Supporting Information

The details algorithms of the triangular self-attention mechanism

For the atom representation for atom i (h_i^{last}) and the pair representation for atom pair ij (q_{ij}^{last}) from the last layer, a series of operations are exerted, including the "Outer product mean" operation to transform atom representation into an update for the pair representation, the "Triangular multiplicative update" operation to update the pair representation by combining information within each triangle of atom pairs ij, ik and jk, the "Triangular self-attention" operation to further update the pair representation, and a transition layer to output the final pair representation.

1. Outer product mean

$$a_i = W_a\left(LN\left(h_i^{last}\right)\right), \quad b_i = W_b\left(LN\left(h_i^{last}\right)\right) \tag{1}$$

$$z_{ij} = q_{ij}^{last} + W_c\left(flatten\left(a_i \otimes b_i\right)\right) \tag{2}$$

where W_a, $W_b \in R^{H \times d}$ and $W_c \in R^{H \times 2H}$ are weight matrices; LN, $flatten$ and $\otimes$ denote the layer normalization, flatten and outer product operations; d and H denotes the hidden dimension and the number of attention head in Transformer encoder, respectively.

2. Triangular multiplicative update

This operation has two symmetric versions, one for the "outgoing" edges and one for the "incoming" edges.

2.1. Triangular multiplicative update using "outgoing" edges

$$a_{ik} = Sigmoid\left(W_{a1}LN\left(z_{ik}\right)\right) \odot W_{a2}LN\left(z_{ik}\right) \tag{3}$$

$$b_{jk} = Sigmoid\left(W_{b1}LN\left(z_{jk}\right)\right) \odot W_{b2}LN\left(z_{jk}\right) \tag{4}$$

$$g_{ij} = Sigmoid\left(W_{g1}LN\left(z_{ij}\right)\right) \tag{5}$$

$$z_{ij} = z_{ij} + g_{ij} \odot W_{g2}LN\left(\sum_k a_{ik} \odot b_{jk}\right) \tag{6}$$

where W_{a1}, W_{a2}, W_{b1}, W_{b2}, W_{g1}, $W_{g2} \in R^{H \times H}$ are weight matrices; LN denotes

layer normalization; $\odot$ denotes inner product; $Sigmoid$ is a nonlinear activation; H denotes the number of attention head in Transformer encoder.

2.2. Triangular multiplicative update using "incoming" edges

$$a_{ki} = Sigmoid\left(W_{a3}LN\left(z_{ki}\right)\right) \odot W_{a4}LN\left(z_{ki}\right) \tag{7}$$

$$b_{kj} = Sigmoid\left(W_{b3}LN\left(z_{kj}\right)\right) \odot W_{b4}LN\left(z_{kj}\right) \tag{8}$$

$$g_{ij} = Sigmoid\left(W_{g3}LN\left(z_{ij}\right)\right) \tag{9}$$

$$z_{ij} = z_{ij} + g_{ij} \odot W_{g4}LN\left(\sum_{k} a_{ki} \odot b_{kj}\right) \tag{10}$$

where W_{a3}, W_{a4}, W_{b3}, W_{b4}, W_{g3}, $W_{g4} \in R^{H \times H}$ are weight matrices; LN denotes layer normalization; $\odot$ denotes inner product; $Sigmoid$ is a nonlinear activation; H denotes the number of attention head in Transformer encoder.

3. Triangular self-attention

This operation also has two symmetric versions, one for the "starting" nodes and one for the "ending" nodes.

