

# タンパク質間相互作用阻害剤設計 のための分子生成手法の開発

兒嶋 佑季    小杉 孝嗣    大上 雅史

東京工業大学 情報理工学院 情報工学系



東京工業大学



- 過去数十年の間, 医薬品開発の効率が低下<sup>[1]</sup>.
- 2000年代初頭から新たな治療標的として,  
**タンパク質間相互作用 (Protein-Protein Interaction, PPI)**  
が注目されるようになる<sup>[2]</sup>.  
⇒ 標的とすることが難しい

## ★ 標的化合物の物理化学的特性の違い

### 一般的な低分子医薬品

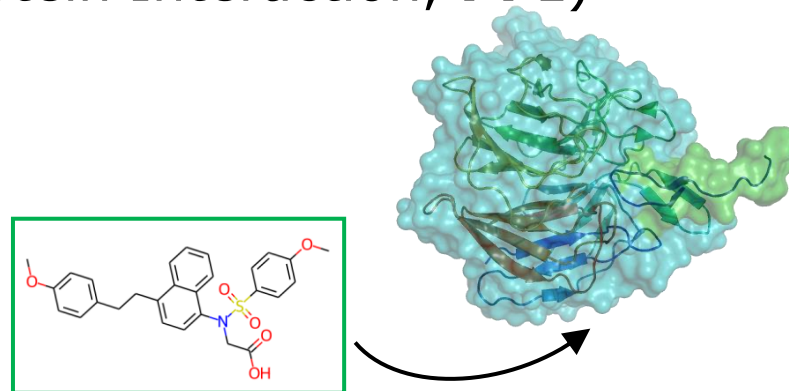
#### Lipinski's "Rule of five (RO5)"<sup>[3]</sup>

- 分子量  $\leq 500$
- 脂溶性を表す指標  $\text{LogP} \leq 5$
- 水素結合ドナーの数  $\leq 5$
- 水素結合アクセプタの数  $\leq 10$

### PPIを標的とする低分子化合物

#### Morelli's "Rule of four (RO4)"<sup>[4]</sup>

- 分子量  $> 400$
- 脂溶性を表す指標  $\text{LogP} > 4$
- 環構造の数  $> 4$
- 水素結合アクセプタの数  $> 4$



[1] Scannell JW, et al. *Nat. Rev. Drug Discov.*, 2012.

[2] Arkin MR, Wells JA. *Nat. Rev. Drug Discov.*, 2004.

[3] Lipinski, CA, et al. *Adv. Drug Deliv. Rev.*, 1997.

[4] Morelli X, et al. *Curr. Opin. Chem. Biol.*, 2011.



- 従来の低分子化合物で構成されるライブラリを用い, PPIを標的としたハイスループットスクリーニング (HTS) を行うとヒット化合物の取得率が著しく低い[5].

→ PPIに特化した化合物ライブラリが必要

- 既存のPPIライブラリ

- Fr-PPIChem[6]
- DLiP[7]
- Enamine PPI Library[8]      等



- 市販の化合物セットから設計されたライブラリ
  - 既知のPPI阻害剤に含まれる母核構造の誘導体を中心
- ⇒ 多様性が限定される傾向にある

[5] Jnoff E, et al. *ChemMedChem*, 2014.

[6] Bosc N, et al. *ACS Chem. Biol.*, 2020.

[7] Ikeda K, et al. [https://skb-insilico.com/static/downloads/DLiP\\_Library\\_Info\\_2020.pdf](https://skb-insilico.com/static/downloads/DLiP_Library_Info_2020.pdf).

[8] Enamine. <https://enamine.net/compound-libraries/targeted-libraries/ppi-library>.



- 既存研究にはなかったこと

- PPIに特化した**バーチャルライブラリ**.
- PPI阻害剤として望ましい特性を持つ分子を、計算機で仮想的に次々に生成。  
→ 生化学実験への効率のよい提案が可能.



