Supporting information of

DMInet: an accurate and highly flexible deep learning framework for drug discovery with membrane selectivity

Guang Chen

Institute of Materials Science, University of Connecticut

Storrs CT, 06269

Contents

S1. Representation of the Martini CG molecules 2

S2. The architecture used for CNN model 3

S3. The complete architecture of the DMInet 4

S4. Training results: loss evolution, and MAE evolution 5

S5. Comparison of PMF profiles of all test samples from DMInet prediction and true values 6

S6. DMInet using three inputs 7

S7. References 8

S1. Representation of the Martini CG molecules

The polar, apolar and nonpolar types of Martini CG beads are listed as follows (Table S1). There are in total 14 different bead types in this data base [1].

Table S1: the CNN component of the DMInet.

Polarity	**Type**
Apolar	C1, C2, C3, C4, C5
Nonpolar	N0, Na, Nd, Nda
Polar	P1, P2, P3, P4, P5

To construct the matric representation of the Martini CG molecules, graph idea of the linear CG molecules is applied. Any type of Martini CG molecules can be represented by the summation of an adjacency matrix and a degree matrix.

S2. The architecture used for CNN model

The DMInet framework was developed under the Tensorflow [2] mainly using Keras package [3]. The following operations were applied on the 2-d images of the Martini CG molecules in the CNN component.

Table S2: the CNN component of the DMInet.

Layer	**Parameters**	**Inputs**
Conv2D	5 filters with 3 by 3 kernel	14 by 14 images
BatchNormalization	–	–
MaxPooling2D	2 by 2	–
Conv2D	3 filters with 3 by 3 kernel	–
BatchNormalization	–	–
MaxPooling2D	2 by 2	–
Flatten	–	–

S3. The complete architecture of the DMInet

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 14, 14, 1)]	0	
conv2d (Conv2D)	(None, 14, 14, 5)	50	input_1[0][0]
batch_normalization (BatchNorma	(None, 14, 14, 5)	20	conv2d[0][0]
max_pooling2d (MaxPooling2D) [0]	(None, 7, 7, 5)	0	batch_normalization[0]
conv2d_1 (Conv2D)	(None, 7, 7, 3)	138	max_pooling2d[0][0]
batch_normalization_1 (BatchNor	(None, 7, 7, 3)	12	conv2d_1[0][0]
max_pooling2d_1 (MaxPooling2D) [0][0]	(None, 3, 3, 3)	0	batch_normalization_1
flatten (Flatten)	(None, 27)	0	max_pooling2d_1[0][0]
input_2 (InputLayer)	[(None, 1)]	0	
concatenate (Concatenate)	(None, 28)	0	flatten[0][0] input_2[0][0]
dense_1 (Dense)	(None, 30)	870	concatenate[0][0]
batch_normalization_2 (BatchNor	(None, 30)	120	dense_1[0][0]
dense_2 (Dense) [0][0]	(None, 20)	620	batch_normalization_2
batch_normalization_3 (BatchNor	(None, 20)	80	dense_2[0][0]
dense_3 (Dense) [0][0]	(None, 10)	210	batch_normalization_3

Fig. S2: the final architecture of DMInet.