3.1. Triangular gated self-attention around starting node

$$Q_{ij}^{h} = W_{Q1}\left(LN\left(z_{ij}\right)\right) \tag{11}$$

$$K_{ik}^{h} = W_{K1}\left(LN\left(z_{ik}\right)\right) \tag{12}$$

$$V_{ik}^{h} = W_{V1}\left(LN\left(z_{ik}\right)\right) \tag{13}$$

$$B_{jk}^{h} = W_{B1}\left(LN\left(z_{jk}\right)\right) \tag{14}$$

$$g_{ij}^{h} = Sigmoid\left(W_{g5}LN\left(z_{ij}\right)\right) \tag{15}$$

$$z_{ij} = z_{ij} + W_{g6}\left(Concat_{h \in 1,\ldots,N_h}\left(g_{ij}^{h} \odot \sum_{k} Softmax_{k}\left(\frac{\left(Q_{ij}^{h}\right)^{T}K_{ik}^{h}}{\sqrt{d_h}} + B_{jk}^{h}\right)V_{ik}^{h}\right)\right) \tag{16}$$

where W_{Q1}, W_{K1}, W_{V1}, W_{B1}, $W_{g5} \in R^{d_h \times H}$, $W_{g6} \in R^{H \times d_h}$ are weight matrices; LN denotes layer normalization; $\odot$ denotes inner product; $Concat$ denotes concatenation operation; $Softmax$ denotes softmax operation; $Sigmoid$ is a nonlinear activation; H denotes the number of attention head in Transformer encoder;

$h \in 1,\ldots,N_h$ denotes the number of attention head here; d_h denotes the dimension of each head here.

3.2 Triangular gated self-attention around ending node

$$Q_{ij}^h = W_{Q2}\left(LN\left(z_{ij}\right)\right) \tag{17}$$

$$K_{ki}^h = W_{K2}\left(LN\left(z_{ki}\right)\right) \tag{18}$$

$$V_{kj}^h = W_{V2}\left(LN\left(z_{kj}\right)\right) \tag{19}$$

$$B_{ki}^h = W_{B2}\left(LN\left(z_{ki}\right)\right) \tag{20}$$

$$g_{ij}^h = Sigmoid\left(W_{g7}LN\left(z_{ij}\right)\right) \tag{21}$$

$$z_{ij} = z_{ij} + W_{g8}\left(Concat_{h \in 1,\ldots,N_h}\left(g_{ij}^h \odot \sum_k Softmax_k\left(\frac{\left(Q_{ij}^h\right)^T K_{ki}^h}{\sqrt{d_h}} + B_{ki}^h\right)V_{kj}^h\right)\right) \tag{22}$$

where W_{Q2}, W_{K2}, W_{V2}, W_{B2}, $W_{g7} \in R^{d_h \times H}$, $W_{g8} \in R^{H \times d_h}$ are weight matrices; LN denotes layer normalization; $\odot$ denotes inner product; $Concat$ denotes concatenation operation; $Softmax$ denotes softmax operation; $Sigmoid$ is a nonlinear activation; H denotes the number of attention head in Transformer encoder; $h \in 1,\ldots,N_h$ denotes the number of attention head here; d_h denotes the dimension of each head here.

4. Transition layer

$$z_{ij} = z_{ij} + W_{T2}\left(RELU\left(W_{T1}\left(LN\left(z_{ij}\right)\right)\right)\right) \tag{23}$$

where $W_{T1} \in R^{2H \times H}$, $W_{T2} \in R^{H \times 2H}$ are weight matrices; LN denotes layer normalization; $RELU$ is a nonlinear activation; H denotes the number of attention head in Transformer encoder.

Table S1. Impacts of two data argumentation strategies on the docking accuracy based on the PDBbind-CrossDocked-Core, APOBind Core and PoseBusters datasets.

Strategy	PDBbind-CrossDocked-Core		APOBind Core		PoseBusters	
	Top1 success rates (%)	Average RMSD (Å)	Top1 success rates (%)	Average RMSD (Å)	RMSD≤ 2.0 Å (%)	RMSD≤ 2.0 Å & PB-Valid (%)
Without data argumentation	**80.91**	**1.543**	**65.94**	**2.094**	**83.4**	**54.4**
CarsiDock	75.09	1.734	50.66	2.778	79.7	47.7

