NEURAL INFORMATION PROCESSING SYSTEMS

Predicting mutational effects on protein-protein binding via a side-chain diffusion probabilistic model

Authors: Shiwei Liu*, Tian Zhu*, Milong Ren, Chungong Yu,

Dongbo Bu, Haicang Zhang#

Presenter: Tian Zhu

Affiliation:
Institute of Computing Technology
Chinese Academy of Sciences

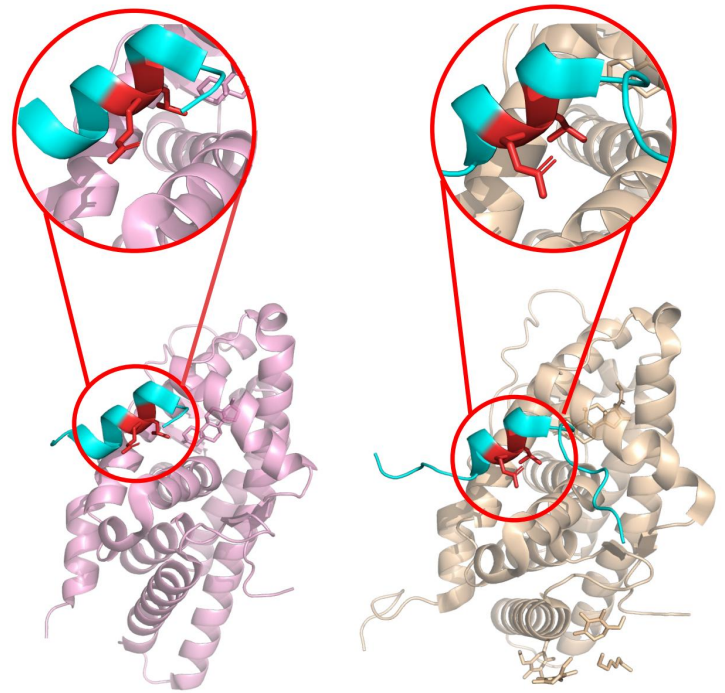
Introduction

NEURAL INFORMATION PROCESSING SYSTEMS

- Many crucial biological processes are regulated by intricate networks of protein-protein interactions.

- In protein engineering, especially in therapeutic discovery, inducing amino acid mutations on the protein-protein interface is essential for modulating binding affinity.

- When different ligands bind to the same receptor, side-chain conformations vary significantly, even though backbone conformations remain consistent.