新規PPI阻害剤の候補化合物を大規模に探索できる  
**de novo分子生成手法**に着目

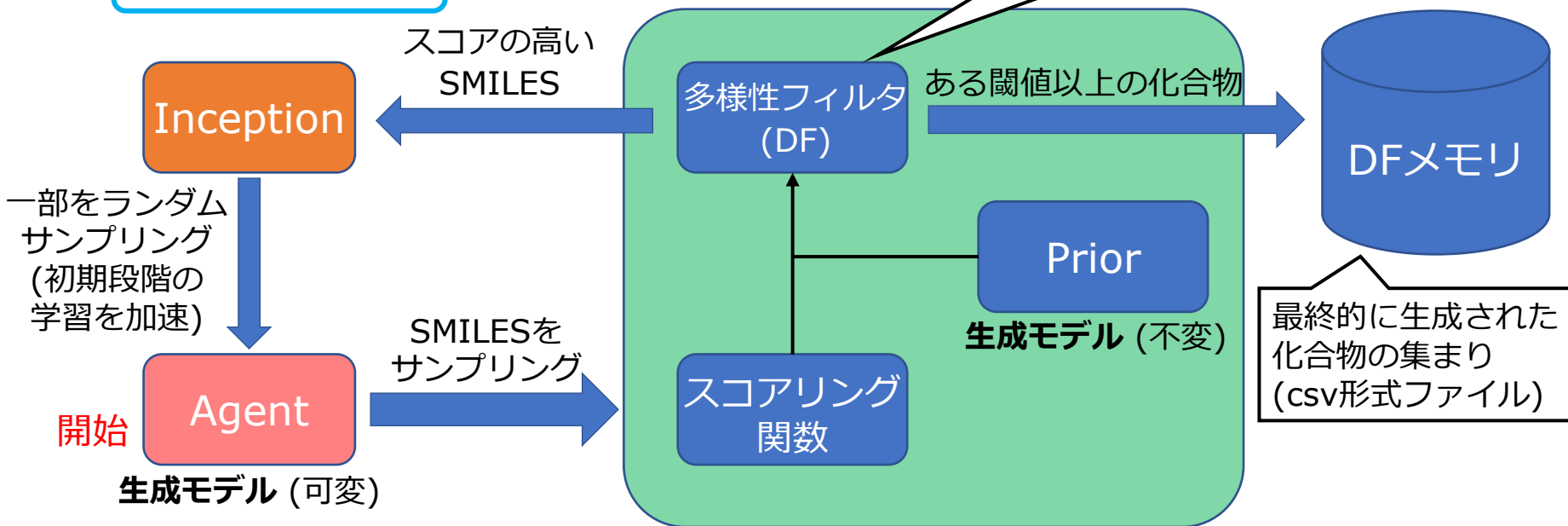
## ■ 研究目的

- 分子生成手法を用いて、新たな探索空間からPPI阻害活性を持つ新規候補化合物を生成する.
- PPI阻害剤としての適正を持つ化合物を収載したバーチャルライブラリの構築.



- 深層強化学習を用いた分子生成ツール<sup>[9]</sup>
  - アーキテクチャ：RNN (LSTM/GRU)

## 強化学習ループ



- 1分子  $x$  に対するスコアリング関数  $S(x) = \frac{\sum_i w_i p_i(x)}{\sum_i w_i}$  ( $\in [0, 1]$ )
  - 各コンポーネント  $p_i(x) \in [0, 1]$ , 重み  $w_i \geq 1$

[9] Blaschke T, et al. "REINVENT 2.0: An AI Tool for De Novo Drug Design", *J. Chem. Inf. Model.*, 2020.

## 1. RO4<sup>[4]</sup>

- 既知のPPI阻害剤データベースの2P2I<sup>[10]</sup>の化合物が以下の経験則によく当てはまった。 ※ 2P2Iは現在は使用不可

### Morelli's "Rule of four (RO4)"

- 分子量 > 400
- 脂溶性を表す指標LogP > 4
- 環構造の数 > 4
- 水素結合アクセプタの数 > 4

## 2. QEPPI<sup>[11]</sup>

- PPI阻害剤としての適正を**定量**評価した指標 ( $\in [0, 1]$ )
- 既知のPPI阻害剤データベースiPPI-DB<sup>[12]</sup>の化合物を学習し、QED<sup>[13]</sup>と同様の手法で設計。

[4] Morelli X, et al. *Curr. Opin. Chem. Biol.*, 2011.

[10] Basse MJ, et al. *Database*, 2016.

[11] Kosugi T, Ohue M. *Int. J. Mol. Sci.*, 2021.

[12] Torchet R, et al. *Bioinformatics*, 2021.

