DCGAN-QM 微型抗体的设计

L.X.ZHOU

lxzhou1939@hotmail.com

本工作的主体由博士研究生 Allison M. Rossetto 完成

綱要

1. 设计思想
2. 蛋白质的全电子结构的量子力学计算
3. 深度卷积的生成对抗网络DCGAN
4. 数据库
5. 计算设计
6. 抗原的电子结构计算结果
7. 微型抗体的设计结果

1. 设计思想

一切疫苗都是促使身体的浆细胞产生抗体。机体产生的抗体是很大的Y形的蛋白质。但我们现在可以通过QM，DCGAN和THPdb数据库可以产生较大的多肽作为微型抗体. 这些跟PD1,PDL1,CTLA-4和covid-19结合得很稳定。这属于蛋白质-蛋白质相互作用，通过电导率的计算，作为抗原的PD1等的电导率提升百倍，成为半导体，从而改变了性质。

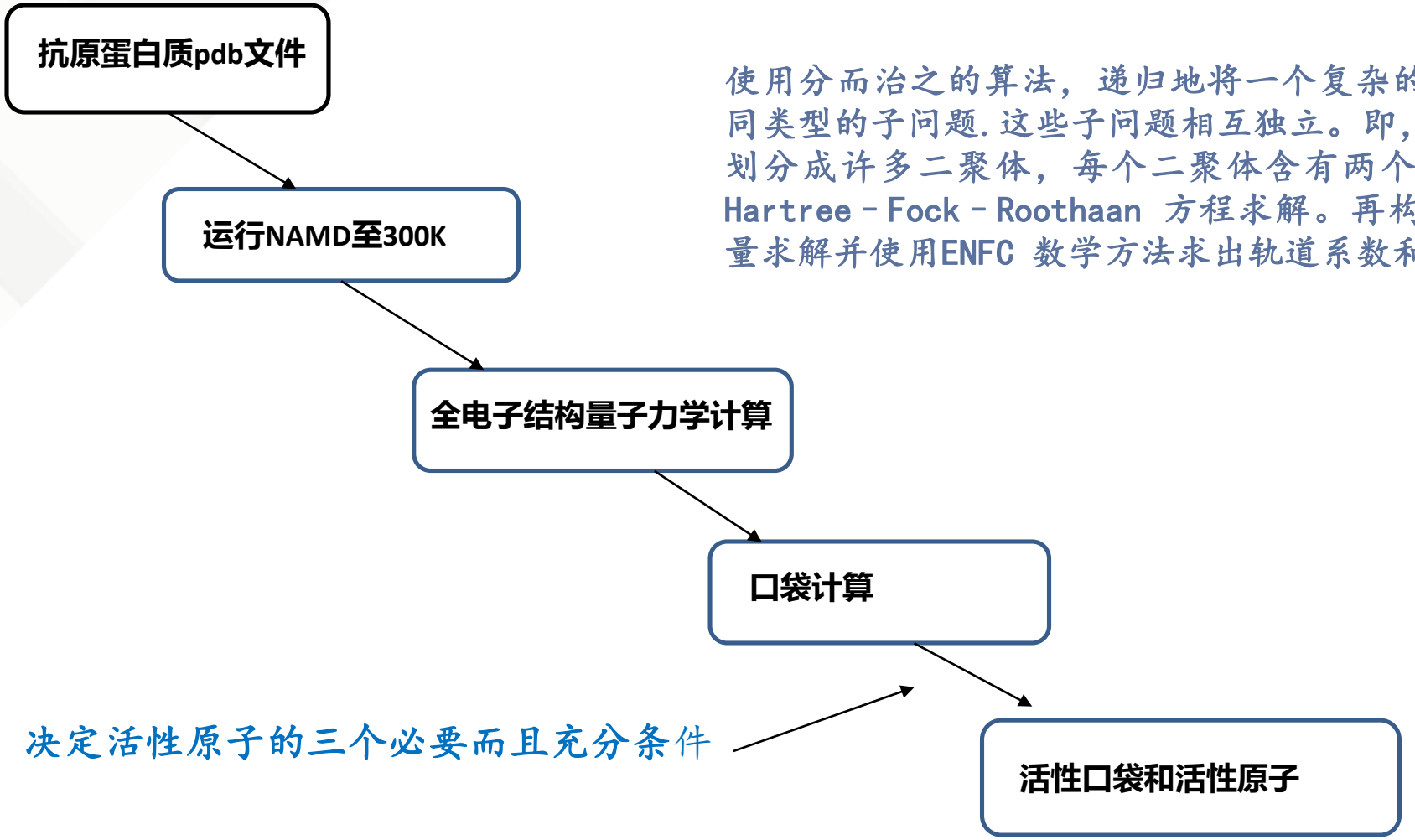
1. 使用全电子结构的量子力学计算来确定抗原蛋白质的活性口袋和活性原子，以便确定抗原-抗体可能发生相互作用的位置。

2. 使用THPdb和FDA数据库,收集和验证药物和蛋白质目标数据.

3. 使用两个DCGAN网络：Squence和Structure产生微型抗体的氨基酸序列和结构.

4. 用pyDockWEB对接抗原和抗体.

