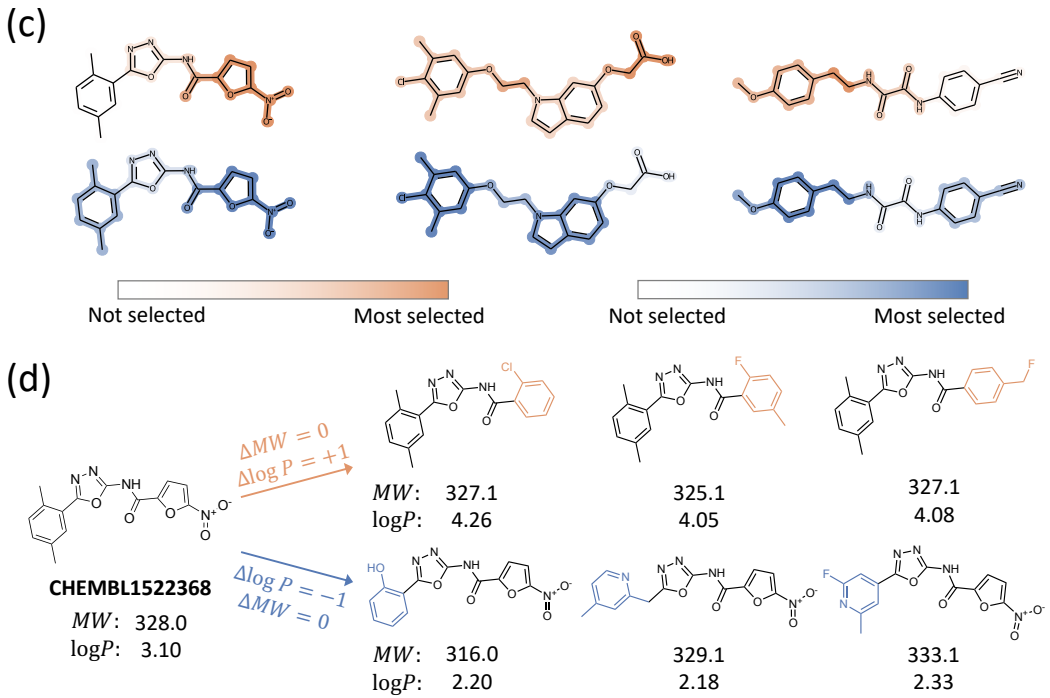


[2] Gaulton, Anna, et al. "The ChEMBL database in 2017." Nucleic acids research 45.D1 (2017): D945-D954.

