Supporting Information

Graph Convolutional Neural Networks for Predicting Drug-Target Interactions

Wen Torng[1] & Russ B. Altman[1,2]

[1]Deparment of Bioengineering, Stanford University, Stanford, CA 94305, [2] Department of Genetics, Stanford University, Stanford, CA, 94305

Wen Torng:

Shriram Room 213

443 Via Ortega Drive

Stanford, CA 94305-4145

Email: wtorng@stanford.edu

Corresponding Author:

Russ B. Altman MD, PhD

Shriram Room 209, MC: 4245

443 Via Ortega Drive

Stanford, CA 94305-4145

Email: russ.altman@stanford.edu

Phone: 650-725-3394

Contact: rbaltman@stanford.edu

Key words: Graph convolutional neural networks, Protein-ligand binding, Pocket representations, Structural Bioinformatics, Drug discovery

Note S1. Stanford XStream and Sherlock Servers

The Stanford XStream GPU cluster is made of 65 compute nodes for a total of 520 Nvidia K80 GPU cards. The Stanford Sherlock cluster includes 6 GPU nodes with dual socket Intel(R) Xeon(R) CPU E5-2640 v2 @ 2.00GHz; 256 GB RAM; 200 GB local storage.

Note S2. Error analysis of MUV - Case Analysis plot

(1) Extent of separation of the actives from the negatives in simple molecular descriptor space

For a given MUV target dataset, to quantify the extent of similarity among the actives and the separation between the actives and negatives, we conceptually view the actives and the negatives as two different clusters and calculate the average silhouette scores of the actives. Specifically, for the active i^{th} active, we calculate the s(i) by the following equations:

- a(i) = average distance between the i^{th} active and all other actives
- b(i) = average distance between the i^{th} active and all negatives

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

The final score for the MUV target is then calculated by

S = average(s(i)) for all actives

Distances between the molecules are calculated by Euclidean distances of the molecules, represented using the Morgan (ECFP) fingerprints.

(2) Average binding site similarity of the MUV target to the DUD-E targets

We quantify pairwise pocket similarities between each MUV target and the DUD-E targets using the PocketFEATURE program. For each MUV target, we then average its pocket similarity score (negative PocketFEATURE score) for all the DUD-E targets to obtain a single final score.

Note S3. Alternative MUV structures to examine model sensitivity to pocket choice

To examine the model sensitivity to pocket conformational changes, we selected alternative structures of the MUV targets --- unbound structures, or bound structures co-crystalized with ligands that are dis-similar to test ligands as below.

(1) For each MUV target, we retrieved PDB structures that map to the same target by Uniprot ID.

(2) For each retrieved structure, if the structure does not have a bound ligand at the known pocket site, we include it as an apo structure. Otherwise, we compute pairwise Dice similarity of the bound ligand to the actives of the corresponding MUV target, and select structures that have maximum bound-ligand Dice similarity lower than a 0.5 threshold to the test actives. The ligands are represented by Morgan Fingerprints with radius of 2 and the Dice similarity was computed using the rdkit.DataStructs.DiceSimilarity module.

Dice similarity between the bound ligands of the alternative structures and the corresponding MUV actives are summarized in Table S5. For Human cAMP-dependent protein kinase (MUV target 548), we included an apo structure, 6BYS, to evaluate our model. We were not able to find alternative structure for MUV target 466 based on the above criteria.

Note S4. Derivation of pocket and ligand importance scores

We calculate the importance scores by following three stages:

- **Saliency of classification score to hidden nodes in the Interaction Layer**

  We first calculate the derivative of the true class score of the pocket-molecule pair with respect to the interaction nodes H_{inter} at H_{inter0}, where H_{inter0} denotes the interaction node values of the given pocket and molecule pair. The derivative is then multiplied by H_{inter0} to obtain the saliency score for each interaction node. By first order Taylor approximation, the saliency score of each node approximates the effect on the true class score when removing the corresponding interaction node.

$$grad_{H_{inter}} = \frac{\partial score_{class}}{\partial H_{inter}} \mid_{H_{inter0}} \quad \text{Eq (1)}$$

$$sali_{H_{inter}} = grad_{H_{inter}} * H_{inter0} \quad \text{Eq (2)}$$

  Interaction nodes that have positive saliency scores ($sali_{H_{inter}}$) are then identified and sorted according to their saliency scores, where a higher saliency score indicate larger contribution. For each pocket-ligand pair, we visualize the top 5 (out of 100)

interaction nodes that have the highest saliency scores.