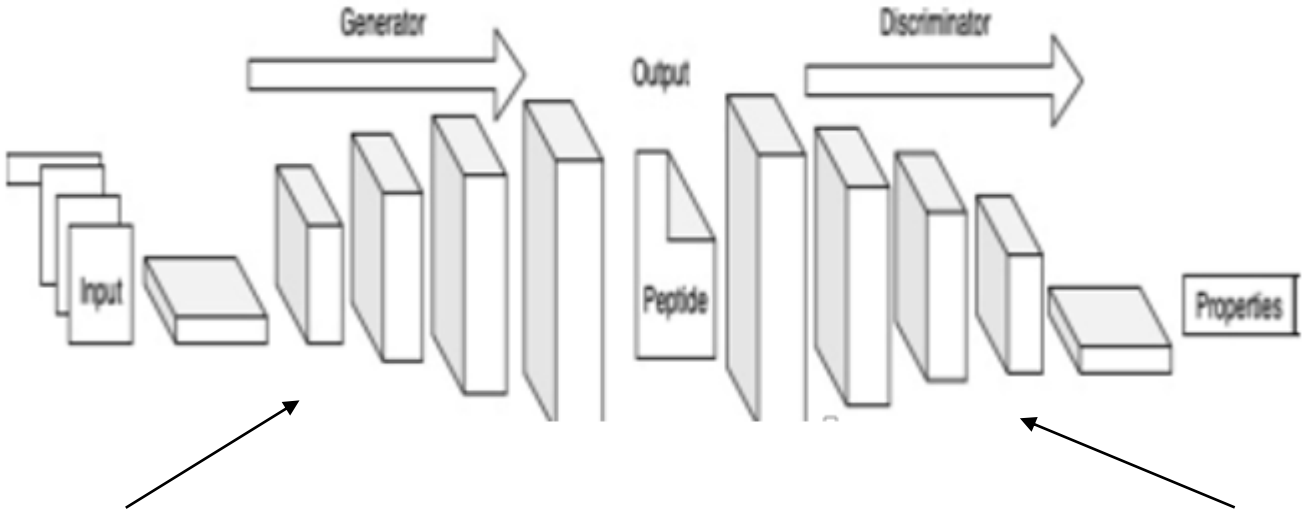
2. 抗原蛋白质的全电子量子力学计算



使用分而治之的算法，递归地将一个复杂的问题分解为许多相同类型的子问题.这些子问题相互独立。即，蛋白质的整个链被划分成许多二聚体，每个二聚体含有两个氨基酸残基。使用Hartree－Fock－Roothaan 方程求解。再构造了完整的哈密顿量求解并使用ENFC 数学方法求出轨道系数和本征值.

决定活性原子的三个必要而且充分条件

3. 深度卷积生成对抗网络DCGAN

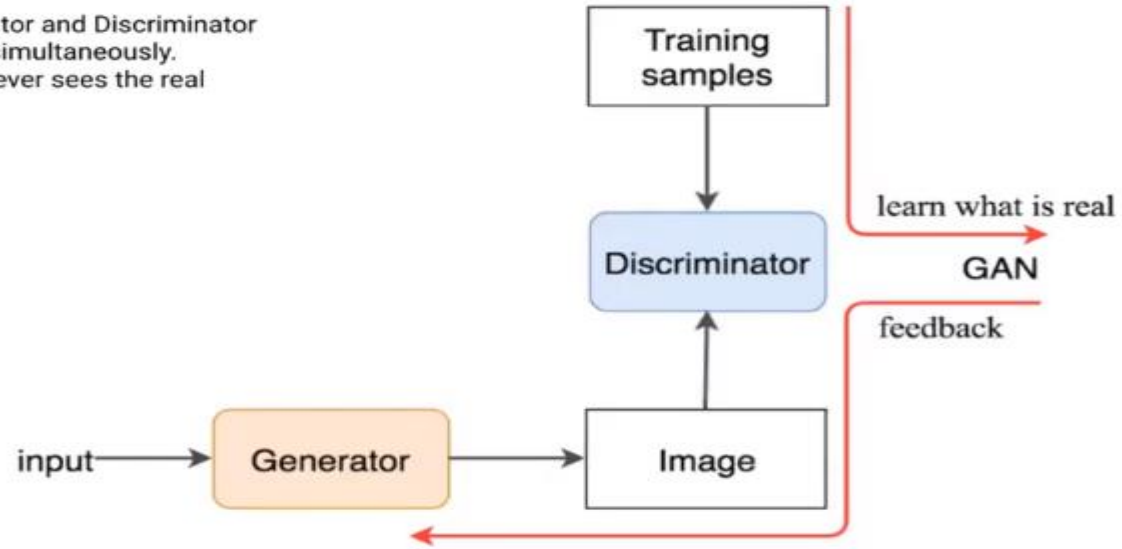


生成网络: 反卷积, 输入噪声, 输出图像

鉴别网络: 卷积,输入图像,输出判断

Training a GAN

- Both Generator and Discriminator are trained simultaneously.
- Generator never sees the real dataset.



Training Sample: Dataset and feature CSV

Main codes: model.py, train.py, prepare.py, predict.py

4. 数据库

THPdb: Therapeutic Protein Database

包含有关这些肽的重要信息，例如它们的描述、序列等。

FDA: 美国食品药物管理局（U.S. Food and Drug Administration）

1. 序列数据：

序列网络使用 FDA 批准的药物及其靶标以及活性原子信息和结合能进行训练。包含药物和目标口袋序列以及结合亲和力和活性原子。您还需要提供目标蛋白的 pdb 文件。

样品输入药物文件：

R94 F95 (目标口袋活性残基)
NH2、CD1 CE1 CZ (目标口袋活性原子)
EDLMKASNLQVSLH (目标口袋序列)
-10.825 (绑定亲和力)
0 (等电点，如果未知写为0)
0 (疏水性（如果已知），如果未知，则将行写为 0)
Y95 (药物的活性残基)
CD2 CE2 CZ (药物的活性原子)
FTGALTTGAKTFPLLNSSGALTSSSVSVPSSSLGTKTYTCNVDHKP (药物序列)