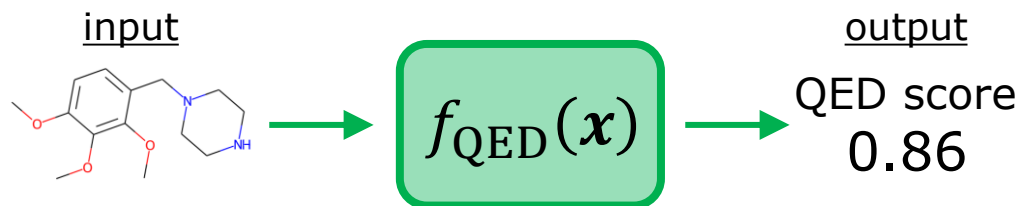
[13] Bickerton GR, et al. *Nat. Chem.*, 2012.



## ● 既存のスコアリング関数

### ➤ QEDスコア

→ そのまま使用可能.



## ● 【提案手法】スコアリング関数の改良

### ① RO4相当スコア

→ そのままでは使用不可.

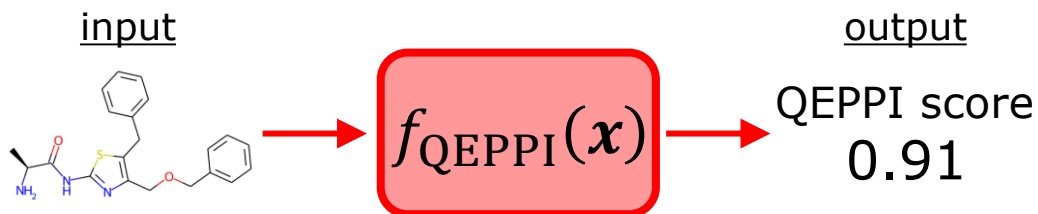
→ ルールを1つ満たしたら0.25加点.

#### Morelli's "Rule of four (RO4)"

- 分子量 > 400
- 脂溶性を表す指標LogP > 4
- 環構造の数 > 4
- 水素結合アクセプタの数 > 4

### ② QEPPIスコア

→ そのまま使用可能.



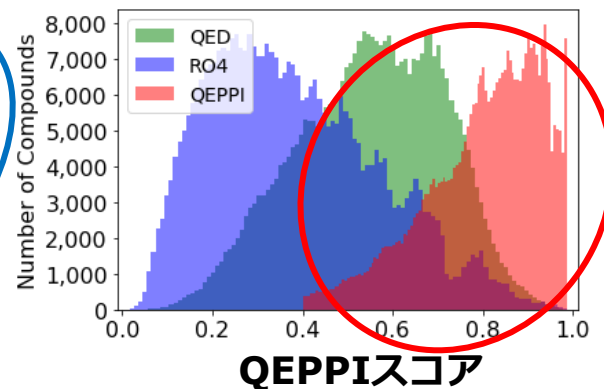
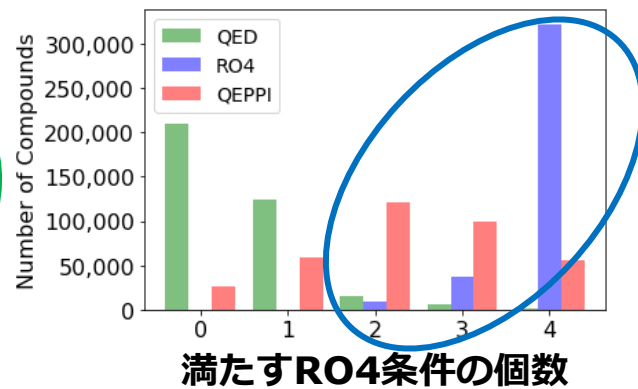
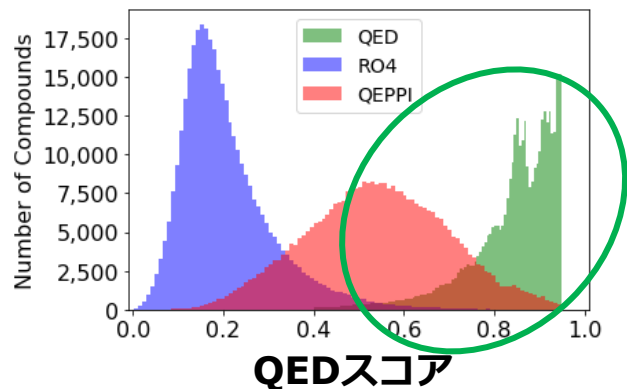


- QED, RO4, QEPPPIの各スコアリング関数を適用し、それぞれ3,000ステップ学習。



生成化合物件数	(既存手法) <b>QED</b> 357,456件	(提案手法1) <b>RO4</b> 368,140件	(提案手法2) <b>QEPPPI</b> 359,722件
---------	-------------------------------	--------------------------------	-----------------------------------