- **Saliency of key interaction node to pocket and molecule fingerprint attributes**

For each identified key interaction node (with index h_{idx}), we derive the contribution of each molecule fingerprint and pocket fingerprint attribute to the interaction node value by similarly calculate the saliency score of each fingerprint attribute to the interaction node value.

$$sali_{h_{idx}} = sali_{H_{inter}}[h_{idx}] \quad \text{Eq (3)}$$

$$grad_{FP_{poc}} = \frac{\partial H_{inter}[h_{idx}]}{\partial FP_{poc}} \mid_{FP_{poc0}} \quad \text{Eq (4)}$$

$$sali_{FP_{poc}} = grad_{FP_{poc}} * FP_{poc0} \quad \text{Eq (5)}$$

$$grad_{FP_{mol}} = \frac{\partial H_{inter}[h_{idx}]}{\partial FP_{mol}} \mid_{FP_{mol0}} \quad \text{Eq (6)}$$

$$sali_{FP_{mol}} = grad_{FP_{mol}} * FP_{mol0} \quad \text{Eq (7)}$$

Where FP_{poc0} denotes the pocket fingerprint values for the given input pocket. FP_{mol0} denotes the molecular fingerprint values for the given input ligand.

Pocket and molecular fingerprint attributes that have positive saliency scores to the key interaction node are then identified, with their saliency scores recorded.

- **Saliency of key pocket / molecule fingerprint to pocket residues and ligand atoms**

For each identified key pocket and molecular fingerprint attribute, we identify the key contributing pocket residues and atoms. Below we describe the procedure to derive contribution of each pocket residue to the pocket fingerprint attribute indexed by fp_{idx}. Let x_v be the node feature of residue node v, and X_{poc} be a matrix containing the node features of all nodes within a pocket graph arranged as columns. Different from the previous procedures, we cannot directly take derivatives of FP_{poc} with respect to X_{poc} to compute the saliency scores. This is because as the Softmax function in Equation (9) reaches saturation, the gradients vanish to 0, prohibiting the gradient to flow freely from FP_{poc} to the input features X_{poc}.

$$S_v = W_{FP_{poc}} x_v + b_{FP_{poc}} \quad \text{Eq (8)}$$

$$FP_{poc} = \sum_{v \in pocket} FP_v = \sum_{v \in pocket} Softmax(S_v) \quad \text{Eq (9)}$$

Instead, for each residue in the pocket, we compute

$$sali_{FP_{(fp_{idx},v)}} = grad_{FP_{poc}}[fp_{idx}] * FP_{\mathrm{v}}[fp_{idx}]_0 \quad \text{Eq(10)}$$

$$grad_{X_{poc}} = \frac{\partial S_v[fp_{idx}]}{\partial X_{poc}} \mid_{\mathrm{X_{poc0}}} \quad \text{Eq(11)}$$

$$sali_{X_{poc}} = grad_{X_{poc}} * X_{poc0} \quad \text{Eq(12)}$$

$$Importance[h_{idx}, fp_{idx}, v] = sali_{h_{idx}} * sali_{FP_{(fp_{idx},v)}} * sali_{X_{poc}} \quad \text{Eq(13)}$$

Where $FP_{\mathrm{v}}[fp_{idx}]_0$ denotes the value of the $fp_{idx}{}^{\mathrm{th}}$ attribute of residue fingerprint of node v in the given pocket graph. X_{poc0} denotes the values of the node features of the given pocket graph.

Contribution of each pocket residue to the $h_{idx}{}^{\mathrm{th}}$ interaction node are then computed and integrated. Specifically,

$$Importance[h_{idx}] = \sum_{fp_{idx} \in \mathrm{pos}(h_{idx})} \sum_{\mathrm{v} \in pocket} Importance[h_{idx}, fp_{idx}, \mathrm{v}] \quad \text{Eq (14)}$$

Where $\mathrm{pos}(h_{idx})$ denotes all the pocket fingerprint indexes which have positive saliency scores for interaction node h_{idx}.

The resulting importance scores of residues in the pocket are then normalized by the maximum score so that all scores have values between 0 to 1. Importance scores of atoms in the ligand are calculated similarly.