结构数据：

3D 结构网络使用大约 25000 个人类蛋白质结构的数据集进行训练。选择符合以下标准的结构：

人肽，大小在 5 到 250 之间， 残留物，无改性或合成残留物， 无 DNA 或 RNA 附着

需要做几个步骤来简化我们的数据，以便更容易地生成新的肽结构。

1. 将所有原子的坐标归一化以(0,0,0)为中心

2、将PDB文件转换为XYZ坐标文件：

3. 从 XYZ 文件中删除文件名和原子标签数量

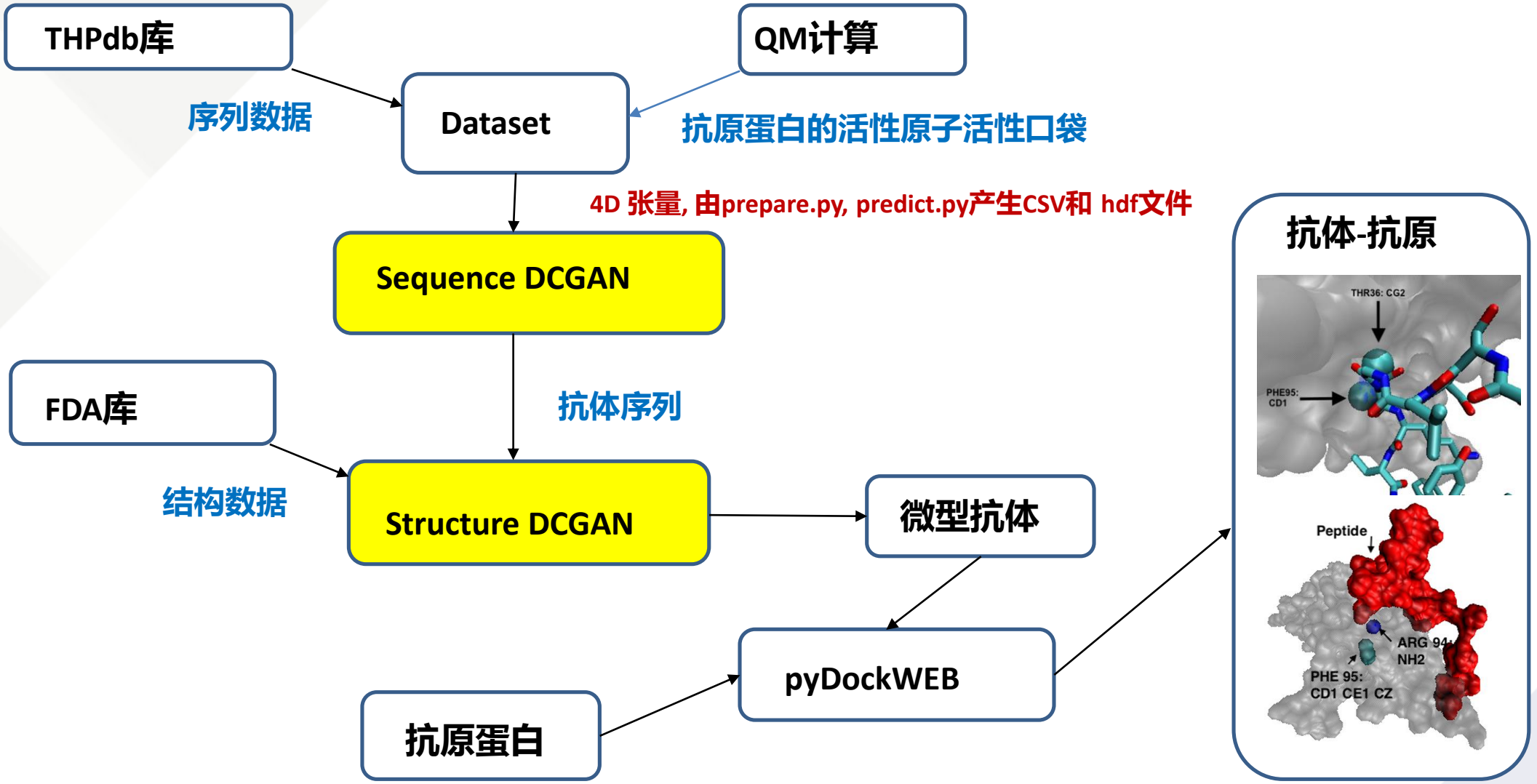
4. 删除多余的空格，在标记之间留下一个空格

5. 附加蛋白质的序列作为最终修改后的 XYZ 文件的第一行

训练结构生成网络的输入的示例：

```
C 7.61800 -11.41100 13.00000
C 7.82800 -12.62800 12.11300
O 8.64300 -12.60000 11.19200
C 8.55100 -11.39300 14.20900
C 8.47500 -8.07000 14.97600
C 9.00400 -8.16200 16.39200
O 9.39700 -7.15600 16.98700
N 9.00100 -11.36700 17.94600
N 7.04500 -13.67500 12.36100
C 7.10700 -14.90100 11.57000
```

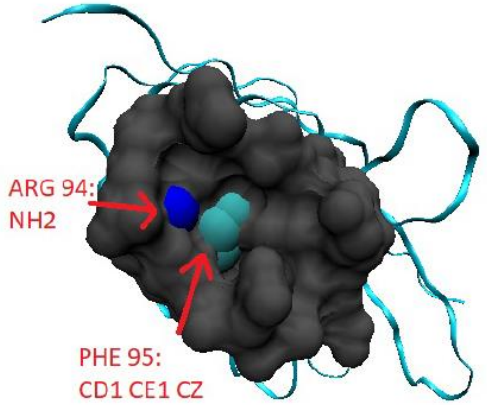
5. 计算设计


THPdb库
QM计算
序列数据
Dataset
抗原蛋白的活性原子活性口袋
4D 张量, 由prepare.py, predict.py产生CSV和 hdf文件
Sequence DCGAN
FDA库
抗体序列
结构数据
Structure DCGAN
微型抗体
pyDockWEB
抗原蛋白
抗体-抗原
THR36: CG2
PHE95: CD1
Peptide
ARG 94
NH2
PHE 95:
CD1 CE1 CZ

6. 抗原的电子结构计算结果

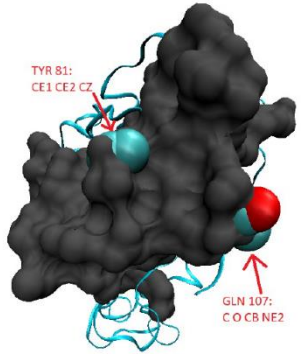
1. PD1

Number	Residue	Active Atoms	Energy Band (eV)	Wave Function
3569	ARG 94	NH2	2.77872	0.993
3555	PHE 95	CD1 CE1 CZ	1.77858	0.993



2. PDL1

Number	Residue	Active Atoms	Energy Band (eV)	Wave Function
6495	TYR 81	CE1,CE2,CZ	1.98602	0.996
6523	GLN 107	C, O, CB, NE2	3.13548	0.966

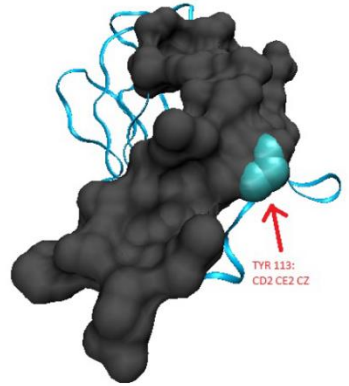


3. CTLA-4

Residue	Active Atoms	Energy Band (eV)	Wave Function
TYR 113	CD2 **CE2** CZ	1.86103	0.9736

