

## Drug repurposing based on network

### ชื่อสมาชิกในกลุ่ม (Team members)

1. มารีนญา	ตะโจปะรัง	6231352421
2. นิชกานต์	ชัยพจนา	6231322621
3. ดรากรณ์	ผดุงพัฒน์นอม	6231323221

## Progress3

เนื่องจากเราต้องการหาความเป็นไปได้ของสูตรยาใหม่ ๆ สำหรับการรับมือโควิด-19 จึงเริ่มจากการหาข้อมูลของยาที่สามารถยับยั้งโรคโควิด-19 ได้ ซึ่งยาที่หามาได้จะมีประสิทธิภาพที่ต่างกัน โดยจะแบ่งคลาสของยาได้จากค่า IC50 ว่าจะต้องใช้ปริมาณยามากเท่าไร ในการยับยั้งโปรตีนโควิด-19 จากนั้นจะคำนวณค่า QED และ LogP ของยาเก็บไว้เพื่อใช้เป็นผลลัพธ์ที่ควรจะได้จากการ generate ยาใด ๆ

ใน progress 3 จะลอง generate ยาโดยใช้ข้อมูลจาก zinc250k เพื่อ generate ยาโดยไม่มีการปรับค่า parameter ต่าง ๆ แล้วลองคำนวณค่า QED และ LogP ที่ได้จากยานั้น ๆ เพื่อดูแนวโน้มและความเป็นไปได้ของยาประเภทต่าง ๆ เทียบกับข้อมูลของยาที่เก็บไว้เบื้องต้น

## ขั้นตอนการทำงาน

### Prepare data

เก็บข้อมูลยาตัวที่ส่งผลต่อ covid-19 จาก chembl database และนำมา filter โดยใช้ standard type เป็น IC50 เพื่อแบ่งกลุ่มยาเหล่านั้นออกเป็น 3 กลุ่ม ได้แก่ active, inactive, intermediate

	molecule_chembl_id	canonical_smiles	standard_value	bioactivity_class
0	CHEMBL187579	<chem>Cc1noc(C)c1CN1C(=O)C(=O)c2cc(C#N)ccc21</chem>	7200.0	intermediate
1	CHEMBL188487	<chem>O=C1C(=O)N(Cc2ccc(F)cc2Cl)c2ccc(I)cc21</chem>	9400.0	intermediate
2	CHEMBL185698	<chem>O=C1C(=O)N(CC2COc3ccccc3O2)c2ccc(I)cc21</chem>	13500.0	inactive
3	CHEMBL426082	<chem>O=C1C(=O)N(Cc2cc3ccccc3s2)c2ccccc21</chem>	13110.0	inactive
4	CHEMBL187717	<chem>O=C1C(=O)N(Cc2cc3ccccc3s2)c2c1cccc2[N+](=O)[O-]</chem>	2000.0	intermediate
...	...	...	...	...
128	CHEMBL2146517	<chem>COC(=O)[C@@]1(C)CCCc2c1ccc1c2C(=O)C(=O)c2c(C)c...</chem>	10600.0	inactive
129	CHEMBL187460	<chem>C[C@H]1COC2=C1C(=O)C(=O)c1c2ccc2c1CCCC2(C)C</chem>	10100.0	inactive
130	CHEMBL363535	<chem>Cc1coc2c1C(=O)C(=O)c1c-2ccc2c(C)cccc12</chem>	11500.0	inactive
131	CHEMBL227075	<chem>Cc1cccc2c3c(ccc12)C1=C(C(=O)C3=O)[C@@H](C)CO1</chem>	10700.0	inactive
132	CHEMBL45830	<chem>CC(C)C1=Cc2ccc3c(c2C(=O)C1=O)CCCC3(C)C</chem>	78900.0	inactive