Table S2. The crucial hyperparameter settings for CarsiDock

Hyperparamters	Settings		
	Pre-training	Fine-tuning	Inference [a]
Weight of distance loss for protein-ligand atom pairs (w_{cross_dist})	1.0	Teacher: 1.0; Student: 1.0	1.0
Weight of distance loss for intramolecular pairs in ligand (w_{lig_dist})	1.0	Teacher: 0.1; Student: 1.0	1.0
Weight for MDN loss (w_{MDN})	1.0	Teacher: 0.1; Student: 0.1	-
Weight for distillation loss ($w_{distillation}$)	-	0.1	-
Dimension of hidden representations (d)	768	768	768
Number of attention heads (H)	16	16	16
Number of layers for protein encoder	6	6	6
Number of layers for ligand encoder	6	6	6
Number of layers for interactive encoder	6	6	6
Number of recycles for interactive encoder	3	3	3
Threshold for the calculation of protein-ligand distance [b]	Training: 8; Prediction:6	Training: 8; Prediction:6	Prediction:6
Learning rate	1e-4	5e-5	1e-3
Batch size	96	16	
Epoch	10	50	
Weight decay	1e-4	1e-6	0
Ratio of warmup phase to the total phase	0.05	0.05	No warmup
Initial learning rate for warmup phase	1e-8	5e-9	No warmup
Optimizer	AdamW	AdamW	LBFGS

[a]: These hyperparameters are employed in the geometry optimization stage where the distance matrices are reconstructed to a binding pose.

[b]: Only the protein-ligand atom pairs within the threshold are considered for loss calculation. The threshold is set to 8.0 for model training while the value turns to 6.0 for prediction.

Table S3. Runtimes of two different versions of CarsiDock on PDBbind-v2016 core set.

Version	Inference time (s)	Conversion time (s)	Total time (s) [a]
CPU	1.96 ± 0.32	158.39 ± 259.02	160.35 ± 259.17
GPU	1.27 ± 0.30	4.68 ± 2.21	5.95 ± 2.43

[a]: The experiment is tested on a single-core single-card NVIDIA Geforce RTX 3090 machine.

Table S4. The impact of the number of initial conformers yielded by RDKit (OpenBabel) on the docking accuracy of CarsiDock on PDBbind-v2016 core set.

Number of initial conformers	Top1 success rates (%)	Average RMSD (Å)
1	68.07	2.513
2	87.37	1.243
3	89.82	1.206
4	89.82	1.191
5	**90.18**	1.182
10	89.82	**1.165**
10 (OpenBabel)	86.92	1.284

Table S5. The impact of initial conformers on the docking accuracy of different docking approaches on PDBbind-v2016 core set.

Methods	Using crystal pose coordinates as initial ligand coordinates		Using 10 conformers yielded with the ETKDG algorithm	
	Top1 success rates (%)	Average RMSD (Å)[a]	Top1 success rates (%)	Average RMSD (Å)
Glide SP	66.67	2.200	64.91	2.206
Glide XP	68.07	2.112	65.61	2.218
AutoDock4	55.79	2.966	46.74	3.449
AutoDock Vina	64.21	2.332	52.28	3.091
Vinardo	61.75	2.743	48.07	3.643
AutoDock-GPU	49.46	3.798	39.86	4.189
Vina-GPU	60.00	2.646	51.23	2.989
Gnina	75.09	1.486	72.63	1.875
DeepDock	36.14	3.892	44.91	3.550
TankBind	70.18	1.866	68.42	1.860
EDM-Dock[b]	45.26	2.686	46.32	2.631
CarsiDock	**94.74**	**0.675**	**89.82**	**1.165**

[a]: the complexes failing in docking are directly omitted to calculate the average RMSD.

[b]. the pose with the lowest RMSD value across the 10 runs are simply employed as the final pose when fed with 10 initial conformers.

