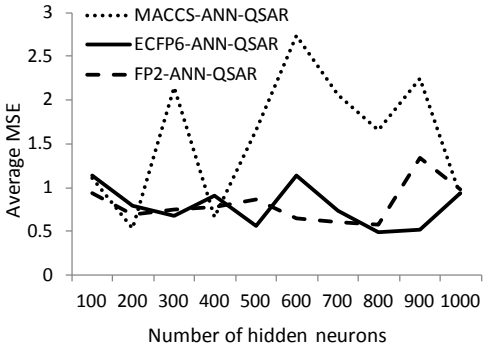
Supporting Information

Molecular Fingerprint-based Artificial Neural Networks QSAR for Ligand Biological Activity Predictions

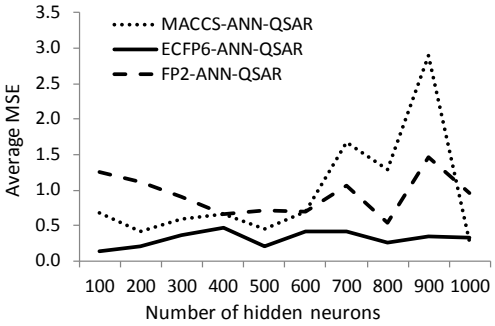
Kyaw-Zeyar Myint, Lirong Wang, Qin Tong and Xiang-Qun Xie.

Contents



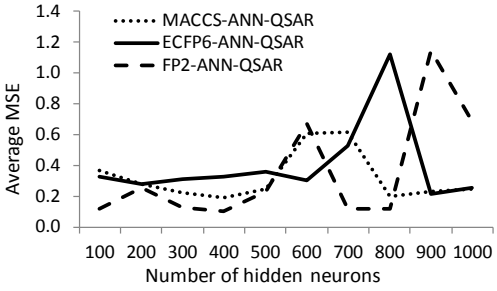
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	800	0.208	0.770	0.489
FP2-ANN-QSAR	800	0.368	0.789	0.579
MACCS-ANN-QSAR	200	0.337	0.742	0.539

Figure SF1. Cross-validation results of three FANN-QSAR models on ACE data set



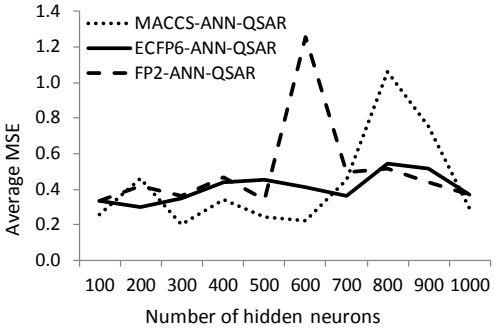
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	100	0.100	0.184	0.142
FP2-ANN-QSAR	800	0.381	0.685	0.533
MACCS-ANN-QSAR	1000	0.176	0.391	0.283

Figure SF2. Cross-validation results of three FANN-QSAR models on AchE data set



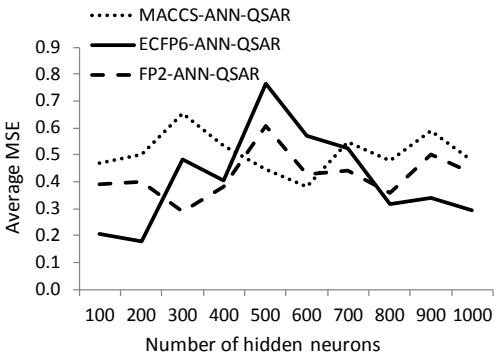
	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	900	0.100	0.332	0.216
FP2-ANN-QSAR	400	0.100	0.104	0.102
MACCS-ANN-QSAR	400	0.100	0.285	0.193

Figure SF3. Cross-validation results of three FANN-QSAR models on BZR data set



	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	200	0.186	0.418	0.302
FP2-ANN-QSAR	100	0.187	0.482	0.335
MACCS-ANN-QSAR	300	0.112	0.293	0.202