Result 1

Task 1: Multi-property control scenario

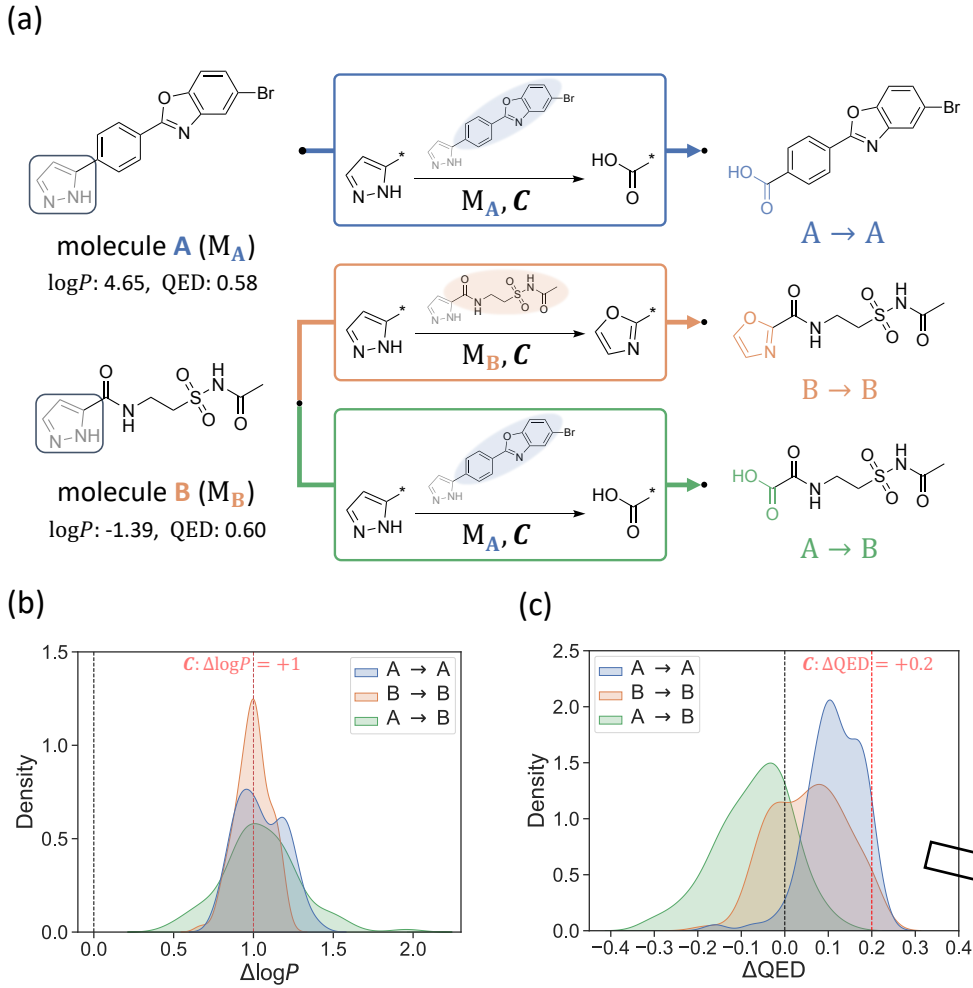




- For C: $\Delta\log P = -1, \Delta\mathrm{MW} = 0$, DeepBioisostere selects a **hydrophobic** moiety and replace it with **hydrophilic** one, decreasing the molecule's $\log P$ value (**more soluble in water**).

Result 2

Task 2: Analysis on selection of insertion fragments



(a) we fixed **pyrazole group** as removal fragments.

- A → A & B → B : DeepBioisostere is directly utilized on molecules A and B.
- A → B: selecting for molecule A and applying to molecule B

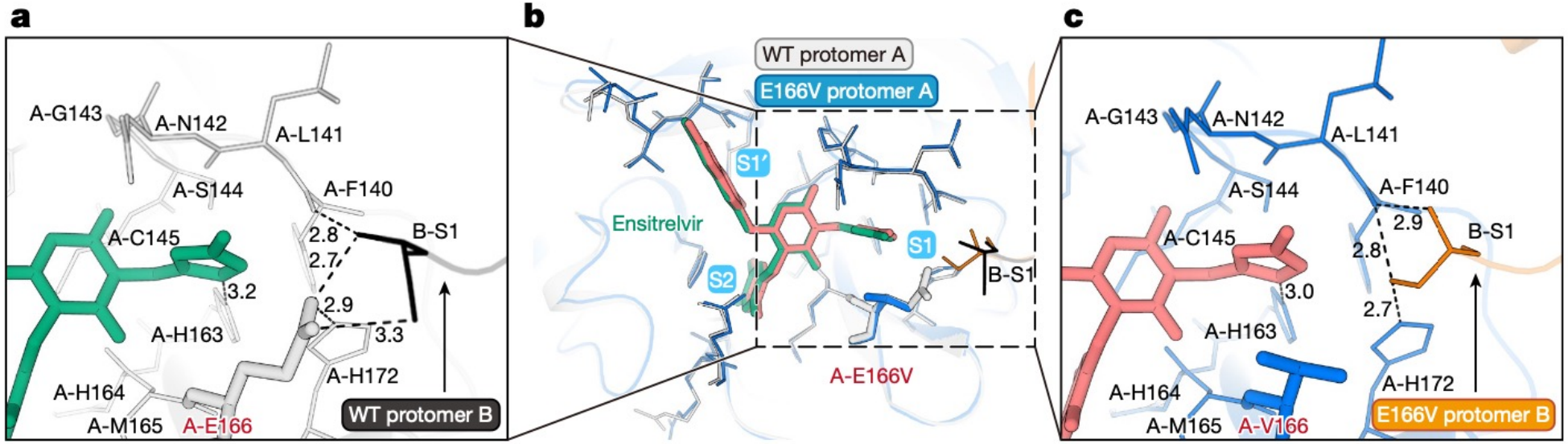
(b) $\Delta MW = 0$ & $\Delta \log P = +1$