Table S1. Dice similarity between bound ligands of DUD-E targets and DUD-E actives

Target	Average	Standard Deviation	Max
AA2AR	0.410	0.131	1.000
ABL1	0.384	0.079	0.859
ACE	0.408	0.149	0.882
ACES	0.266	0.075	0.889
ADA	0.406	0.168	1.000
ADA17	0.349	0.128	0.821
ADRB1	0.309	0.091	0.729
ADRB2	0.244	0.079	1.000
AKT1	0.212	0.080	0.929
AKT2	0.273	0.170	1.000
ALDR	0.435	0.162	0.933
AMPC	0.469	0.232	1.000
ANDR	0.250	0.177	0.898
AOFB	0.270	0.087	0.718
BACE1	0.265	0.054	0.600
BRAF	0.410	0.176	1.000
CAH2	0.168	0.044	0.354
CASP3	0.320	0.107	0.651
CDK2	0.310	0.067	0.719
COMT	0.519	0.148	1.000
CP2C9	0.349	0.056	0.523
CP3A4	0.260	0.078	0.664
CSF1R	0.339	0.147	0.845
CXCR4	0.401	0.315	1.000
DEF	0.325	0.136	1.000
DHI1	0.311	0.110	0.882
DPP4	0.289	0.120	0.894
DRD3	0.268	0.059	0.794
DYR	0.395	0.096	0.884
EGFR	0.430	0.104	0.856
ESR1	0.519	0.191	0.948
ESR2	0.387	0.115	1.000
FA10	0.335	0.098	1.000
FA7	0.479	0.118	1.000
FABP4	0.499	0.242	1.000
FAK1	0.467	0.162	1.000
FGFR1	0.297	0.050	0.420
FKB1A	0.388	0.104	0.694
FNTA	0.379	0.075	0.712
FPPS	0.478	0.118	1.000
GCR	0.324	0.066	0.526
GLCM	0.137	0.081	0.261

GRIA2	0.245	0.112	1.000
GRIK1	0.272	0.111	0.758
HDAC2	0.398	0.151	0.859
HDAC8	0.432	0.119	1.000
HIVINT	0.310	0.044	0.422
HIVPR	0.390	0.090	0.688
HIVRT	0.243	0.074	0.637
HMDH	0.508	0.185	0.901
HS90A	0.315	0.074	0.529
HXK4	0.366	0.126	1.000
IGF1R	0.371	0.099	0.909
INHA	0.353	0.235	1.000
ITAL	0.377	0.181	0.835
JAK2	0.381	0.140	1.000
KIF11	0.427	0.116	0.741
KIT	0.267	0.079	1.000
KITH	0.404	0.207	1.000
KPCB	0.444	0.149	0.695
LCK	0.360	0.105	0.804
LKHA4	0.467	0.094	0.763
MAPK2	0.349	0.054	0.497
MCR	0.192	0.153	1.000
MET	0.458	0.127	0.854
MK01	0.407	0.108	0.610
MK10	0.344	0.082	0.736
MK14	0.315	0.067	1.000
MMP13	0.449	0.104	1.000
MP2K1	0.299	0.155	1.000
NOS1	0.091	0.087	0.625
NRAM	0.261	0.087	1.000
PA2GA	0.373	0.132	1.000
PARP1	0.375	0.071	0.736
PDE5A	0.343	0.140	0.913
PGH1	0.303	0.094	0.855
PGH2	0.419	0.151	0.915
PLK1	0.313	0.077	0.839
PNPH	0.305	0.113	0.699
PPARA	0.423	0.101	0.940
PPARD	0.425	0.081	0.783
PPARG	0.436	0.084	0.783
PRGR	0.335	0.083	0.768
PTN1	0.289	0.114	0.595
PUR2	0.534	0.063	0.660
PYGM	0.325	0.047	0.421
PYRD	0.547	0.214	0.922

RENI	0.298	0.095	0.613
ROCK1	0.261	0.066	0.416
RXRA	0.111	0.034	0.204
SAHH	0.668	0.118	1.000
SRC	0.330	0.086	0.591
TGFR1	0.493	0.136	1.000
THB	0.412	0.124	0.764
THRB	0.308	0.062	0.534
TRY1	0.316	0.076	0.488
TRYB1	0.331	0.071	0.512
TYSY	0.409	0.113	0.894
UROK	0.330	0.082	0.904
VGFR2	0.362	0.119	1.000
WEE1	0.620	0.113	0.859
XIAP	0.446	0.091	0.683

We computed pairwise Dice similarity between the bound-ligand of each DUD-E target to the corresponding DUD-E actives using the rdkit.DataStructs.DiceSimilarity module. The ligands are represented by Morgan Fingerprints with radius of 2.