- 各場合の生成化合物群それぞれに対し、QED, RO4, QEPPPIのスコアをすべて計算しスコアごとに分布を確認。



⇒ 各々で適用したスコアが高くなるように正しく学習できている。



## 結果2. 生成後期の化合物の分布

9

生成化合物件数

(既存手法) **QED**  
357,456件

(提案手法1) **RO4**  
368,140件

(提案手法2) **QEPPi**  
359,722件

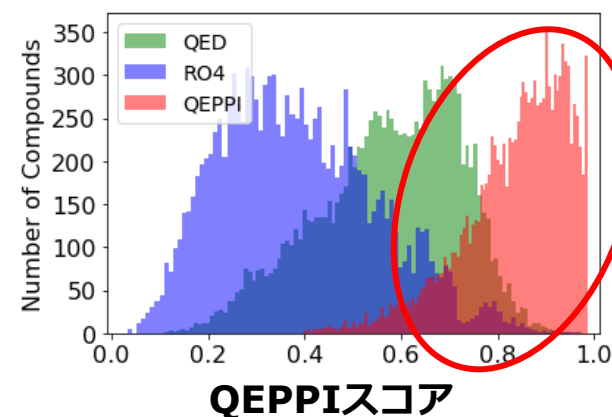
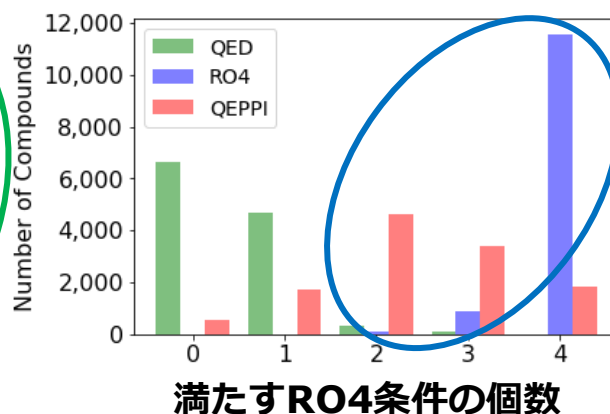
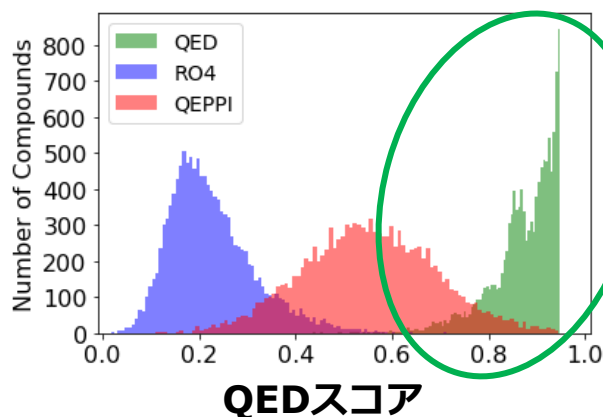
得られた化合物のうちより高いスコアの分子を獲得するために、  
3,000ステップ中の最後の100ステップ実行分 (2,901~3,000ステップ) を抽出。