(c) $\Delta MW = 0$ & $\Delta QED = +0.2$

A → B fails to increase QED of molecule A.

Result 3

Task 3: Case study on a drug optimization for a mutant with drug-resistance.



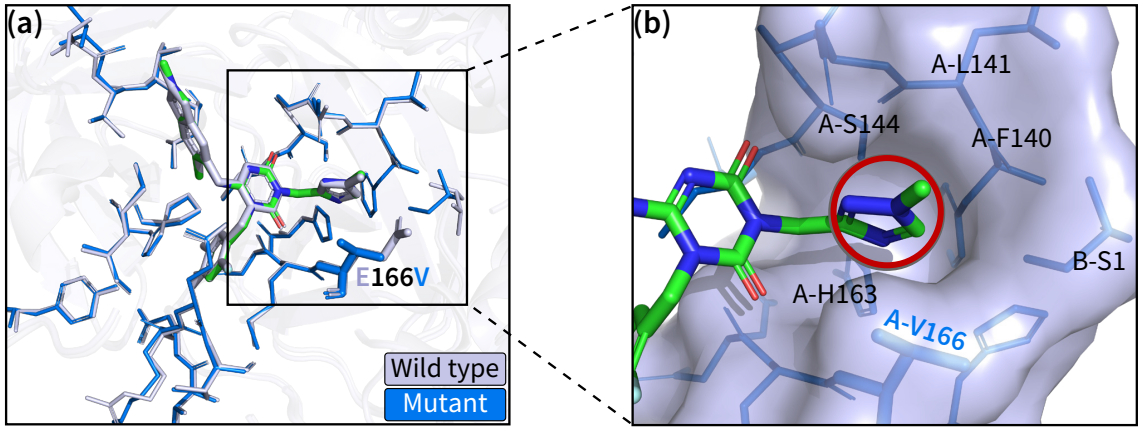
- *Ensitrelvir* is an oral SARS-CoV-2 Main protease inhibitor in clinical study for treating CODIV-19.[3]
- The molecular mechanism of resistance to Ensitrelvir of a mutant E166V has been reported.[4]

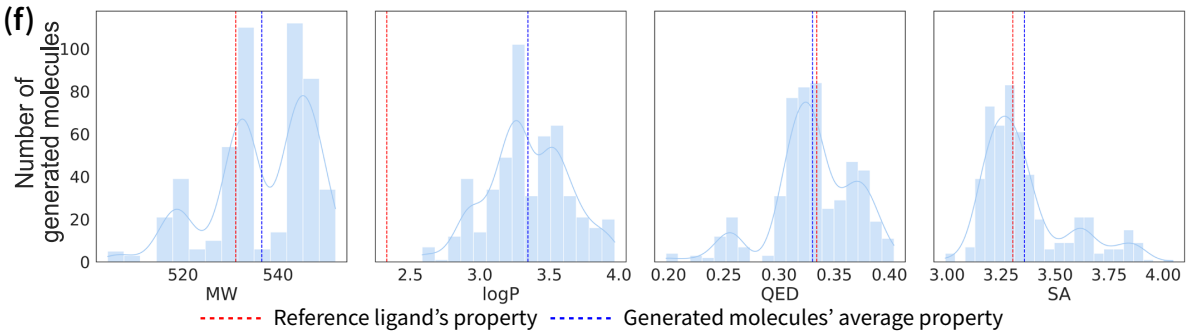
  E: glutamic acid (with minus charge) → V: valine (hydrophobic)