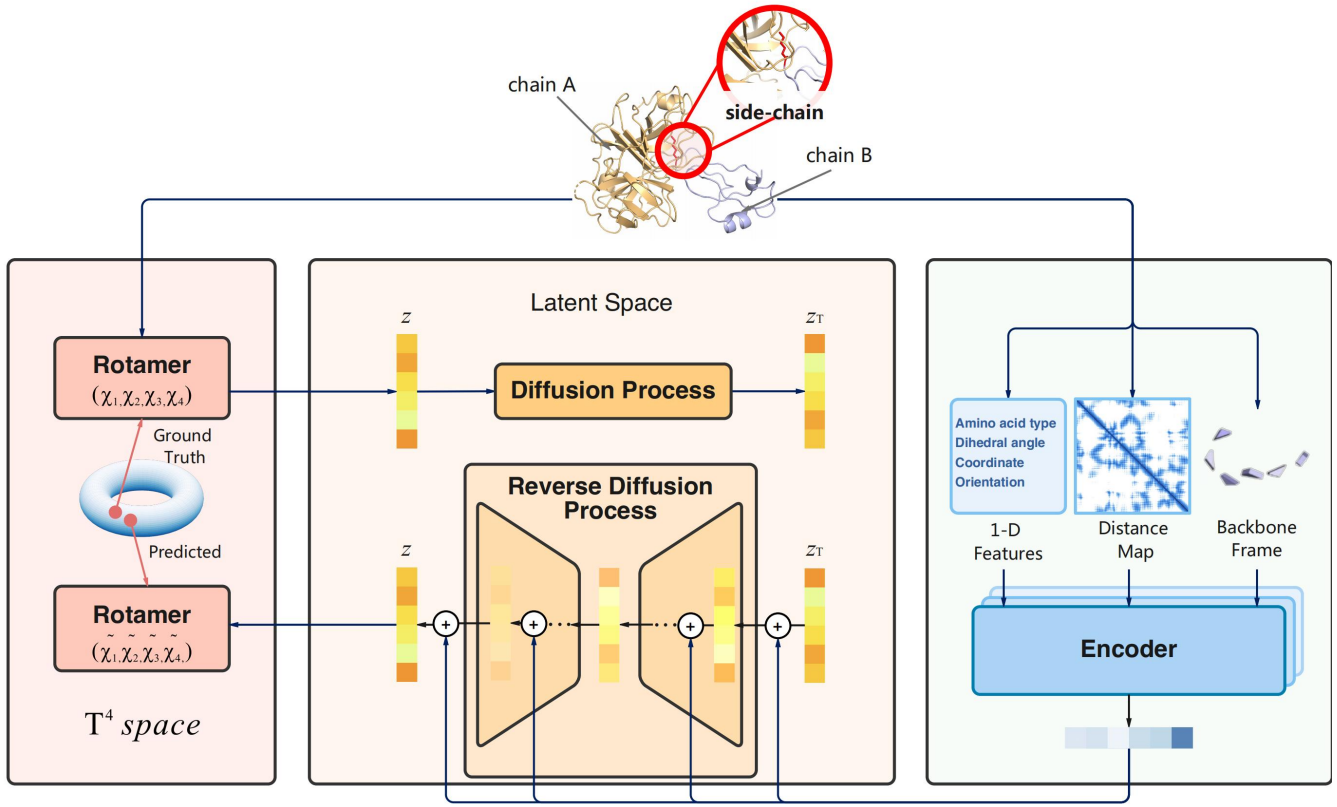
Motivation

NEURAL INFORMATION PROCESSING SYSTEMS

- When docking with various ligands, the side-chain conformations at the protein interface can change.
  - Is it possible to use a generative model to capture the uncertainty in side-chain conformations?
  - Can the generative model enhance our predictions of mutational effects on protein-protein interactions?

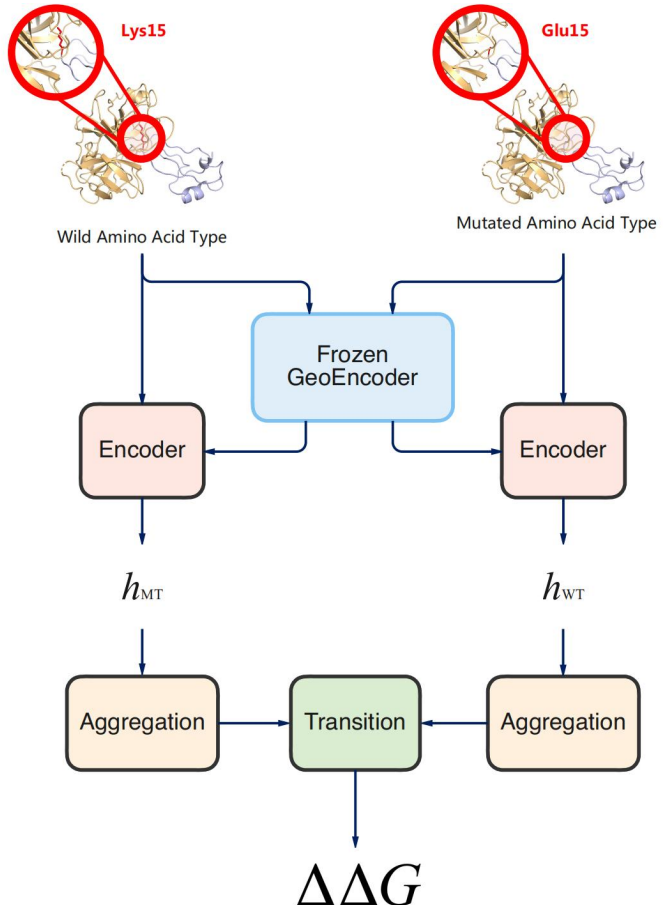
Method - SidechainDiff

- It's the first probabilistic model employing conditional Riemannian diffusion for protein side-chain conformations.
- Additionally, it captures the structural context of mutations at the protein-protein interface.



Method - DiffAffinity

- A model that leverages the representations from SidechainDiff to forecast changes in binding free energy due to mutations.



Results

- DiffAffinity surpasses all baseline models in nearly every metric, achieving state-of-the-art results on the SKEMPI2 dataset.