抽出化合物件数

(既存手法) **QED**  
11,714件

(提案手法1) **RO4**  
12,547件

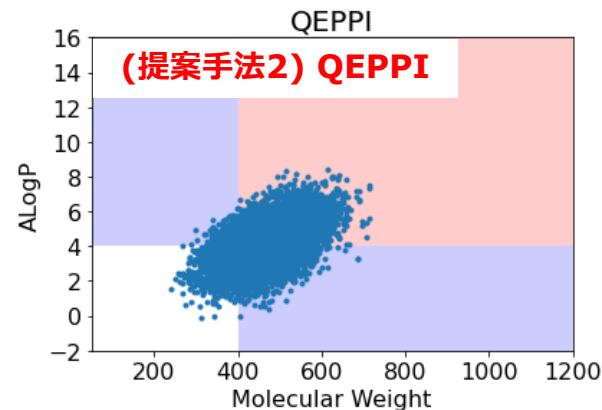
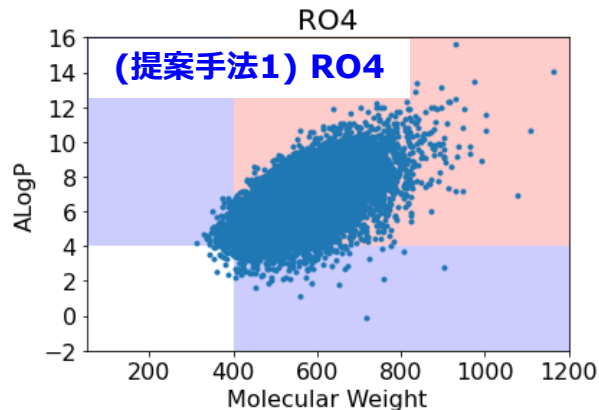
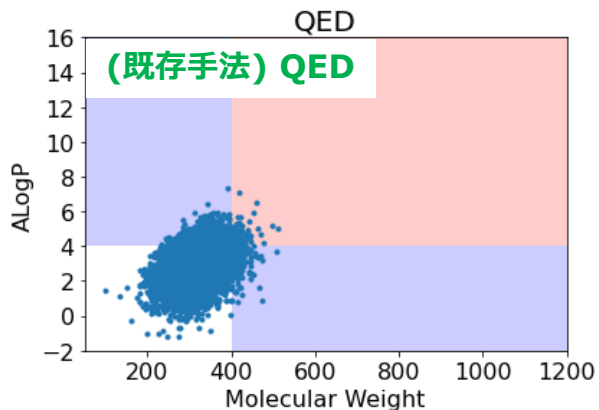
(提案手法2) **QEPPi**  
12,097件



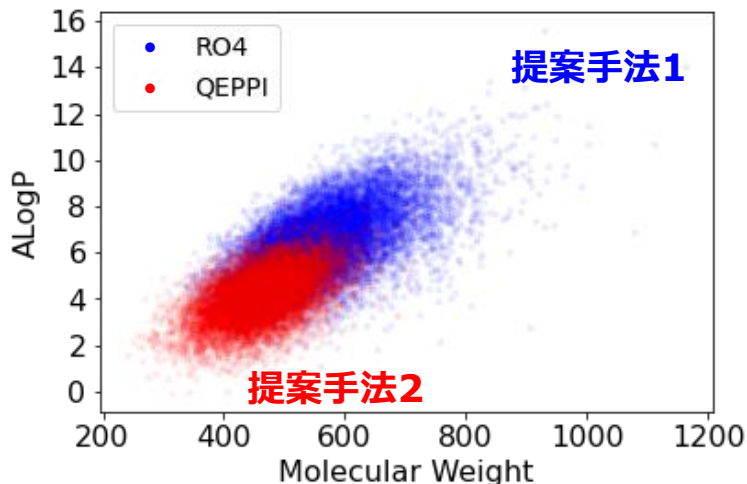
⇒ 全3,000ステップの場合と同様の各スコアの分布



- 各スコアリング関数適用時の分子量-LogP散布図 (横軸: 分子量, 縦軸: LogP)



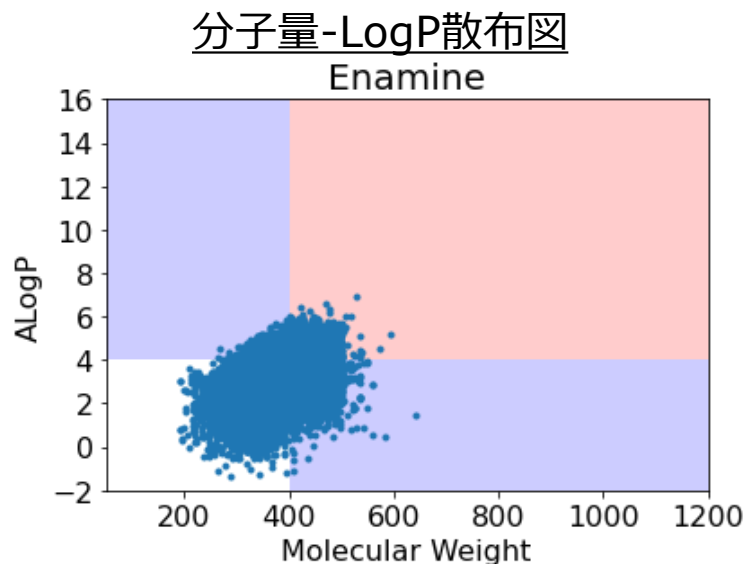
- RO4(青)およびQEPPI(赤)の重ね合わせ散布図



- (既存手法) QED
  - 分子量・LogPともに小さい。
  - 10,547件(90.0%)の分子がRO4の2条件とも満たさない。
- (提案手法1) RO4
  - 12,029件(95.9%)の分子がRO4の2条件をともに満たす。
- (提案手法2) QEPPI
  - 分子量は85.3%, LogPは58.6%の分子がRO4の条件を満たす。
  - RO4のように過度に分子量やLogPが高い分子は生成されていない。

比較対象： **Enamine PPI Library**<sup>[8]</sup>

- 20種類以上のタンパク質複合体の構造から特定の部分構造を抽出.
- その部分構造に対する結合が期待される母核構造を持つリガンドの集まり.



