Supporting Information for:

Binding Affinity Prediction by Pairwise Function Based on Neural Network

Fangqiang Zhu[1*], Xiaohua Zhang[1], Jonathan E. Allen[2], Derek Jones[2], Felice C. Lightstone[1]

1. Biochemical and Biophysical Systems Group, Biosciences and Biotechnology Division, Physical and Life Sciences Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA
2. Global Security, Computing Applications Division, Computing Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, Livermore, CA 94550, USA

[*] Correspondence: zhu13@llnl.gov

Table S1. Model training with different initial parameters.

Model #	Stage 1 (Adam)		Stage 2 (L-BFGS)	
	Train set	*Dev set*	*Train set*	*Dev set*
1	5.8	5.4	2.6	2.7
2	6.3	6.4	2.7	2.9
3	3.2	3.4	2.4	2.6
4	9.1	8.6	2.9	3.0
5	13.1	12.8	9.7	9.1
6	3.5	3.6	2.6	2.7
7	4.4	4.7	3.2	3.6
8	5.3	5.5	2.3	2.8
9	2.8	2.9	2.3	2.5
10	3.0	3.2	2.3	2.8
11	3.0	3.1	2.1	2.6
12	4.2	4.2	2.5	2.7
13	2.8	3.1	2.2	2.7
14	3.5	3.7	2.4	2.9
15	6.3	6.3	5.6	5.8
16	12.2	13.0	12.1	12.9
17	12.5	13.3	12.1	12.9
18	3.7	3.7	2.0	2.4
19	6.6	6.7	3.5	3.6
20	4.1	4.0	2.2	2.7

Each of the twenty models was trained using a different set of random initial parameters for the neural network. The training consists of two stages (see Methods in the main text): a stochastic gradient descent minimization with the Adam algorithm, followed by a full-batch L-BFGS minimization. The mean square errors on the *test set* and *dev set* after each stage are listed in the table for each model. Model 18 has the best performance among all these twenty models and is thus taken as the final model in this study.

Table S2. Performance of the pairwise model on randomly shuffled affinity data.

	RMSE	Pearson's r	Spearman's ρ
Train set	2.26	0.10	0.10
Test set	2.54	-0.04	-0.05

All the affinities in the dataset here were randomly shuffled first. We then trained a model using the *train set* and evaluated its performance on both the *train set* and the *test set*.