Figure SF4. Cross-validation results of three FANN-QSAR models on COX2 data set



	Optimal no. of hidden neurons	Training Set MSE	Test Set MSE	Average MSE
ECFP6-ANN-QSAR	200	0.100	0.261	0.180
FP2-ANN-QSAR	300	0.241	0.340	0.291
MACCS-ANN-QSAR	600	0.207	0.557	0.382

Figure SF5. Cross-validation results of three FANN-QSAR models on DHFR data set

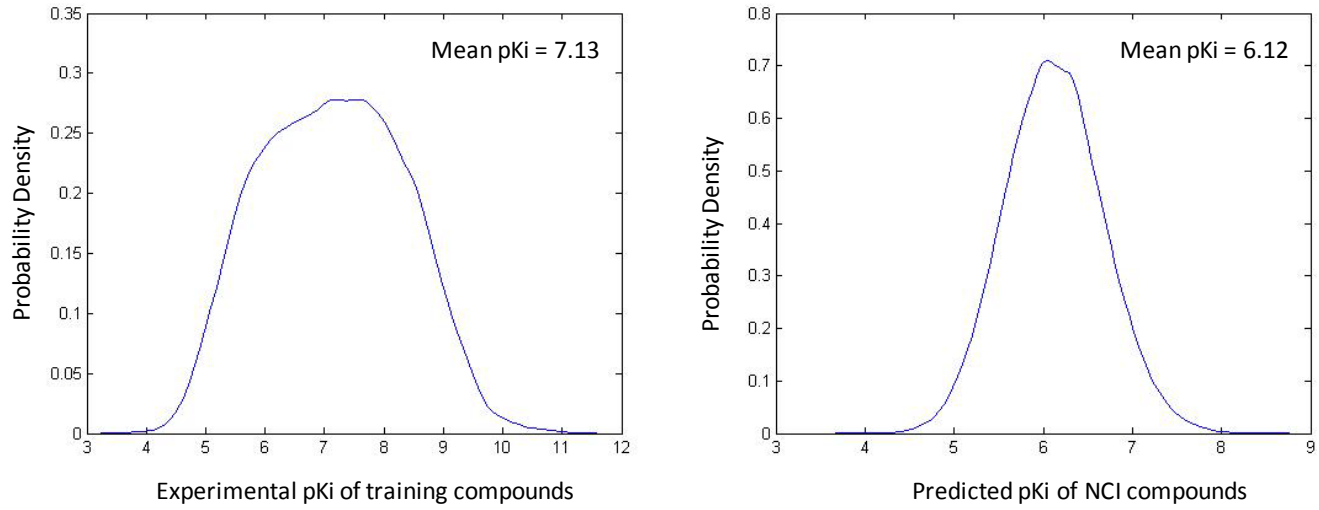


Figure SF6. Distributions of experimental binding affinities of training compounds (CBID database) and predicted binding affinities for NCI compounds

Round	Training MSE	Validation MSE	Avg. MSE	Optimal No. of Hidden Neurons	r^2 training	r^2 validation
1	0.169	0.733	0.451	1000	0.87	0.47
2	0.126	0.832	0.479	500	0.91	0.49
3	0.362	0.645	0.504	600	0.74	0.53
4	0.262	0.555	0.409	400	0.81	0.59
5	0.332	0.776	0.554	800	0.74	0.49
6	0.214	0.830	0.522	200	0.84	0.41
7	0.207	0.622	0.415	500	0.85	0.56
8	0.181	0.764	0.472	400	0.87	0.51
9	0.255	0.563	0.409	800	0.82	0.60
10	0.227	0.713	0.470	900	0.84	0.52
11	0.217	0.636	0.427	600	0.83	0.51
12	0.239	0.832	0.535	600	0.82	0.50
13	0.185	0.729	0.457	800	0.87	0.43
14	0.137	0.656	0.396	400	0.90	0.54
15	0.389	0.565	0.477	900	0.70	0.55
16	0.159	0.727	0.443	500	0.87	0.50
17	0.303	0.709	0.506	500	0.78	0.50
18	0.182	0.515	0.348	600	0.87	0.64
19	0.219	0.681	0.450	900	0.83	0.53
20	0.208	0.567	0.388	1000	0.85	0.53

Table ST1. Twenty rounds of model training and cross-validation results using CB_2 ligand data set before testing compounds from NCI database