Table S2. Dice similarity between bound ligands in MUV targets and MUV actives

Target	Average	Standard Deviation	Max
846	0.329	0.060	0.447
600	0.068	0.045	0.155
692	0.064	0.032	0.147
859	0.233	0.055	0.478
852	0.087	0.020	0.122
548	0.257	0.055	0.343
713	0.245	0.065	0.365
466	0.215	0.046	0.321
689	0.275	0.045	0.351
832	0.108	0.038	0.185

We computed pairwise Dice similarity score of the bound-ligand of each MUV target to the corresponding MUV actives using the rdkit.DataStructs.DiceSimilarity module. The ligands are represented by Morgan Fingerprints with radius of 2.

Table S3. List of functional atoms used to determine the functional centers representing each residue type

Residue Type	Site 1	Site 2
Glycine (G)	CA	-
Cysteine (C)	SG	-
Arginine (R)	CZ	-
Serine (S)	OG	-
Threonine (T)	OG1	-
Lysine (K)	NZ	-
Methionine(M)	SD	-
Alanine (A)	CB	-
Leucine (L)	CB	-
Isoleucine (I)	CB	-
Valine (V)	CB	-
Aspartic acid (D)	OD1, CG, OD2	-
glutamic acid (E)	OE1, CD, OE2	-
histidine (H)	NE2, ND1	-
Asparagine (N)	OD1, CG, ND2	-
Proline (P)	N, CA, CB, CD, CG	-
Glutamine (Q)	OE1, CD, NE2	-
Phenylalanine (F)	CG, CD1, CD2, CE1, CE2, CZ	-
Tryptophan (W)	CD2, CE2, CE3, CZ2, CZ3, CH2	NE1
Tyrosine (Y)	CG, CD1, CD2, CE1, CE2, CZ	OH

For each residue type, the average location of the listed functional atoms for a given site is used to represent the corresponding amino acid environment. Tyrosine and Tryptophan are each represented by two sites due to their larger sizes.

Table S4. Network architecture and parameters of unsupervised pocket graph autoencoders.

Pocket Graph-Autoencoder	Input	Parameters	Output
Layer1 Autoencoder I	Pocket Graph G_{pocket} With residue embedding from FEATURE program ∈ [N_{res},480]	$W_{self_{L1}}$ ∈ R [480,200] , $W_{deg_{y_{L1}}}$ ∈ R [480,200] , max degree = 20	Residue embedding ∈ R [N_{res},200]
Layer1 Autoencoder II		$W_{FP_{L1}}$ ∈ R [480,512]	$FP_{res_{L1}}$ ∈ R [N_{res},512]
Layer2 Autoencoder I	Pocket Graph G_{pocket} With residue embedding from Layer 1 ∈ [N_{res},200]	$W_{self_{L2}}$ ∈ R [200,100] , $W_{deg_{y_{L2}}}$ ∈ R [200,100] , max degree = 20	Residue embedding ∈ R [N_{res},100]
Layer2 Autoencoder II		$W_{FP_{L2}}$ ∈ R [200,512]	$FP_{res_{L2}}$ ∈ R [N_{res},512]
Average Layer	$FP_{poc_{L1}} = \frac{1}{N_{res}} \sum_{res} FP_{res_{L1}}$, ∈ [1,512] $FP_{poc_{L2}} = \frac{1}{N_{res}} \sum_{res} FP_{res_{L2}}$, ∈ [1,512]		
Output Layer	$FP_{poc} = FP_{poc_{L1}} + FP_{poc_{L2}}$, ∈ [1,512]		

Our pocket graph-autoencoder comprises two graph-autoencoder layers, each including two autocoders. In each layer, Autoencoder I takes in the pocket graph, and compresses local graph neighborhood information in the previous layer into new residue embeddings using convolutional filters. Autoencoder II takes in embeddings of residue nodes within the same layer, and integrates them into a fixed-size pocket fingerprint. The residue embeddings in Layer 1 are the FEATURE vectors which describe the amino acid environment for each key residue, whereas the residue embeddings in Layer 2 are the output from Autoencoder I in Layer 1. The input and output columns describe the input and output of each module respectively. The parameter column describes the learnable parameters in each module. The bias terms are omitted here for simplicity.

