Transformer-based generative model accelerating the development of novel BRAF inhibitors

Lijuan Yang[1,2,3,4], Guanghui Yang[1,4], Zhitong Bing[1,4], Yuan Tian[5,1], Yuzhen Niu[6*], Liang Huang[2] and Lei Yang[1,4*]

1. Institute of modern physics, Chinese Academy of Science, Lanzhou 730000, China

2. School of Physics and Technology, Lanzhou University, Lanzhou 730000, China

3. School of Physics, University of Chinese Academy of Science, Beijing 100049, China

4. Advanced Energy Science and Technology Guangdong Laboratory, Huizhou 516000, China

5. School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China

6. Shandong Provincial Research Center for Bioinformatic Engineering and Technique, School of Life Sciences, Shandong University of Technology, Zibo 255000, China

* Corresponding to: Yuzhen Niu, Email address: niuyzh12@lzu.edu.cn
* Corresponding to: Lei Yang, Email address: lyang_imp@outlook.com

Table S1: Experimental results of the percentage of small molecules generated when GRUs of different layers are stacked.

	No. of layers =1	No. of layers =2	No. of layers =4
Invalid	4.368%±0.361%	4.104%±0.264%	4.270%±0.245%
Redundant	0.038%±0.015%	0.034%±0.009%	0.038%±0.024%
Retained	95.594%±0.351%	95.862%±0.268%	95.692%±0.241%

Table S2: Test results of the predictive model on the PDBbind dataset.

	Number of samples	RMSE	R
Train	10680	1.034	0.85
Test	1186	1.463	0.68

Table S3: The percentage changes of molecules generated during reinforcement learning.

	Aver reward=10.0		Aver reward=14.0		Aver reward=15.0	
	Transformer	GRU	Transformer	GRU	Transformer	GRU
Invalid	1.080%±0.141%	3.508%±0.160%	1.483%±0.145%	2.772%±0.104%	1.170%±0.052%	2.690%±0.167%
Redundant	14.473%±0.202%	9.584%±0.262%	19.247%±0.720%	20.626%±0.430%	30.273%±0.525%	33.828%±0.616%
Retained	84.447%±0.097%	86.877%±0.502%	79.270%±0.740%	76.427%±0.429%	68.557%±0.552%	63.482%±0.456%

Table S4: Details of the architecture and hyperparameters used for training the generative and predictive models.

	Generative model		Transfer learning		Reinforcement learning		Predictive model
Hyperparameter	Transformer-based model	GRU-based model	Transformer-based model	GRU-based model	Transformer-based model	GRU-based model	
Type of neural network	Transformer	RNN	Transformer	RNN	Transformer	RNN	RNN
No. of layers	12	1	12	1	12	1	3
Hidden size	768	1500	768	1500	768	1500	256
Stacked memory	--	1500	--	1500	--	1500	--
Activation function	GELU	TanH	GELU	TanH	GELU	TanH	ReLU
Optimizer	AdamW	Adam	AdamW	Adam	AdamW	Adam	Adam

| Loss function	Cross Entropy Loss	Cross Entropy Loss	Cross Entropy Loss	Cross Entropy Loss	Cross Entropy Loss	Cross Entropy Loss	MSELoss
Initial learning rate	0.00015	0.0005	0.00015	0.0005	0.00015	0.0005	0.001
Batch size	48	128	48	128	50	50	48
Gradient clipping	Yes	Yes	Yes	Yes	Yes	Yes	Yes
No. of epochs	500	500	100	100	60	150	150
GPU details	1 Tesla A100 GPU	1 Tesla A100 GPU	1 Tesla A100 GPU	1 Tesla A100 GPU	1 Tesla A100 GPU	1 Tesla A100 GPU	1 Tesla A100 GPU
Implementation framework	PyTorch	PyTorch	PyTorch	PyTorch	PyTorch	PyTorch	PyTorch
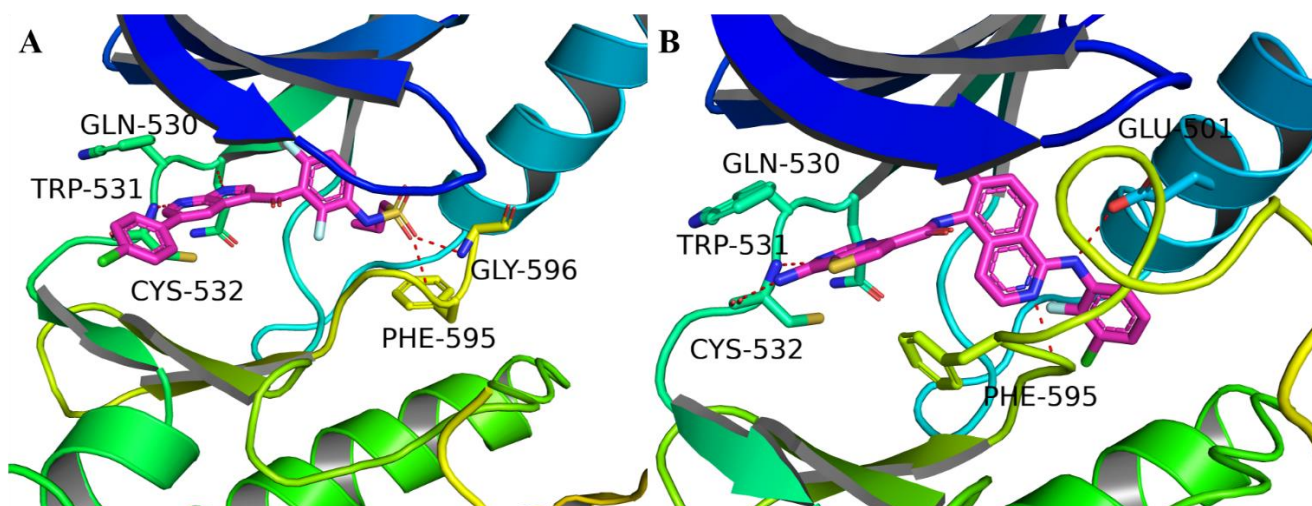
Figure S1. Two binding pockets of BRAF. A. the ATP binding pocket; B. the allosteric binding pocket.