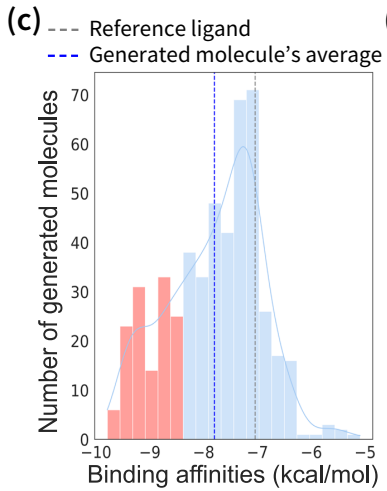
[3] Unoh, Yuto, et al. "Discovery of S-217622, a noncovalent oral SARS-CoV-2 3CL protease inhibitor clinical candidate for treating COVID-19." *Journal of medicinal chemistry* 65.9 (2022): 6499-6512.
[4] Duan, Yinkai, et al. "Molecular mechanisms of SARS-CoV-2 resistance to nirmatrelvir." *Nature* 622.7982 (2023): 376-382.

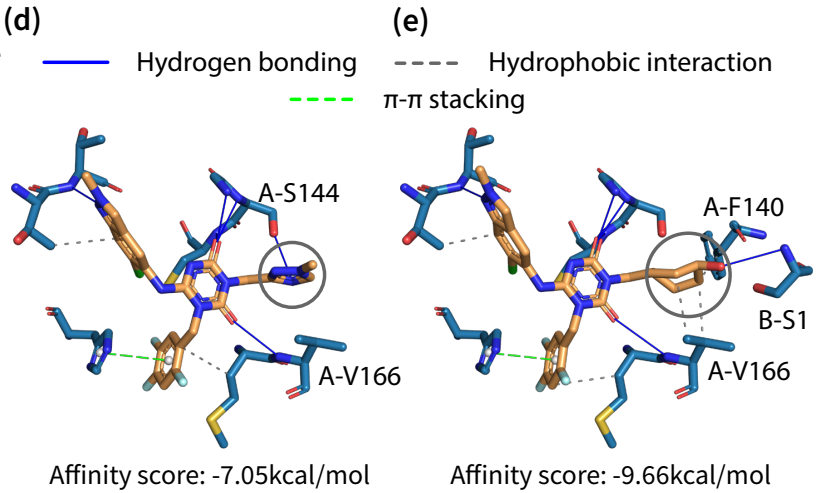
Result 3

Task 3: Case study on a drug optimization for a mutant with drug-resistance.









We designated the triazole moiety as the removal fragment (b) and used property control condition of $\Delta\mathbf{MW} = \mathbf{0}, \Delta\mathbf{log}\,\boldsymbol{P} + \mathbf{1}, \Delta\boldsymbol{SA} = \mathbf{0}, \Delta\mathbf{QED} = \mathbf{0}$.

Generated molecules show better binding affinity scores, evaluated by SMINA[5] (c), while retaining other crucial properties such as QED and SAscore (f).

[5] Koes, D., Baumgartner, M. & Camacho, C. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. J. Chem. Inform. Model. 53, 1893–1904 (2013).

Thanks to...

ICL Team

Prof. Woo Youn Kim

NRF 한국연구재단 National Research Foundation of Korea

e-mail: novainco98@kaist.ac.kr

Visit wooyoun.kaist.ac.kr!