S4. Training results: loss evolution, and MAE evolution

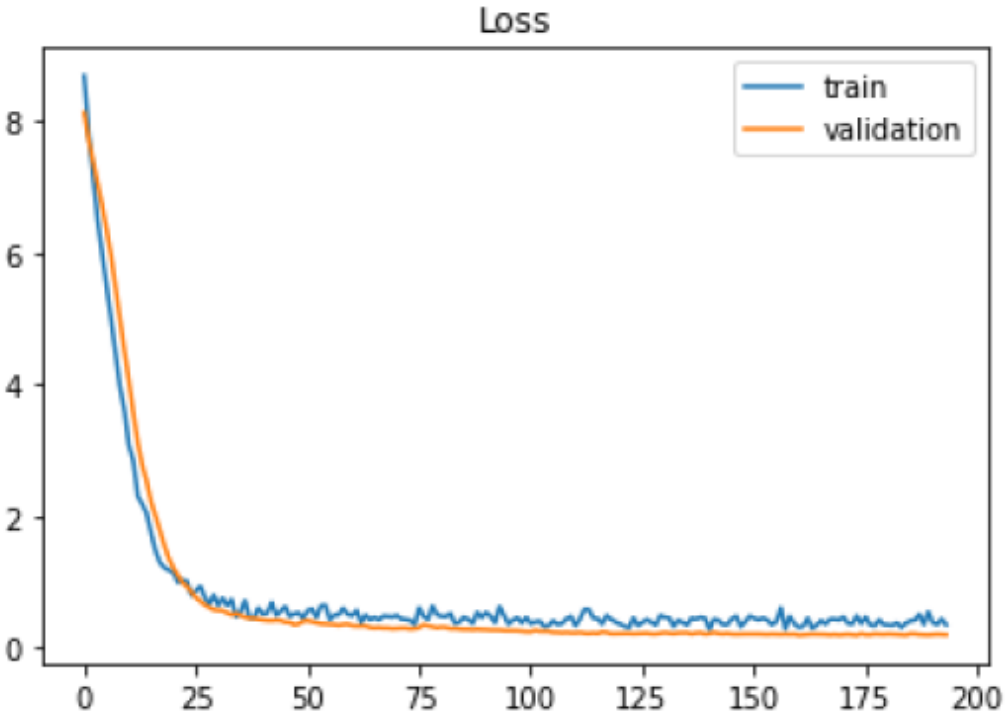

Fig S3: the learning curves of loss during model training

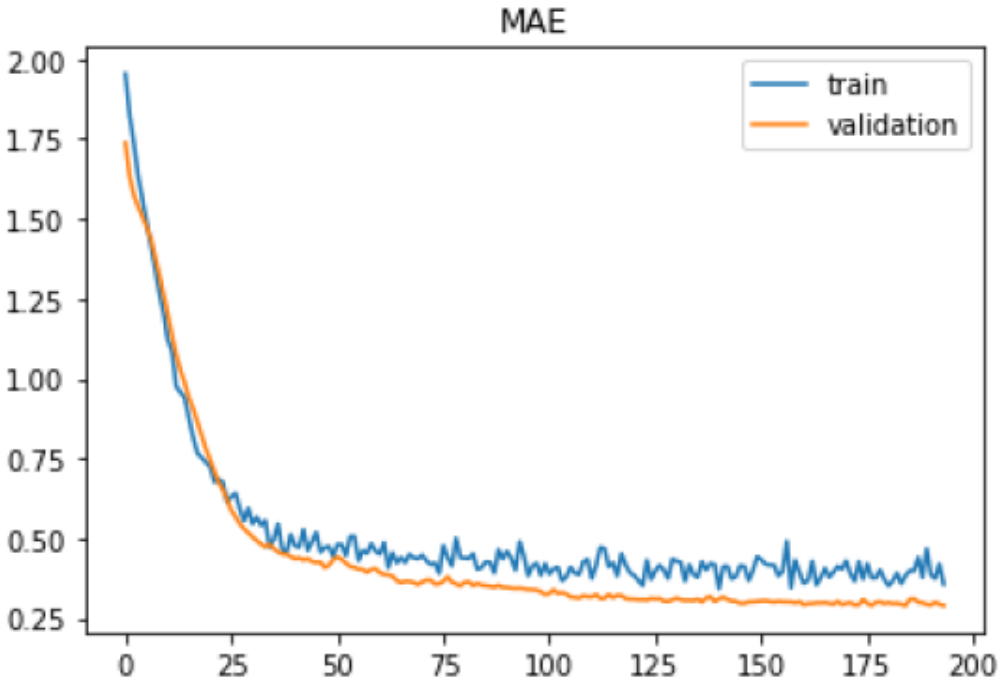

Fig S4: the learning curves of MAE metric during model training.

S5. Comparison of PMF profiles of all test samples from DMInet prediction and true values