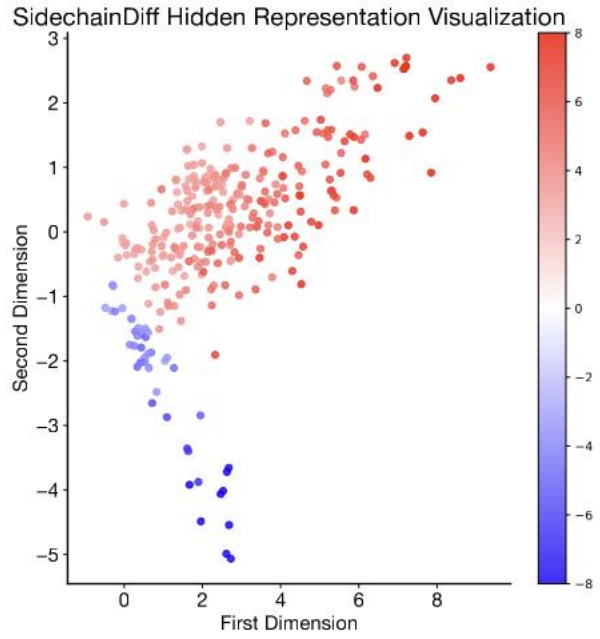
Table 1: Evaluation of prediction on the SKEMPI2 dataset

Method	mutations	Overall						Per-Structure	
		Pearson	**Spearman**	**RMSE**	**MAE**	**AUROC**	**AUPRC**	**Pearson**	**Spearman**
FoldX	all	0.319	0.416	1.959	1.357	0.671	0.839	0.376	0.375
	single	0.315	0.361	1.651	1.146	0.657	0.839	0.382	0.360
	multiple	0.256	0.418	2.608	1.926	0.704	0.841	0.333	0.340
Rosetta	all	0.311	0.346	1.617	1.131	0.656	0.810	0.328	0.298
	single	0.325	0.367	1.183	0.987	0.674	0.834	0.351	0.418
	multiple	0.199	0.230	2.658	2.024	0.621	0.798	0.191	0.083
FelxDDG	all	0.402	0.427	1.587	1.102	0.675	0.866	0.414	0.386
	single	0.425	0.431	1.457	0.997	0.677	0.874	0.433	0.435
	multiple	0.398	0.419	1.765	1.326	0.669	0.854	0.401	0.363
ESM-1v	all	0.192	0.157	1.961	1.368	0.541	0.735	0.007	−0.012
	single	0.191	0.157	1.723	1.192	0.549	0.770	0.042	0.027
	multiple	0.192	0.175	2.759	2.119	0.542	0.678	−0.060	−0.128
ESM-IF	all	0.319	0.281	1.886	1.286	0.590	0.768	0.224	0.202
	single	0.296	0.287	1.673	1.137	0.605	0.776	0.391	0.364
	multiple	0.326	0.335	2.645	1.956	0.637	0.754	0.202	0.149
ESM2	all	0.133	0.138	2.048	1.460	0.547	0.738	0.044	0.039
	single	0.100	0.120	1.730	1.210	0.541	0.734	0.019	0.036
	multiple	0.170	0.163	2.658	2.021	0.566	0.746	0.010	0.010
ESM2*	all	0.623	0.498	1.615	1.179	0.721	0.887	0.362	0.316
	single	0.625	0.468	1.357	0.986	0.707	0.879	0.391	0.342
	multiple	0.603	0.529	2.15	1.67	0.758	0.909	0.333	0.304
DDGPred	all	0.630	0.400	**1.313**	**0.995**	0.696	0.892	0.356	0.321
	single	0.652	0.359	1.309	0.936	0.656	0.884	0.351	0.318
	multiple	0.591	0.503	2.181	1.670	0.759	0.913	0.373	0.385
RDE-Net	all	0.632	0.527	1.601	1.142	0.731	0.887	0.415	0.376
	single	0.637	0.491	1.341	0.961	0.720	0.885	0.413	0.385
	multiple	0.601	0.567	2.157	1.631	0.768	0.898	0.390	0.360
Linear	all	0.326	0.305	1.954	1.399	0.642	0.857	0.222	0.222
	single	0.318	0.293	1.649	1.175	0.651	0.854	0.209	0.202
	multiple	0.277	0.288	2.593	1.961	0.629	0.867	0.193	0.195
DiffAffinity*	all	0.646	0.538	1.578	1.113	0.742	0.741	0.415	0.392
	single	0.657	0.523	1.312	0.931	**0.742**	0.741	0.417	0.396
	multiple	0.613	0.542	2.133	1.606	0.750	0.750	0.407	0.379
DiffAffinity	all	**0.669**	**0.556**	1.535	1.093	**0.744**	**0.896**	**0.422**	**0.397**
	single	**0.672**	**0.523**	**1.288**	**0.923**	0.733	**0.887**	**0.429**	**0.409**
	multiple	**0.650**	**0.602**	**2.051**	**1.540**	**0.784**	**0.921**	**0.414**	**0.387**