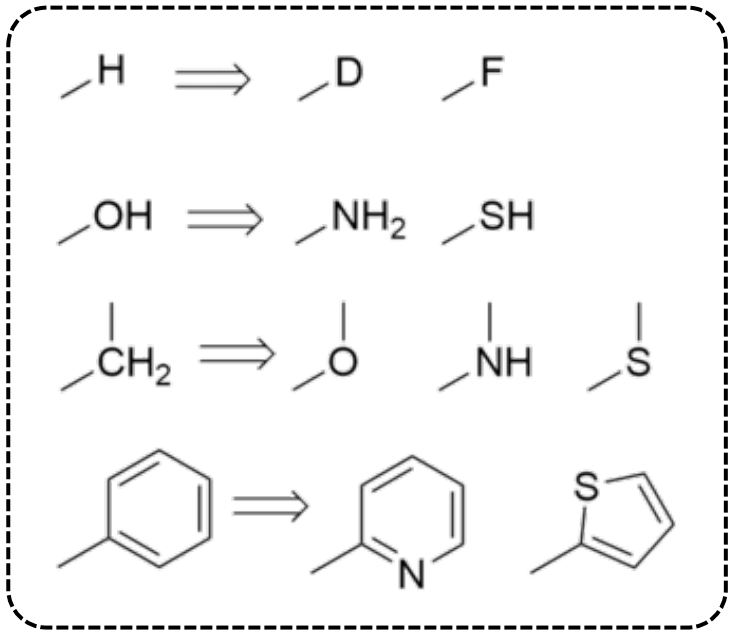
감사합니다.

Appendix

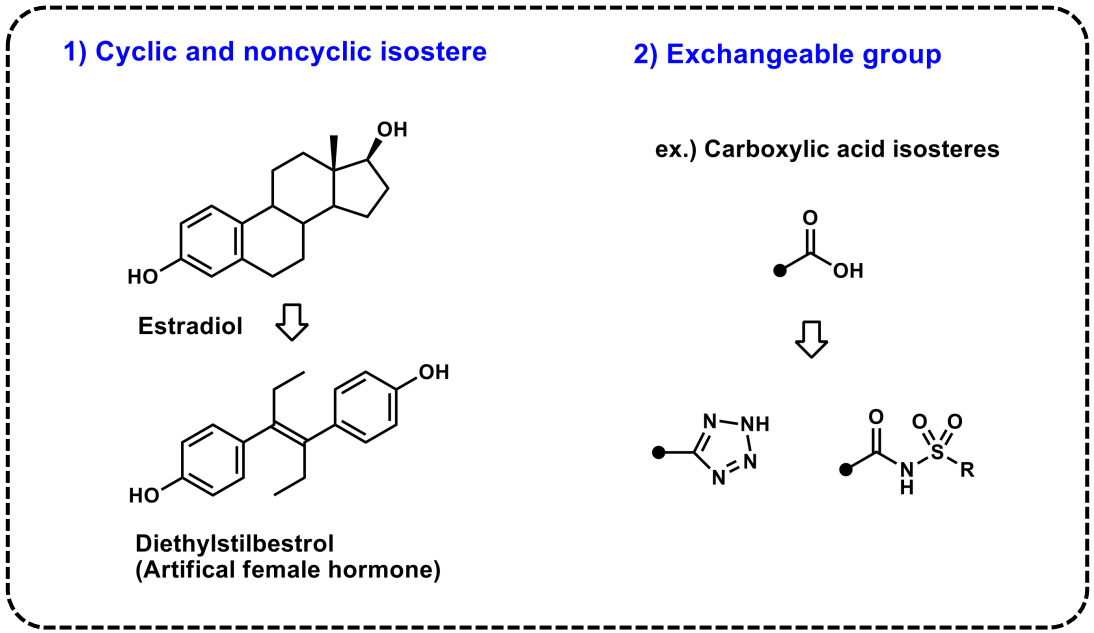
A. Definition of Bioisostere

Bioisosteres

Some examples of classical and non-classical bioisosteres.