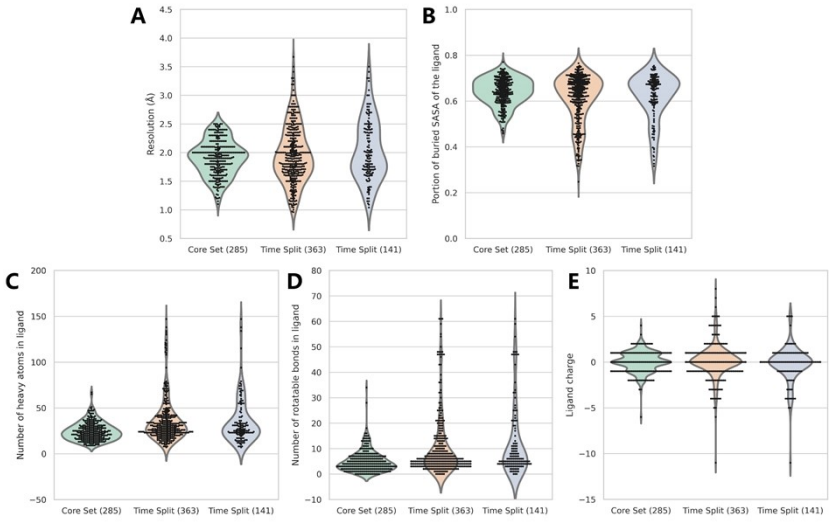

Figure S1. Multiple properties including (A) the X-ray resolution of the complex structure, (B) portion of buried SASA of the ligand, (C) number of heavy atoms in ligand, (D) number of rotatable bonds in ligand, and (E) ligand net charge of the PDBbind-v2016 core set, time-split set of the PDBbind-v2020 dataset, and new receptors on the time-split set.

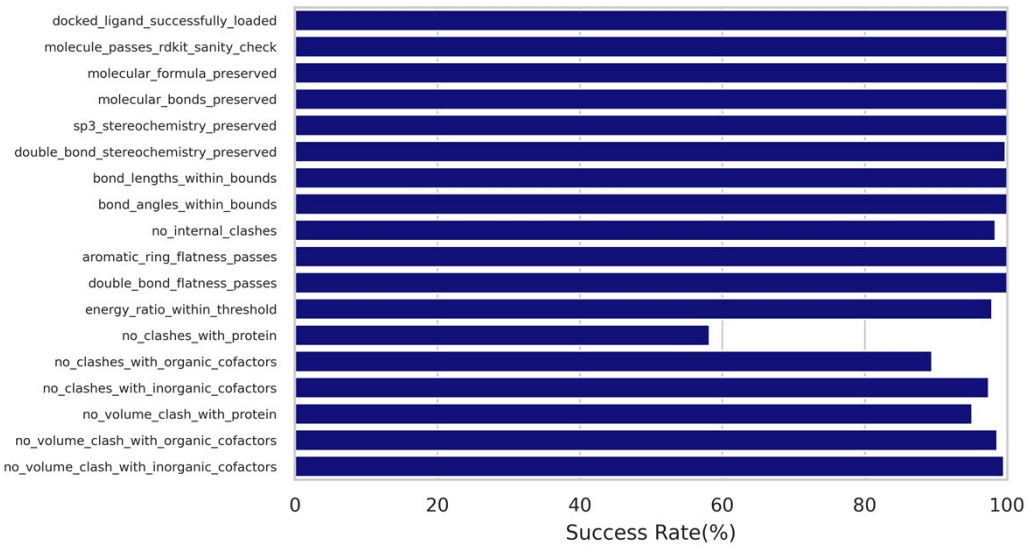

Figure S2. Success rates of CarsiDock passing the different checks in PoseBusters benchmark set.