4. ACE

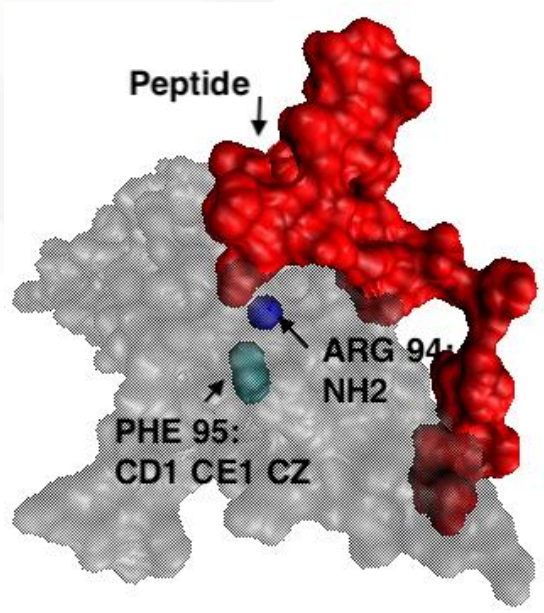
Active Residue	Active Atoms	Energy Band (eV)	Wave Function
LEU 85	N,C,O,CB	0.995	1.24681



7. 微型抗体的设计结果

RESULTS: EFFECTS OF TRANSFER LEARNING ON PD-1

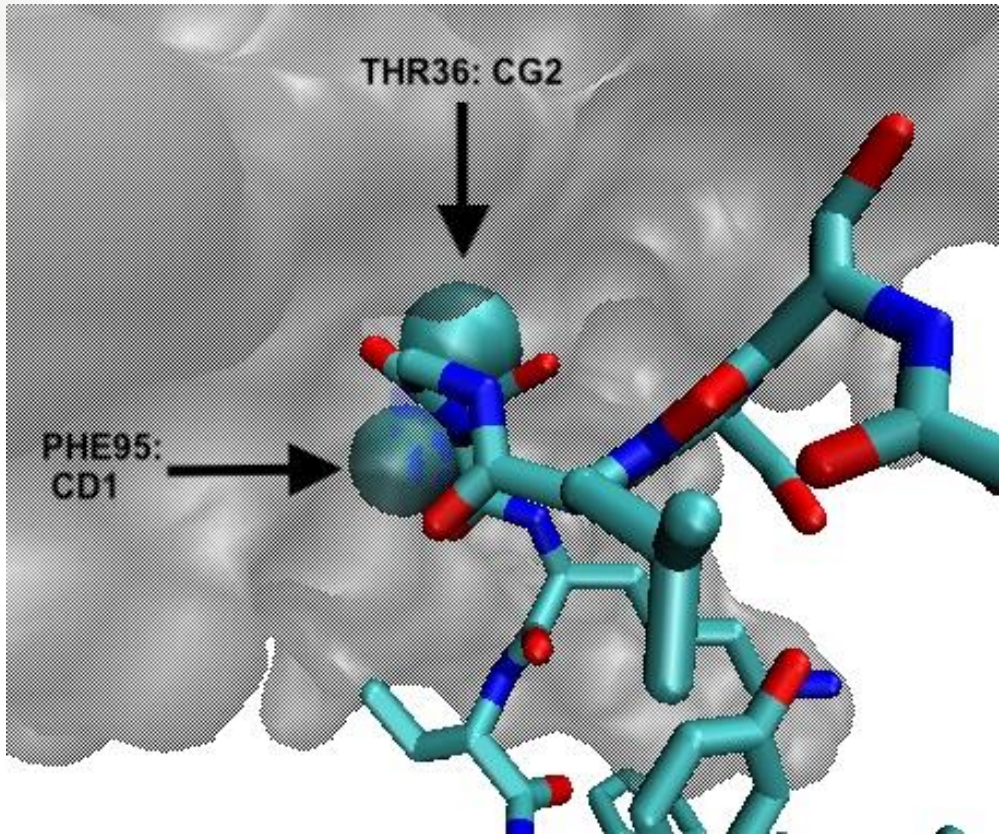
- PD-1 transfer learning results



Rank	Predicted Binding Affinity	Sequence
1	-18.140	TATTTTTITETTAGTATRTRRHTGTRAHHKTTTTTDHD
2	-17.637	ATTGDARITRTGARATATADTTTDTADNATEDTDKTR
3	-16.733	GTTDTTTRATEGAGTTTTTATRTHHATAGTTKTRT
4	-16.684	TTTTEADDRDTTTTTDTTDTHAKRDTGKDDTDTAD
5	-16.644	TEDTGAKGRTDTDTTRTGGARATGHDATRTTHTHG
6	-16.633	TTTRDATFTDKATTTTTTAGHTDGTARTATAAEAH
7	-15.549	AATAAHADTTTTAEATRDGTTTRTTATATTKTT
8	-15.279	ATKADTTTTATTATTTDTTTGTTAGGGTHDAT
9	-14.732	HHTATTGTFADTFAKTKDTDTKREAATATAT
10	-11.927	ADDRDTTTTTDTTDTHAKRDTGKDD

RESULTS: MULTI-GAN PD-1 RESULTS

- PD-1 generated peptide results



GANDALF Peptide Analysis	
Number of residues	45
Most Active Atoms	LYS 11: **NZ**
Distance (nM)	3.012
Docking Binding Affinity (kCal/Mol)	-9.191
Predicted Binding Affinity (kCal/Mol)	-10.231
Sequence	YPGGIRLGDRVFGVEV EEEVDDLPRVGGIIFGV VIILVYEVGGEV

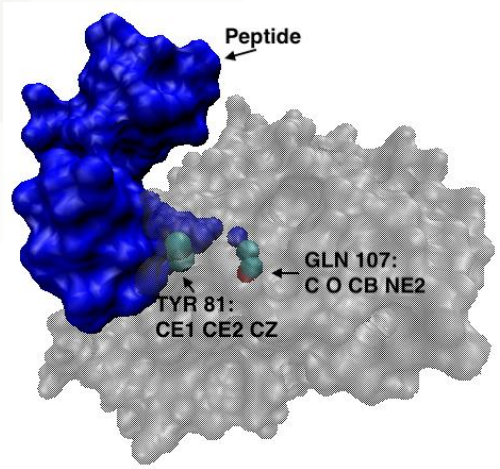
RESULTS: MULTI-GAN PD-1 RESULTS

- PD-1 generated peptide results

Comparison of Best Generated Peptide and FDA Approved drugs for PD-1					
	Binding Affinity	Distance(nM)	Percent Difference of Binding Affinity	SD of Distances	Similarity
Best Generated Peptide	-9.191	3.012	0.00	0.00	100.00
Nivolumab	-10.825	14.249	-1.634	7.946	69.57
Pembrolizumab	-18.810	6.090	-9.619	2.176	69.57