Classical

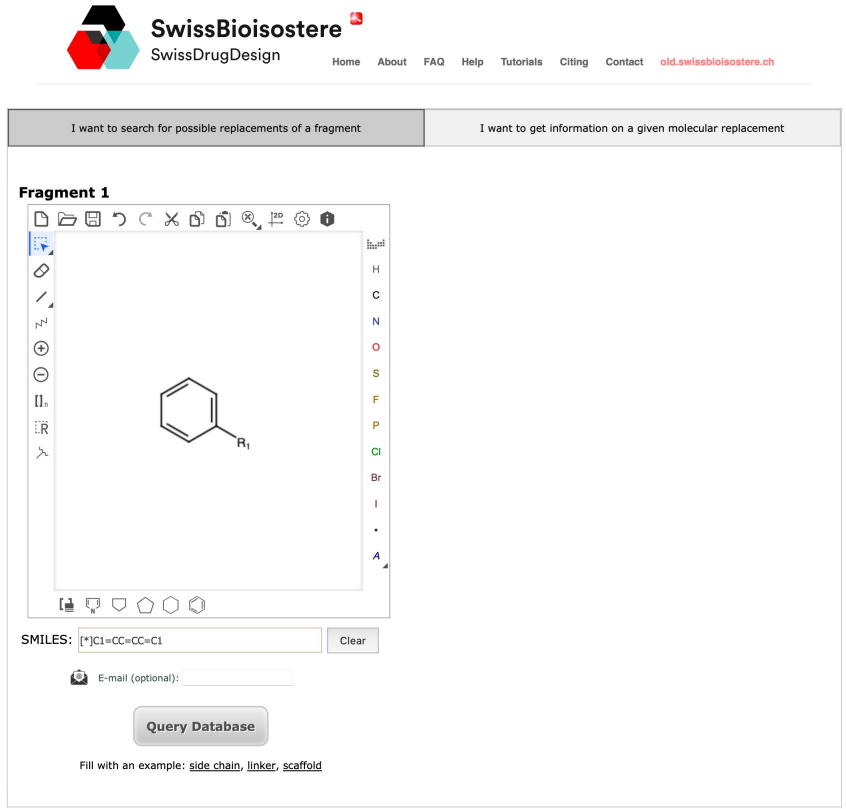


Non-classical

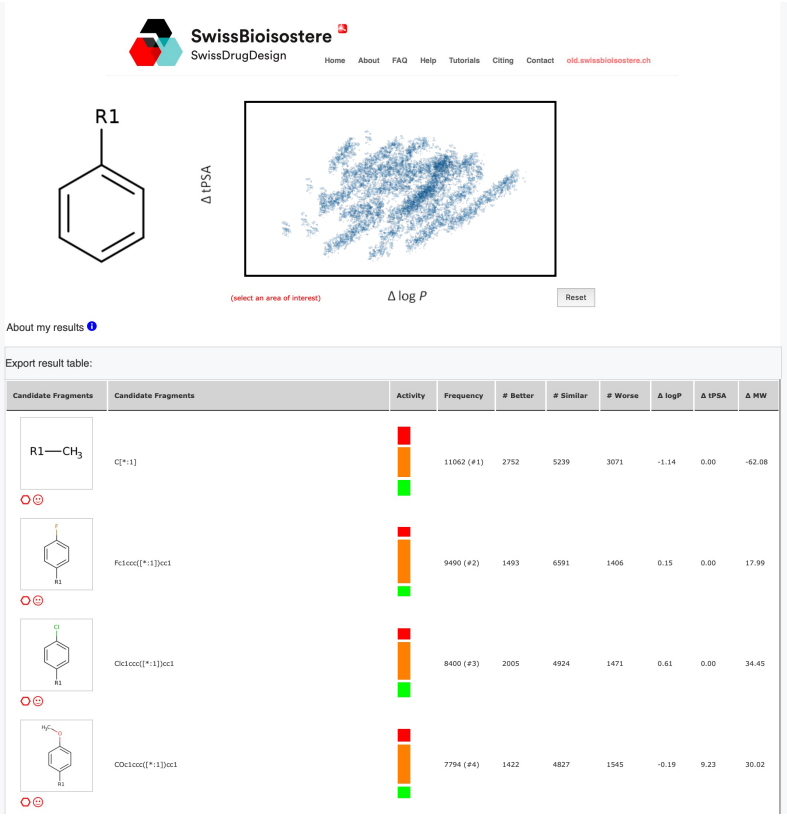
B. In silico bioisosteric replacement

SwissBioisostere

Database-mining approach to discover bioisosteric replacements.