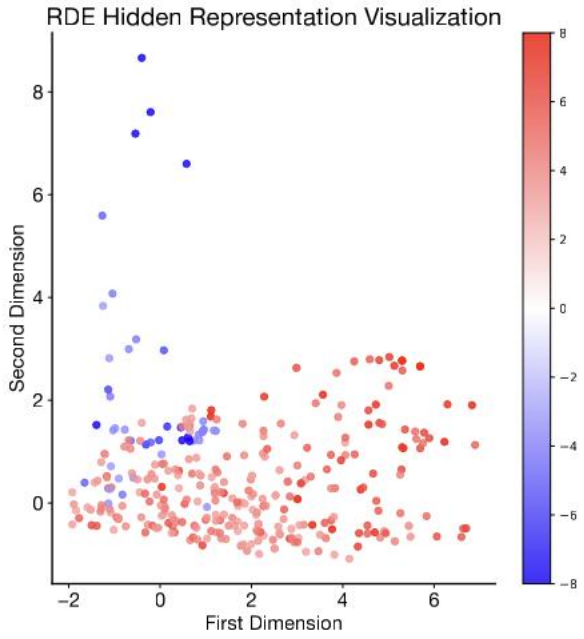
[1] Linear is short of DiffAffinity-Linear just for saving space.

Results

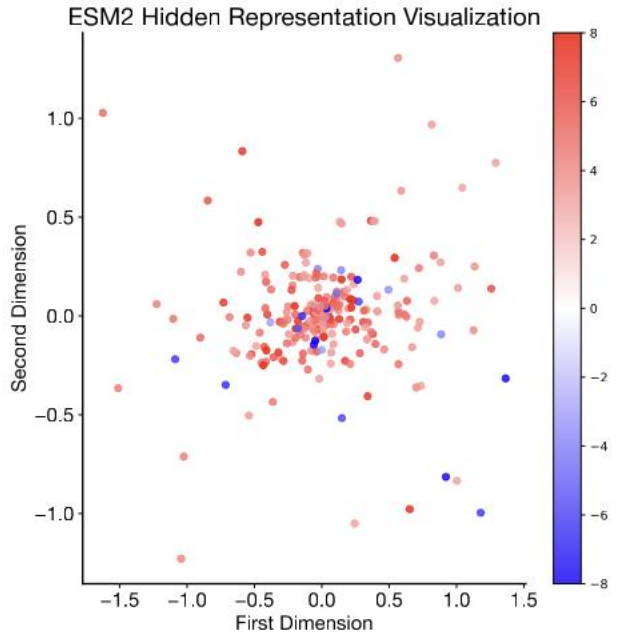
- Representations from SidechainDiff are more adept at distinguishing data based on varying binding free energy values.



(a)



(b)



(c)

Results

- DiffAffinity aids in optimizing antibodies against the SARS-CoV-2 spike protein and in mutational scanning of the spike protein's receptor binding domain (RBD).
  - Predicting the impact of every single-point mutation on the binding affinity of the ancestral Wuhan-Hu-1 RBD (PDB ID:6M0J).
  - Optimizing human antibodies targeting the SARS-CoV-2 spike protein (PDB ID:7FAE).