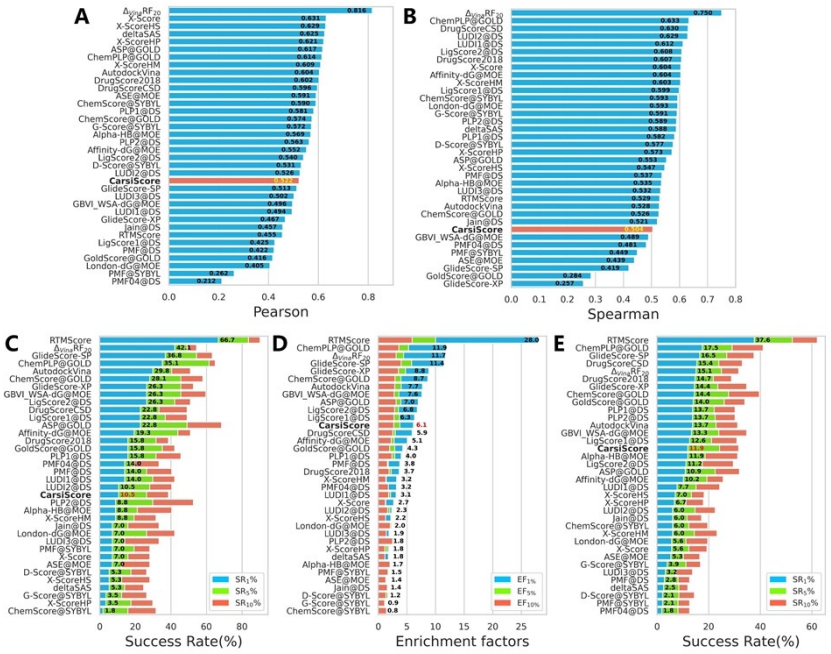

Figure S3. Performances of scoring functions on the CASF-2016 benchmark, including (A) the scoring power in terms of the Pearson correlation coefficient, (B) the ranking power in terms of the Spearman correlation coefficient, and the screening powers in terms of (C) success rates and (D) enrichment factors in the forward screening and (E) success rates in the reverse screening.

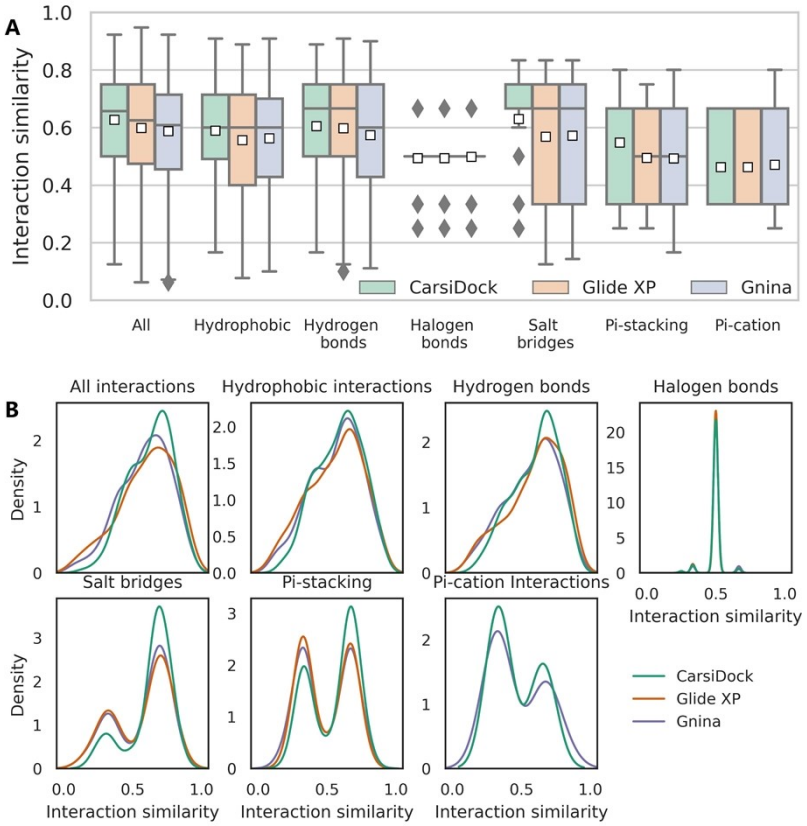

Figure S4. (A) Interaction similarity at the residue level and (B) the corresponding distributions for seven types of interactions of the poses predicted by three docking programs, including all interactions, hydrophobic interactions, hydrogen bonds, halogen bonds, salt bridges, pi-stacking, and pi-cation interactions. The white square in the box plot represents the mean value of each statistics.