Query



C. Training data / Pre-processing

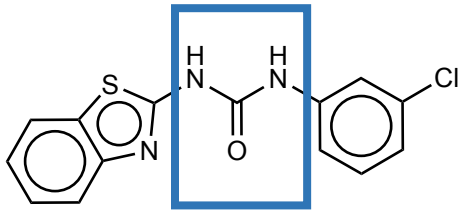
Definition of MMP

1) # of atoms of variable part ≤ 12

2) # of atoms of variable part ≤ # of atoms of common part

ONLY HEAVY ATOMS

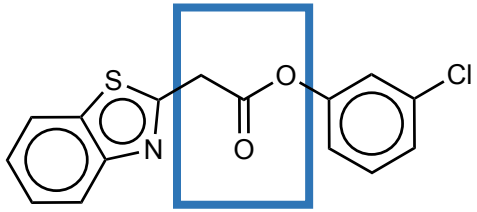
M_1



variable part: 4 < 12
variable part (4) < # common part (16)

PASS!

M_2



variable part: 4 < 12
variable part (4) < # common part (16)

C. Training data / Pre-processing

Definition of MMP

3) Only BRICS bonds are regarded as cuttable bonds.

- Makes bioisosteric replacements designed by our model more **synthetically feasible**.

4) The number of bond cuts is not limited.

- Enables a linker change or **scaffold-hopping**.

C. Training data / Pre-processing

- To see the **generalizability** of our model to unseen bioisosterism, we have split the data into training / val / test while the **insertion fragments are not overlapped**.
- The number of obtained MMP was 10,650,360, split into about 8:1:1.

D. Training objective

- We define a chemical modification task as learning the joint distribution of an original molecule, M, a modified molecular structure, M', and a property control condition, $\boldsymbol{C}$: $p(M, M', \boldsymbol{C})$
- We factorize the joint probability as $p(M, M', \boldsymbol{C}) = p(M, \boldsymbol{C})p(M'|M, \boldsymbol{C})$.

1) $p(M, \boldsymbol{C})$: The likelihood of an original molecule M and a property control condition $\boldsymbol{C}$ for that molecule, arising from former experimental observations.

2) $p(M'|M, \boldsymbol{C})$: conditional probability of a modified structure to achieve the given property control condition $\boldsymbol{C}$ from an original structure M → Training objective of DeepBioisostere

D. Training objective

- $p(M'|M, \boldsymbol{C})$ depends on the bond cleavage rule, $\boldsymbol{R}$ → we denote it as $p_{\boldsymbol{R}}(M'|M, \boldsymbol{C})$
- For an MMP according to $\boldsymbol{R}$, (M, M'), their conditional probability can be factorized with the conditional probabilities of corresponding three modification components:

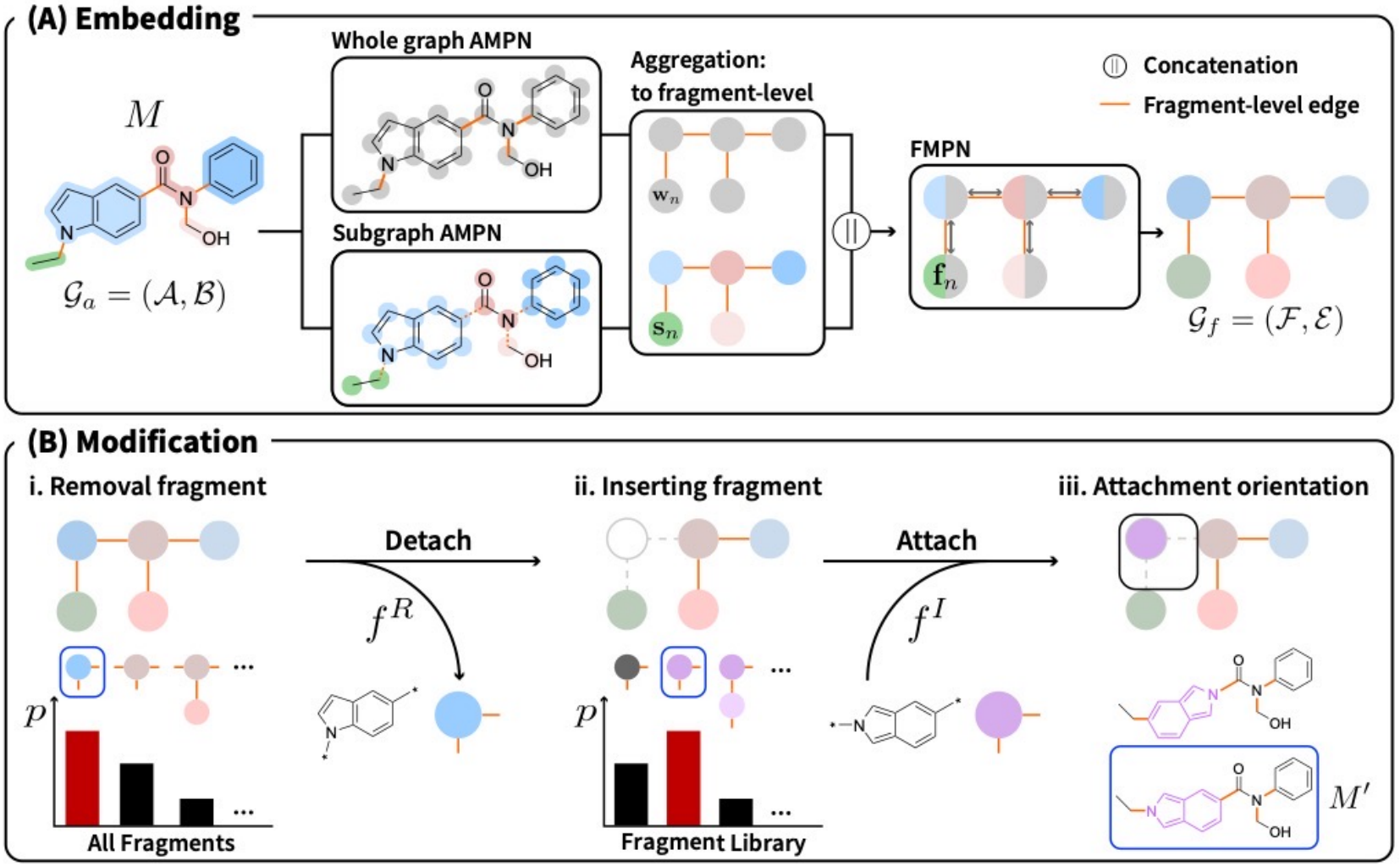
$$p_{\boldsymbol{R}}(M'|M, \boldsymbol{C}) = p_{\boldsymbol{R}}(f^R, f^I, A|M, \boldsymbol{C})$$

Q. (M, M') and (f^R, f^I, A) are uniquely mapped (1:1 mapping)?

$$= p_{\boldsymbol{R}}(f^R|M, \boldsymbol{C}) p_{\boldsymbol{R}}(f^I|M, f^R, \boldsymbol{C}) p_{\boldsymbol{R}}(A|M, f^R, f^I, \boldsymbol{C}),$$

where f^R is a removal fragment, f^I is an insertion fragment, and A is an attachment orientation.

E. Model architecture



AMPN:

Atom-Message Passing Network

FMPN:

Fragment-Message Passing Network

Fragment-level edge:

conforming BRICS rules