RESULTS: EFFECTS OF TRANSFER LEARNING ON PDL-1

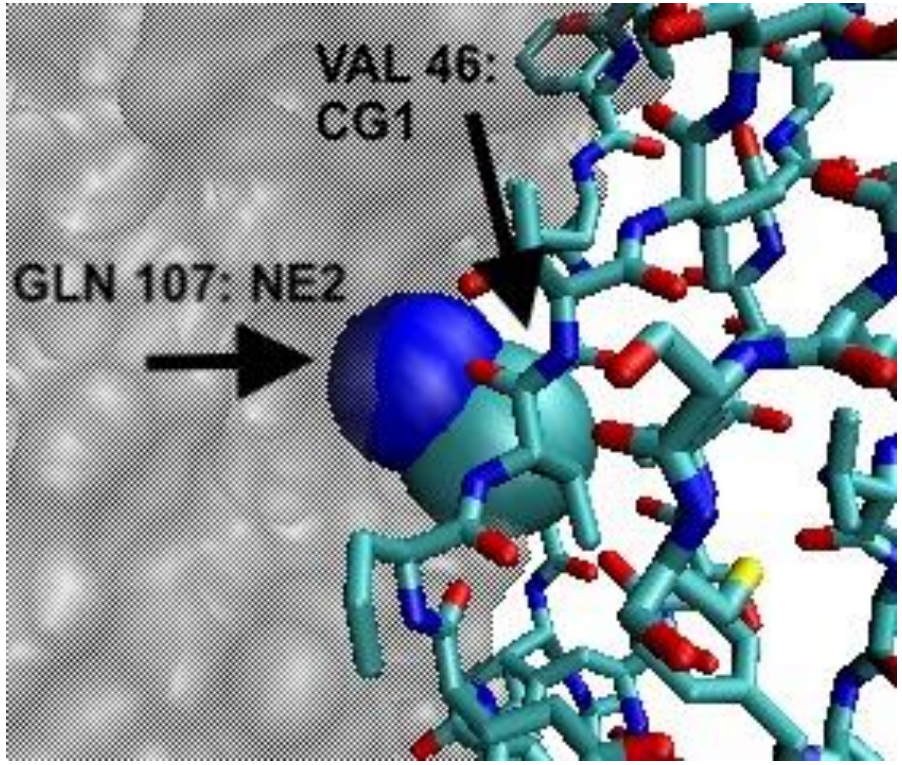
- PDL-1 transfer learning results



Rank	Predicted Binding Affinity	Sequence
1	-20.885	ATTTTTHRATDTGTATDTEAKAATGTADRTDTDRTTHTTTATGT
2	-20.463	ATTKTTADKTTTATTTATDATTRTRTTTTATAGTHTRATGART
3	-19.992	TGTTTGTATTTDTTTTTTTRTDADTGTTAADTGDKTEEAEDT
4	-19.550	AAAITTTAEATATRTRTRREDGTTDTAATTARRETAHTAAE
5	-19.128	EKATTTTTGTGRTDTGTGTATHGKTTTTTTATTTRATTTT
6	-19.055	RATETTGATAKGAHGTTATGATKTTTATTDHDTTGTGTAD
7	-18.994	TTTTKTATTTRTTTTADATTTKTTTGTAATRTTATAGTRD
8	-17.138	TTTTTETGDATTAATKATRTGAAADTETTRATKRRE
9	-16.716	TTTTTTGGIIATETRTTATDTHTARTETRKAAAAA
10	-14.276	TTDTTTTTTTRTDADTGTTAADTGDKTEEA

RESULTS: MULTI-GAN PDL-1 RESULTS

- PDL-1 generated peptide results



GANDALF Peptide Analysis	
Number of residues	78
Most Active Atoms	VAL 46: N **CG1**
Distance (nM)	4.197
Docking Binding Affinity (kCal/Mol)	−22.227
Predicted Binding Affinity (kCal/Mol)	−25.582
Sequence	VFVGVLFEYGDRVGFPDDV VIVLDVIDGIDDVERILVPPG PIFLELFEYGDRVGFPDDVV VLDVIDGIDDVERILVGG

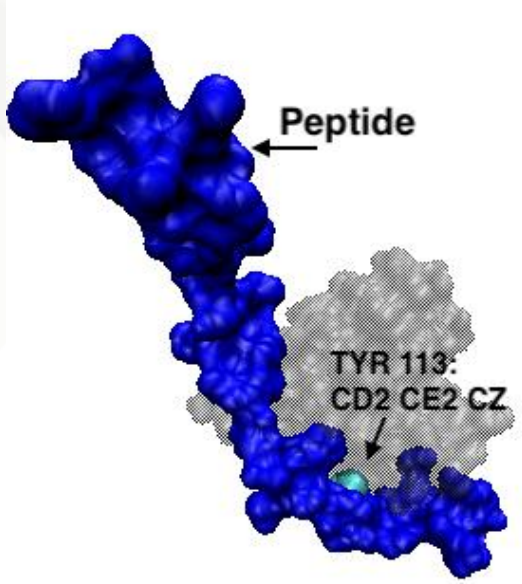
RESULTS: MULTI-GAN PDL-1 RESULTS

- PDL-1 generated peptide results

Comparison of Best Generated Peptide and FDA Approved drugs for PDL1					
	Binding Affinity	Distance(nM)	Percent Difference of Binding Affinity	SD of Distances	Similarity
Generated Peptide	-22.277	4.197	0.000	0.000	100.00
Avelumab	-10.581	4.264	11.696	0.047	89.61
Atezolizumab	-13.775	5.118	8.502	0.651	88.31
Durvalumab	-21.347	4.579	0.931	0.270	88.31