Table 2: Pearson correlation coefficient in SARS-COV-2 binding affinity

Method	Pearson
FoldX	0.385
RDE-Net	0.438
DiffAffinity*	0.295
DiffAffinity	**0.466**

Table 3: Rankings of the five favorable mutations on the human antibody against SARS-CoV-2 receptor-binding domain (RBD) by various competitive methods

Method	TH31W	AH53F	NH57L	RH103M	LH104F
FoldX	**4.25%**	**14.57%**	**2.43%**	27.13%	63.77%
RDE-Net	**5.06%**	**12.15%**	55.47%	50.61%	**9.51%**
DiffAffinity*	**7.29%**	**0.81%**	**19.03%**	84.21%	28.54%
DiffAffinity	**7.28%**	**3.64%**	**18.82%**	81.78%	**10.93%**

Results

- SidechainDiff outperforms energy-based approaches and matches the performance of RDE and AttnPacker

Table 4: Evaluation of predicted side-chain conformations with more comparisons

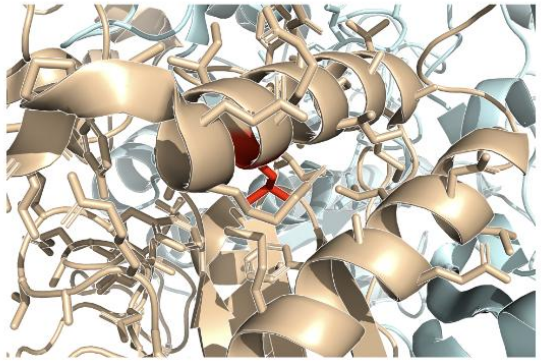
Method	$\chi^{(1)}$	$\chi^{(2)}$	$\chi^{(3)}$	$\chi^{(4)}$	average	clash number
Proportion	100%	82.6%	28.2%	12.6%	-	
SCWRL4	24.33°	32.84°	47.42°	56.15°	28.21°	118.44
Rosetta	23.98°	32.14°	48.49°	58.78°	28.09°	113.31
RDE	**17.13°**	28.47°	**43.54°**	58.62°	**21.99°**	34.41
DLPacker	21.40°	29.27°	50.77°	74.64°	26.42°	62.45
AttnPacker	19.65°	**25.17°**	47.61°	**55.08°**	22.35°	57.31
DiffAffinity	18.00°	27.41°	43.97°	56.65°	22.06°	**27.70**

Table 5: Accuracy of side-chain conformations stratified based on the contact number

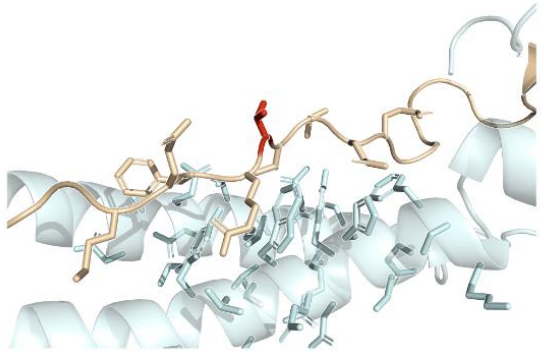
contact number	average χ
$1 \sim 7$	28.70
$8 \sim 10$	22.39
$11 \sim 19$	15.43

Results

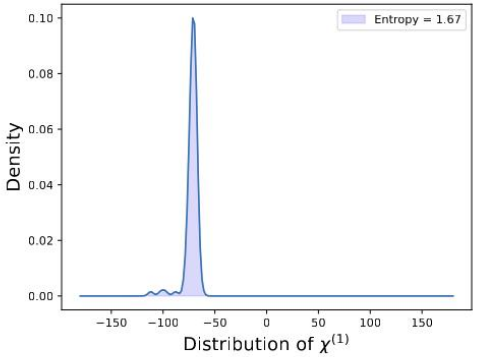
- SidechainDiff accurately forecasts the plausible variations in side-chain conformations.



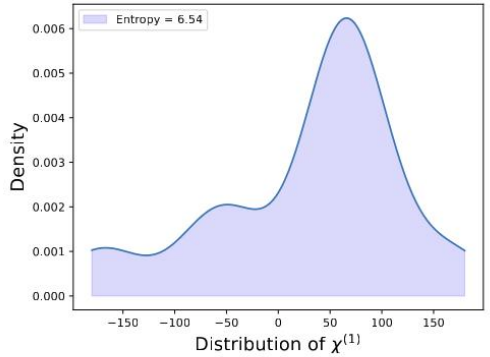
(a)



(b)



(c)



(d)


NEURAL INFORMATION
PROCESSING SYSTEMS


ICT


CHINESE ACADEMY OF SCIENCES

Thanks