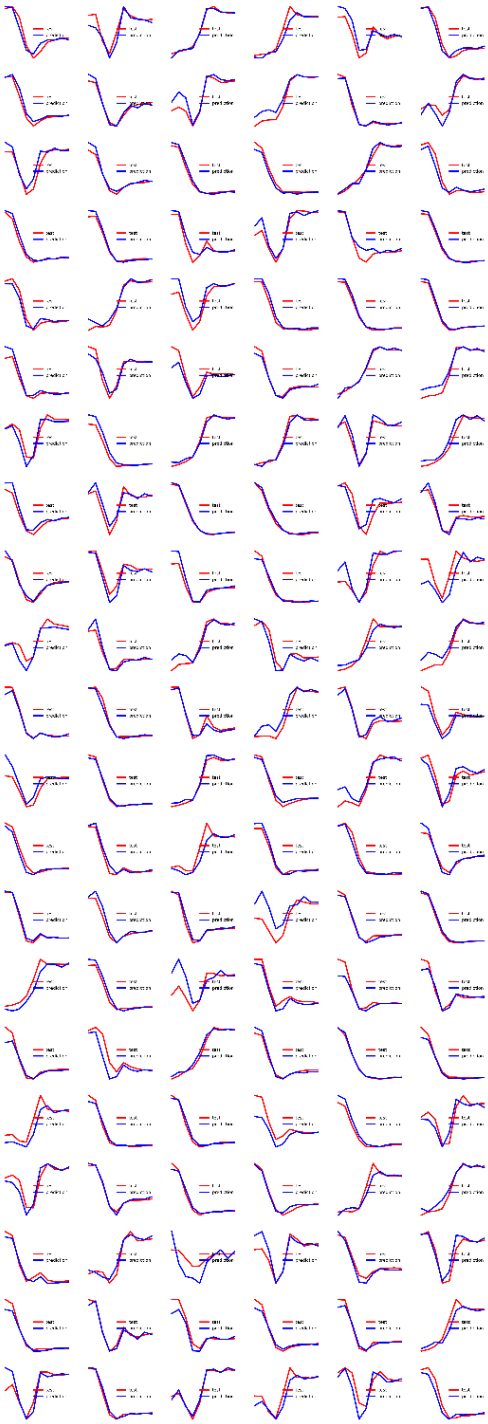

Fig. S5: the comparison of DMInet prediction and ground true on total test samples.

S6. DMInet using three inputs

Using two inputs (Martini CG molecules and membrane) and three inputs (Martini CG molecules, membrane, and spatial distance along the membrane thickness) are compared. Fig. S6 is the comparison of the free energy barrier between DMInet prediction and true values using three inputs.

Using two inputs can give better results, and thus two inputs are selected in the final DMInet model.

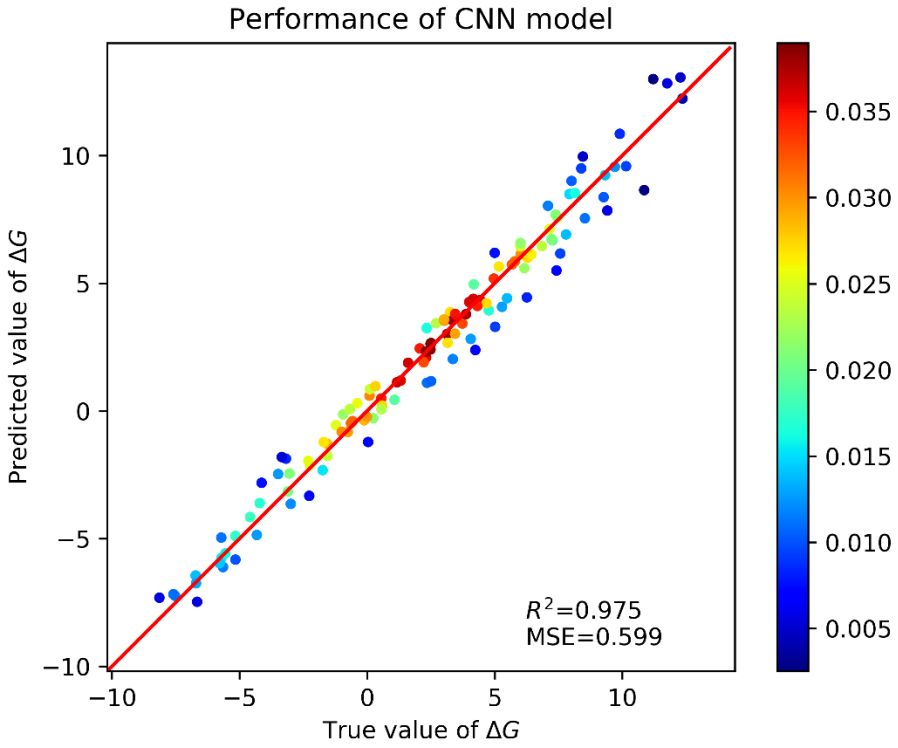

Fig. S6: the R^2 score of DMInet using three inputs.

S7. References

[1] Hoffmann, Christian, et al. "Molecular dynamics trajectories for 630 coarse-grained drug-membrane permeations." Scientific Data 7.1 (2020): 1-7.

[2] Abadi, Martín, et al. "Tensorflow: A system for large-scale machine learning." 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). 2016.

[3] F. Chollet, et al., Keras, 2015. URL: https://keras.io/

[4] Carpenter, Timothy S., et al. "A method to predict blood-brain barrier permeability of drug-like compounds using molecular dynamics simulations." Biophysical journal 107.3 (2014): 630-641.