RESULTS: EFFECTS OF TRANSFER LEARNING ON CTLA-4

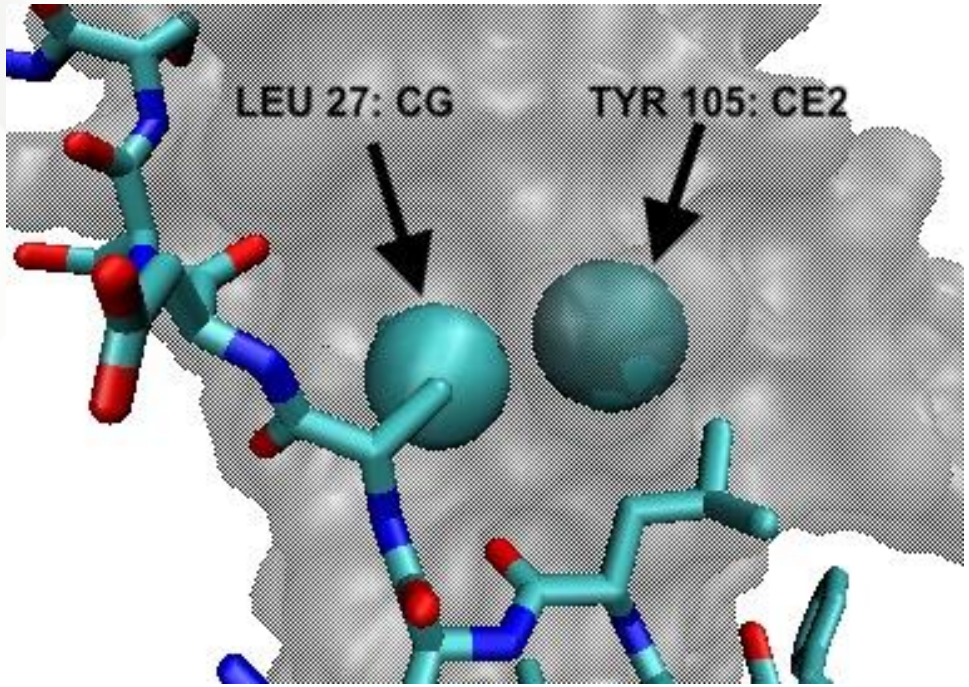
- CTLA-4 transfer learning results



Rank	Predicted Binding Affinity	Sequence
1	-19.999	RTDTKTTTTRETKTTTTTKRKARTAKTTKTTATGTTTATEAT
2	-19.461	TAHTHARTTTRRTTATDTTATTTRGTAATRTTKAGHTAHKT
3	-19.430	TTETRGDTAAKTRAATRTKKKTGTDATRTDDTTRATTTTTR
4	-19.069	TRTDTTTDTTTTTDDAKKDTTGDTTKGTTDRETTAHAADT
5	-18.619	ATRATTTDTATDRATTTADTTTVATRATTDTTDHDTGTA
6	-18.553	TRDTTRTARTTGTKTTATDDEADTATAHTTTTAATRKDA
7	-17.928	KARTETRTTETKGTADAADRAAATHRTTTTTTTTTIKAD
8	-17.164	AITTATIGRKAKTDRKKTATKTRATAAATAGTTTDT
9	-15.183	TTETAATTTTDITTKKAHTTTDGAETTTARAT
10	-13.392	TDTTTTTDDAKKDTTGDTTKGTTDRETT

RESULTS: MULTI-GAN CTLA-4 RESULTS

- CTLA-4 generated peptide results



GANDALF Peptide Analysis	
Number of residues	49
Most Active Atoms	LEU 27: **CG**
Distance (nM)	8.180
Docking Binding Affinity (kCal/Mol)	-25.402
Predicted Binding Affinity (kCal/Mol)	-25.221
Sequence	GRDIRGAEYELDRVVG RYLVIAVFVIGIVGIYVVI FGFGGIRAYLFDGA

RESULTS: MULTI-GAN CTLA-4 RESULTS

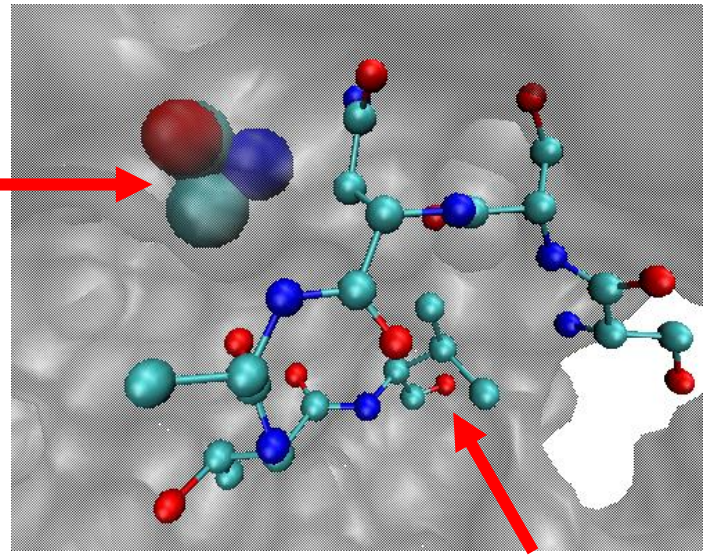
- CTLA-4 generated peptide results

Comparison of Best Generated Peptide and FDA Approved drugs for CTLA-4					
	Binding Affinity	Distance(nM)	Percent Difference of Binding Affinity	SD of Distances	Similarity
Generated Peptide	-25.402	8.1796	0.000	0.000	100.00
Ipilimumab	-31.720	14.249	-6.318	4.292	26.67
Tremelimumab	-38.208	6.090	-12.806	1.478	26.67

RESULTS: COVID-19 TARGETS

- ACE2 top 5 generated peptides
- Comparable peptide inhibitor has a docking binding affinity of -19.843.
  - Larger at 26 residues