Table S5. Dice similarity between bound ligands in alternative MUV structures and MUV actives

MUV Target	Alternative PDB of the MUV target	Bound ligand	Dice similarity to MUV actives		
			Average	Standard Deviation	Max
832	1KYN	KTP	0.303	0.079	0.444
859	6OIJ	IXO	0.111	0.029	0.165
692	4QJR	PIZ	0.054	0.027	0.130
600	4QJR	PIZ	0.058	0.037	0.130
689	2XYU	Q9G	0.228	0.056	0.422
548	6BYR	ATP	0.168	0.039	0.262
548	5BX7	4W1	0.266	0.068	0.392
548	5BX6	495	0.287	0.059	0.396
713	5U2D	OBH	0.248	0.061	0.343
713	5T92	77W	0.279	0.068	0.414

For each MUV alternative structure, we compute pairwise Dice similarity between its bound ligand and the actives of the corresponding MUV target and summarized the average, maximum and standard deviation of the similarity scores.

Table S6. Reconstruction errors of the pocket graph-autoencoder layers

Pocket graph-autoencoder		Input Attribute Size	Reconstruction Error[1]	Percentage of Error[2]
Layer 1	Autoencoder I	480	1.623	0.169%
	Autoencoder II	480	4.345	0.905%
Layer 2	Autoencoder I	200	2.817	0.705%
	Autoencoder II	200	6.020	3.010%

1: The reconstruction error for Autoencoder I is defined as

$Error_{reconstruct} = \frac{1}{N}\sum_p \frac{1}{R_p}\sum_r \sum_i [\left(v'_{x(p,r,i)} - v_{x(p,r,i)}\right)^2 + \left(v'_{n(p,r,i)} - v_{n(p,r,i)}\right)^2]$, where $v_{x(p,r,i)}$ denotes the true value of attribute i of the node embedding of residue r in pocket p, $v'_{x(p,r,i)}$ denotes the reconstructed value of attribute i of the node embedding of residue r in pocket p by Autoencoder I. $v_{n(p,r,i)}$ denotes the value of attribute i of the neighborhood vector of residue r in pocket p, $v'_{n(p,r,i)}$ denotes the reconstructed value of the neighborhood vector of residue r in pocket p by Autoencoder I. N is the total number of pockets, and R_p is the number of residues in pocket p.

The reconstruction error for Autoencoder II is defined as

$Error_{reconstruct} = \frac{1}{N}\sum_p \frac{1}{R_p}\sum_r \sum_i \left(v''_{x(p,r,i)} - v_{x(p,r,i)}\right)^2$, where $v_{x(p,r,i)}$ denotes the true value of attribute i of the node embedding of residue r in pocket p, $v''_{x(p,r,i)}$ denotes the reconstructed value of attribute i of the node embedding of residue r in pocket p by Autoencoder II. N is the total number of pockets, and R_p is the number of residues in pocket p.

2: The percentage of error is defined as $\frac{Error_{reconstruct}}{Error_{max}}$, where $Error_{max}$ of Autoencoder I is defined as $\sum_i 2 * v_{\max\_i}{}^2$ and $Error_{max}$ of Autoencoder II is defined as $\sum_i v_{\max\_i}{}^2$, where $v_{\max\_i}$ denotes the maximum possible value of attribute i in vector v.

Table S7. AUC Performance on MUV dataset using alternative structures

Target	Original MUV PDB	Alternative PDB	AUC Performance		
			Original Pocket	Alternative Pocket	Dummy Pocket
832	1AU8	1KYN	0.530	0.549	0.552
859	5CXV	6OIJ	0.619	0.629	0.536
692	1YOW	4QJR	0.536	0.538	0.446
600	1YOW	4QJR	0.583	0.577	0.372
689	2Y6O	2XYU	0.717	0.699	0.6
548	3POO	6BYR	0.698	0.669	0.390
	3POO	5BX7	0.698	0.758	0.390
	3POO	5BX6	0.698	0.766	0.390
	3POO	6BYS (Apo structure)	0.698	0.733	0.390
713	5TN7	5U2D	0.591	0.612	0.532
	5TN7	5T92	0.591	0.597	0.532

For all MUV targets, our model showed comparable performance using the original and alternative structures of the same targets as input, suggesting that the model is generally robust to the choice of input pocket. Both original and alternative pockets generally performed significantly better than the dummy pockets.