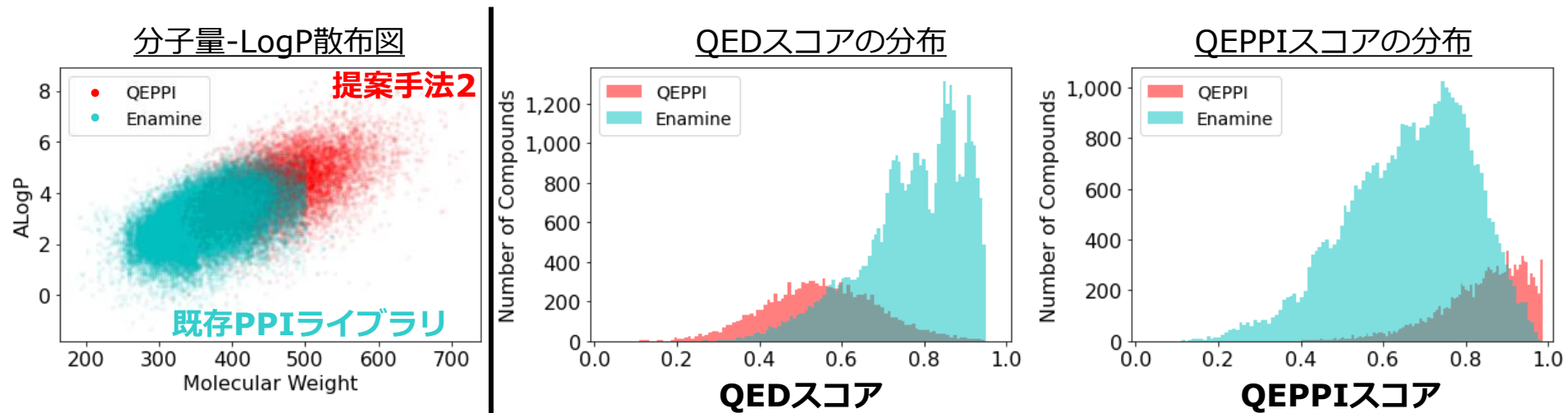
- 2,615件(6.4%)の分子がRO4の2条件をともに満たす.
- 提案手法と比べると割合が低く、QEDと似たような分布.

考察として、以下の2つの化合物集団を比較した.

● **提案手法2 (QEPPi)** に基づく生成後期化合物 (12,097件)

● **Enamine PPI Library** に含まれる全化合物 (40,640件)

[8] Enamine. <https://enamine.net/compound-libraries/targeted-libraries/ppi-library>.



## ● 提案手法2 (QEPPi) に基づく生成化合物

- より多様な化合物空間を探索できている。
- QEPPiが高い (= 近年のPPI阻害剤の傾向により適合した化合物の生成ができている[11]) 。

Enamineと同様、提案手法2に基づく生成化合物にもPAINSフィルタ<sup>[14]</sup>をかけた。

⇒ 最終的に、11,252件をバーチャルライブラリに収載した。

[11] Kosugi T, Ohue M. *Int. J. Mol. Sci.*, 2021.

[14] Baell JB, Holloway GA. *J. Med. Chem.*, 2010.



## ■ 結論

- REINVENTでは各スコアリング関数が高くなるように正しく学習することができた。特にQEPPPIを適用した場合に、より多様な化合物空間を探索できている。
- 本研究で設計したバーチャルライブラリは、既存のPPIライブラリを補完したものと考えられる。

## ■ 今後の課題

- 生成化合物の多様性の拡大
  - 他の既存PPIライブラリとの比較
  - 新たなスコアリング関数の提案
- 生成化合物の合成可能性についての更なる検討  
→ 将来的に効率よくスクリーニングを行えるような有用なライブラリの構築

## 謝辞

本研究は以下の支援を受けて行われた。

- JST ACT-X (No. JPMJAX20A3)
- JSPS科研費 基盤研究(B) (No. 20H04280)
- 公益財団法人 みずほ学術振興財団