LEU 85 N, C, O, CB

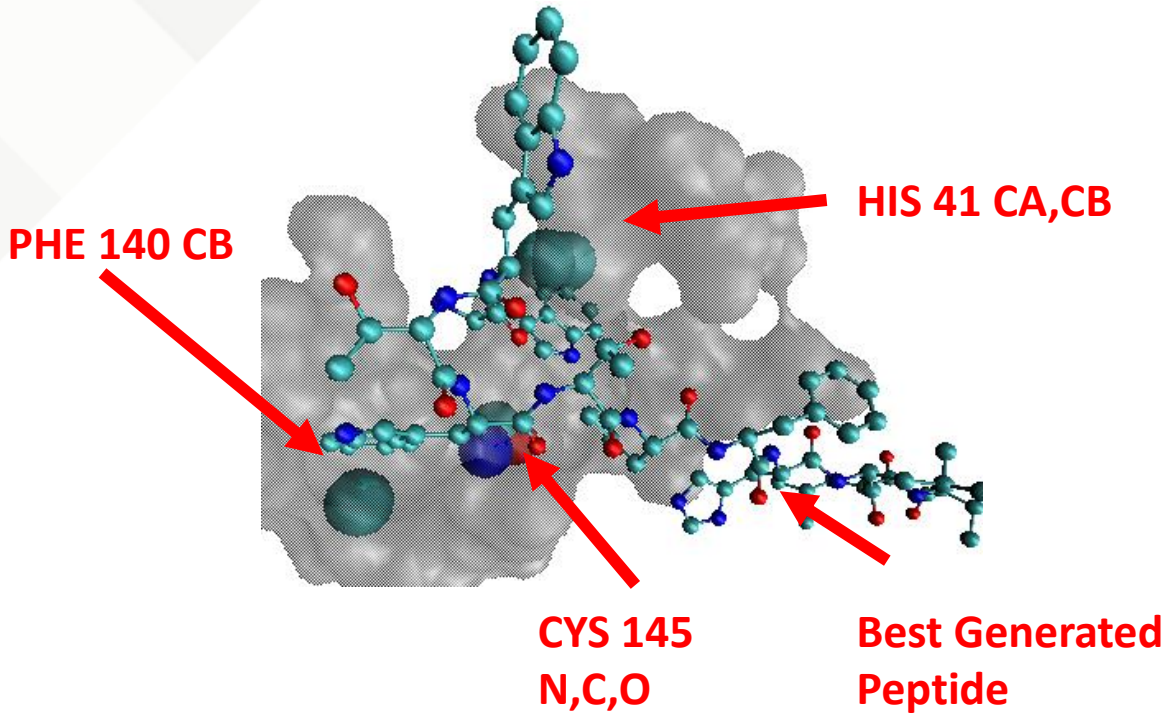


Best Generated Peptide

Top 5 generated peptides for ACE2				
Rank	Docking Binding Affinity	Predicted Affinity	Size	Sequence
1	-29.880	-27.507	7	SSNATV
2	-28.640	-29.962	11	SSNNAVWTASY
3	-27.906	-28.378	10	CSNATVWAAS
4	-25.033	-27.048	16	CSSNNAVWTASFVTWP
5	-22.661	-22.412	5	SSNTV

RESULTS: COVID-19 TARGETS

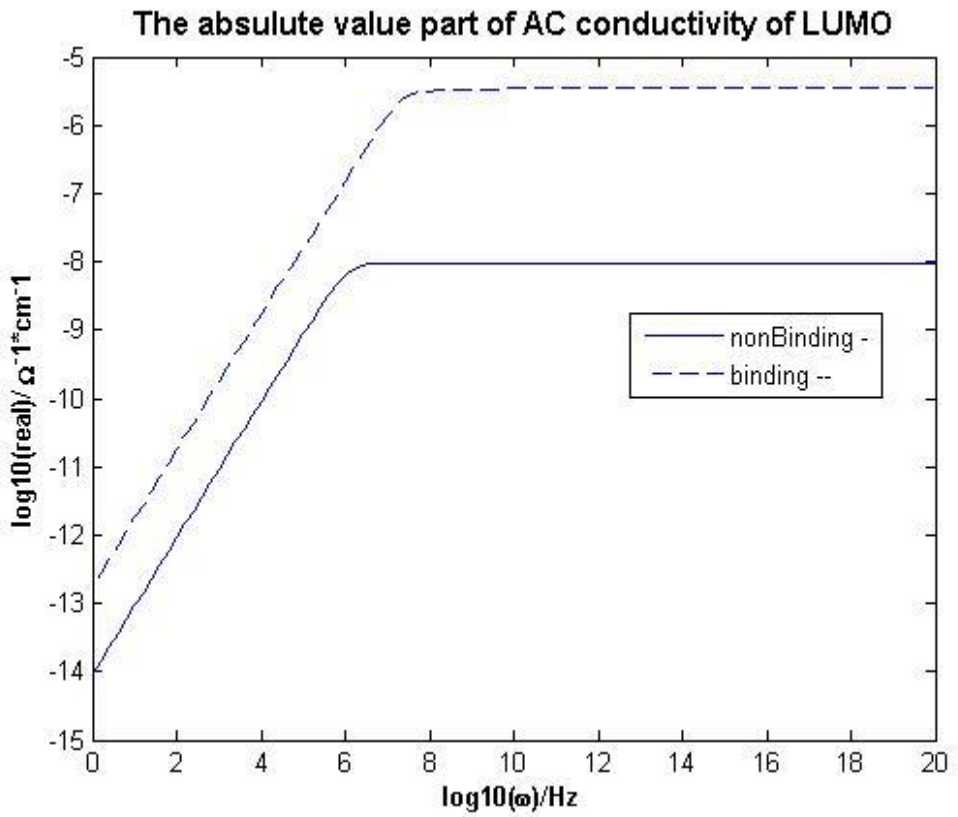
- M^{Pro} top 5 generated peptides
- Small molecule inhibitor has a docking binding affinity of -5.501
  - Based on α- ketoamide



Top 5 generated peptides for M^{Pro}				
Rank	Docking Binding Affinity	Predicted Affinity	Size	Sequences
1	-41.038	-47.738	11	WWTWTPFHLLV
2	-39.513	-40.570	6	WWTFHL
3	-38.098	-36.433	15	VFWTWTPFHLLVRWK
4	-35.883	-35.595	5	WFQIL
5	-21.068	-27.216	11	WTVWRFIHLV