### Calculate QED and LogP

หลังจากได้ข้อมูลยาที่แบ่งกลุ่มแล้ว จะนำมาคำนวณค่า QED และ LogP โดยใช้ rdkit  
หลังจากได้ค่า QED และ LogP ทุกตัวมา จะนำไปคิดค่าทางสถิติต่าง ๆ เพื่อใช้เปรียบเทียบกับยาอื่น ๆ  
โดยเก็บในลักษณะดังนี้

	molecule_chembl_id	canonical_smiles	standard_value	bioactivity_class	MW	LogP	NumHDonors	NumHAcceptors	QED
0	CHEMBL187579	<chem>Cc1noc(C)c1CN1C(=O)C(=O)c2cc(C#N)ccc21</chem>	7200.0	intermediate	281.271	1.89262	0.0	5.0	0.757559
1	CHEMBL188487	<chem>O=C1C(=O)N(Cc2ccc(F)cc2Cl)c2ccc(I)cc21</chem>	9400.0	intermediate	415.589	3.81320	0.0	2.0	0.487042
2	CHEMBL185698	<chem>O=C1C(=O)N(CC2COC3CCCCC3O2)c2ccc(I)cc21</chem>	13500.0	inactive	421.190	2.66050	0.0	4.0	0.485762
3	CHEMBL426082	<chem>O=C1C(=O)N(Cc2cc3cccc3s2)c2cccc21</chem>	13110.0	inactive	293.347	3.63080	0.0	3.0	0.683944
4	CHEMBL187717	<chem>O=C1C(=O)N(Cc2cc3cccc3s2)c2c1cccc2[N+](=O)[O-]</chem>	2000.0	intermediate	338.344	3.53900	0.0	5.0	0.348717

	count	mean	std	min	25%	50%	75%	max
bioactivity_class								
active	15.0	3.777084	1.056293	2.4107	2.81550	3.70040	4.314350	6.10120
inactive	104.0	3.969391	1.455105	-0.0554	3.22032	3.99644	4.808425	7.05154
intermediate	14.0	3.594304	1.165892	1.1272	3.31852	3.68656	4.025030	6.30470

### ค่าทางสถิติของ LogP

	count	mean	std	min	25%	50%	75%	max
bioactivity_class								
active	15.0	0.627761	0.156704	0.206778	0.613257	0.674688	0.731628	0.766180
inactive	104.0	0.465817	0.197367	0.035506	0.305394	0.438330	0.650963	0.924827
intermediate	14.0	0.569472	0.157902	0.286913	0.490015	0.553261	0.664966	0.862078

### ค่าทางสถิติของ QED

## Generate drug

ใช้ torchdrug เพื่อ generate ยา โดยใช้ RGCN เป็น model และใช้ GCPN ในการ generate ยาใหม่ขึ้น และใช้ dataset เป็น zinc250k เมื่อได้ผลของยาในรูปของ smile แล้วจะนำมาคำนวณค่า QED และ LogP ของยานั้น ๆ เพื่อใช้เปรียบเทียบกับค่า QED และ LogP ของยาที่มีประสิทธิภาพต่อ โพรตีนโควิด-19

```
count      32.000000
mean        0.708784
std         0.152630
min         0.242727
25%         0.667417
50%         0.740529
75%         0.820531
max         0.893178
Name: QED, dtype: float64
```

ค่าทางสถิติของ QED

```
count      32.000000
mean        3.021954
std         1.036792
min         0.995300
25%         2.135275
50%         2.895135
75%         3.785300
max         5.013800
Name: LogP, dtype: float64
```

ค่าทางสถิติของ LogP

prepare data:

[https://colab.research.google.com/drive/1\\_yQ0BwggR7-VzNRf1AYGomxAgebN1OUO?usp=sharing](https://colab.research.google.com/drive/1_yQ0BwggR7-VzNRf1AYGomxAgebN1OUO?usp=sharing)

generate and calculate new QED and LogP:

[https://colab.research.google.com/drive/1\\_VOMIIS19Mbv-HVsR0vpHycg2OGUC9AJ?usp=sharing](https://colab.research.google.com/drive/1_VOMIIS19Mbv-HVsR0vpHycg2OGUC9AJ?usp=sharing)

Training data:

[https://colab.research.google.com/drive/1O4i2mhDOw7XJgjOB5JiW0luOIFCWNM\\_t?usp=sharing](https://colab.research.google.com/drive/1O4i2mhDOw7XJgjOB5JiW0luOIFCWNM_t?usp=sharing)

data folder:

[https://drive.google.com/drive/folders/1eMmlsfq7Ga0Op\\_GJsSaFi9deOJvqX746?usp=sharing](https://drive.google.com/drive/folders/1eMmlsfq7Ga0Op_GJsSaFi9deOJvqX746?usp=sharing)

ใน progress ถัดไปจะเป็นการ tune parameter เพื่อให้ได้ผลลัพธ์ที่ดีขึ้น และอาจลองวัดผลโดยใช้ binding affinity