抗原结合抗体后的电导率

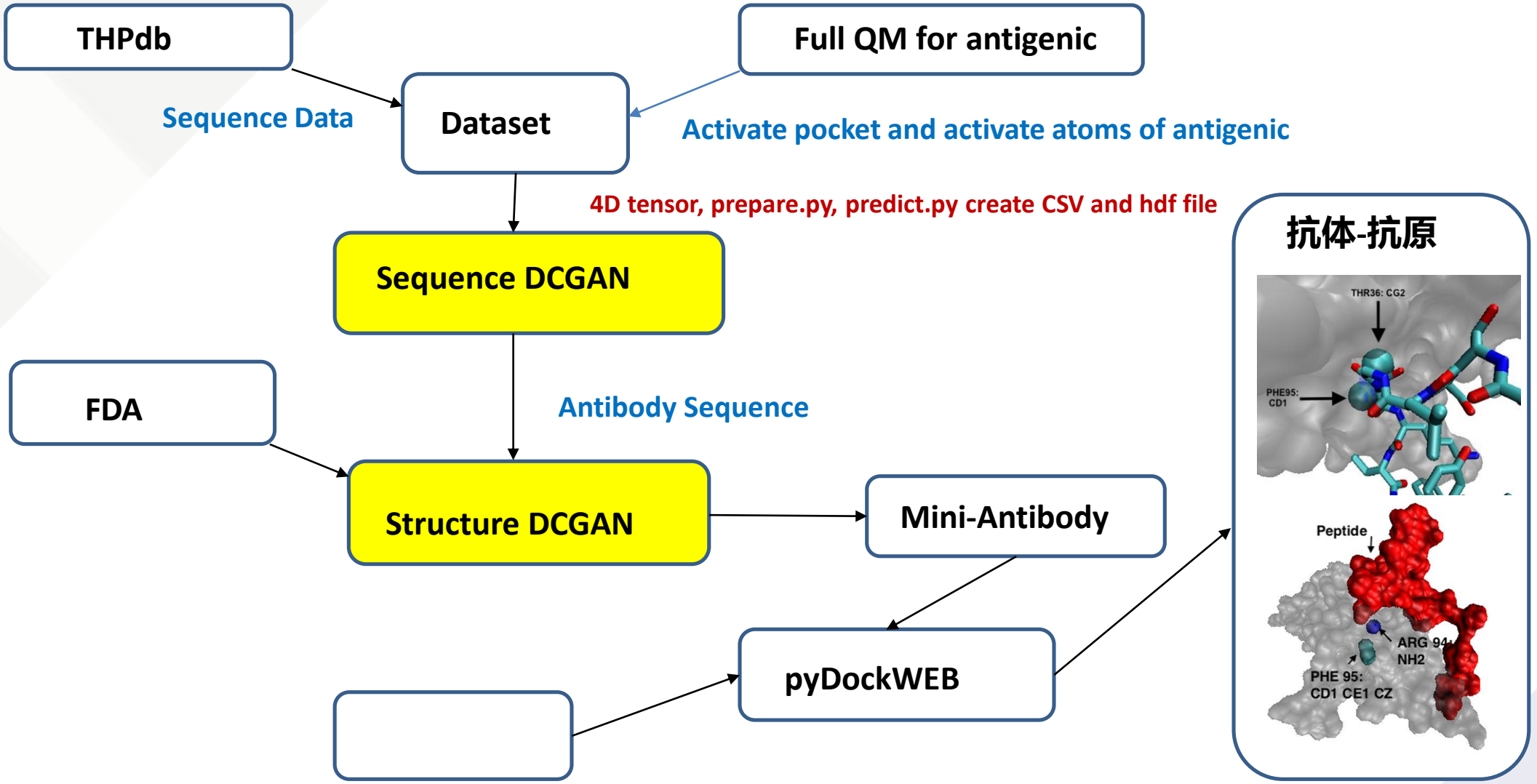
抗体与抗原的相互作用属于蛋白质-蛋白质相互作用，因此我们也计算了抗原结合抗体后的电导率。 以PD1 为例，计算结果是PD1的电导率增加了100倍，绝缘体变成了半导体，从而引入了载流子，导致性能发生变化，这从根本上解释了抗体的作用。



Questions?

REFERENCES

1. DiMasi, Joseph A., Henry G. Grabowski, and Ronald W. Hansen. "Innovation in the pharmaceutical industry: new estimates of R&D costs." Journal of health economics 47 (2016): 20-33.
2. Ferreira, Leonardo G., Ricardo N. Dos Santos, Glaucius Oliva, and Adriano D. Andricopulo. "Molecular docking and structure-based drug design strategies." Molecules 20, no. 7 (2015): 13384-13421.
3. Wallach, Izhar, Michael Dzamba, and Abraham Heifets. "AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery." arXiv preprint arXiv:1510.02855 (2015).
4. Zhang, Lu, Jianjun Tan, Dan Han, and Hao Zhu. "From machine learning to deep learning: progress in machine intelligence for rational drug discovery." Drug discovery today 22, no. 11 (2017): 1680-1685.
5. Van Montfort, Rob LM, and Paul Workman. "Structure-based drug design: aiming for a perfect fit." Essays in biochemistry 61, no. 5 (2017): 431-437.
6. Nero, Tracy L., Michael W. Parker, and Craig J. Morton. "Protein structure and computational drug discovery." Biochemical Society Transactions 46, no. 5 (2018): 1367-1379.
7. Yu, Yang, Zhiqiang Gong, Ping Zhong, and Jiaxin Shan. "Unsupervised representation learning with deep convolutional neural network for remote sensing images." In International Conference on Image and Graphics, pp. 97-108. Springer, Cham, 2017.
8. Popova, Mariya, Olexandr Isayev, and Alexander Tropsha. "Deep reinforcement learning for de novo drug design." Science advances 4, no. 7 (2018): eaap7885.
9. Merk, Daniel, Lukas Friedrich, Francesca Grisoni, and Gisbert Schneider. "De novo design of bioactive small molecules by artificial intelligence." Molecular informatics 37, no. 1-2 (2018): 1700153.
10. Neil, D., Segler, M., Guasch, L., Ahmed, M., Plumbley, D., Sellwood, M. and Brown, N., 2018. Exploring deep recurrent models with reinforcement learning for molecule design.
11. Zhang, Genwei, Sebastian Pomplun, Alexander Robert Loftis, Andrei Loas, and Bradley L. Pentelute. "The first-in-class peptide binder to the SARS-CoV-2 spike protein." bioRxiv (2020).
12. Zhang, Linlin, Daizong Lin, Xinyuanyuan Sun, Ute Curth, Christian Drosten, Lucie Sauerhering, Stephan Becker, Katharina Rox, and Rolf Hilgenfeld. "Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors." Science 368, no. 6489 (2020): 409-412.
13. Shereen, Muhammad Adnan, Suliman Khan, Abeer Kazmi, Nadia Bashir, and Rabeea Siddique. "COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses." Journal of Advanced Research (2020).


THPdb
Full QM for antigenic
Sequence Data
Dataset
Activate pocket and activate atoms of antigenic
4D tensor, prepare.py, predict.py create CSV and hdf file
Sequence DCGAN
FDA
Antibody Sequence
Structure DCGAN
Mini-Antibody
pyDockWEB
抗体-抗原
THR36: CG2
PHE95: CD1
Peptide
ARG 94
NH2
PHE 95:
CD1 CE1 CZ