Supporting Information

Random Forest Refinement of Pairwise Potentials for Protein-Ligand Decoy Detection

Jun Pei,[§] *Zheng Zheng,* [§‡] *Hyunji Kim,* [§a] *Lin Frank Song,* [§] *Sarah Walworth,* [§b] *Margaux R. Merz,*[§c] *and Kenneth M. Merz Jr.** [§†]

[§]Department of Chemistry, Michigan State University, 578 S. Shaw Lane, East Lansing, Michigan 48824, United States

[†]Institute for Cyber Enabled Research, Michigan State University, 567 Wilson Road, East Lansing, Michigan 48824, United States

Table. s1 Summary of peak positions and number of probability functions at each peak positions in GARF

Peak position / Å	number of Probability functions
2.5	2
2.6	2
2.7	5
2.8	14
2.9	9
3.0	6
3.1	5
3.2	2
3.3	14
3.4	23
3.5	23
3.6	23
3.7	53
3.8	43
3.9	34
4.0	36
4.1	24
4.2	28
4.3	38
4.4	30
4.5	8
4.6	12
4.7	13
4.8	17
4.9	8
5.0	4
5.1	4

Table. s2 Accuracy values for different training set sizes from RF models with original and uniform GARF

Training size / %	Original GARF			Uniform GARF		
	Averaged accuracy	Highest accuracy	Lowest accuracy	Averaged accuracy	Highest accuracy	Lowest accuracy
5	0.883	0.926	0.833	0.900	0.929	0.867
10	0.923	0.939	0.912	0.916	0.936	0.879
15	0.919	0.941	0.884	0.933	0.948	0.913
20	0.926	0.947	0.887	0.936	0.946	0.902
25	0.938	0.953	0.918	0.939	0.960	0.919
30	0.943	0.953	0.935	0.945	0.965	0.933
35	0.947	0.962	0.928	0.942	0.960	0.914
40	0.945	0.959	0.924	0.946	0.951	0.933
45	0.941	0.951	0.925	0.949	0.969	0.922
50	0.953	0.966	0.938	0.944	0.957	0.935
55	0.952	0.973	0.916	0.945	0.968	0.923
60	0.954	0.975	0.935	0.950	0.968	0.937
65	0.945	0.964	0.937	0.949	0.965	0.921

70	0.954	0.980	0.922	0.955	0.972	0.924
75	0.953	0.980	0.922	0.952	0.977	0.924
80	0.953	0.969	0.942	0.953	0.980	0.918

Table. s3 Atom types in the GARF potential database[a]

Atom type	definition
Protein atom types	
C	sp^2 carbonyl carbon and aromatic carbon with hydroxyl substituent in tyrosine
C*	sp^2 aromatic carbon in 5-membered ring with one substituent
CA	sp^2 aromatic carbon in 6-membered ring with one substituent
CB	sp^2 aromatic carbon at junction between 5- and 6-membered rings
CC	sp^2 aromatic carbon in 5-membered ring with one substituent and next to a nitrogen atom
CN	sp^2 aromatic junction carbon in between 5- and 6-membered rings
CR	sp^2 aromatic carbon in 5-membered ring between two nitrogen atoms and bonded to one hydrogen atom (in HIS)
CT	sp^3 carbon with four explicit substituents
CV	sp^2 aromatic carbon in 5-membered ring bonded to one nitrogen atom and bonded to an explicit hydrogen
CW	sp^2 aromatic carbon in 5-membered ring bonded to one N−H group and an explicit hydrogen
N	sp^2 nitrogen in amide group
N2	sp^2 nitrogen in base NH_2 group or arginine NH_2
N3	sp^2 nitrogen with four substituents
NA	sp^2 nitrogen in 5-membered ring with hydrogen attached
NB	sp^2 nitrogen in 5-membered ring with lone pairs
O	carbonyl oxygen
O2	carboxyl oxygen
OH	alcohol oxygen
S	sulfur in disulfide linkage or methionine
SH	sulfur in cysteine
Ligand atom types	
C.3	sp^3 carbon without polar group substituent
C.2	sp^2 carbonyl carbon without polar group substituent
C.1	sp carbon
C.ar	sp^2 aromatic carbon without polar group substituent
O.3	alcohol oxygen
O.3P	ether oxygen
O.2	carbonyl oxygen
O.co2	carboxylate oxygen
O.2v	sulfate/phosphate oxygen
N.2	sp/sp^2/aromatic nitrogen
N.1h	sp^3 nitrogen with one hydrogen atom attached
N.2h	sp^3 nitrogen with two hydrogen atoms attached
N.3h	sp^3 nitrogen with three hydrogen atoms attached
P	Phosphorus
F	Fluorine

Cl	Chlorine
Br	Bromine
I	Iodine
C.cat	carbon cation
S.3	thiol/thioether sulfur
S.o	sulfoxide sulfur
C.3X	sp^3 carbon with polar group substituent
C.2X	sp^2 carbonyl carbon with polar group substituent
C.arX	sp^2 aromatic carbon with polar group substituent

a. This table is as same as Table 2 in GARF paper, reference 1.

Table s4. Summary of importance factor for each atom pair in CASF-2013[2]

Atom pair name	importance factor	Atom pair name	importance factor	Atom pair name	importance factor	Atom pair name	importance factor
CT__C.2x	0.0435	CR__C.2	0.0017	CV__N.1h	0.0007	SH__O.2	0.0002
CT__C.2	0.0409	CB__C.3x	0.0017	NA__C.arx	0.0007	S__S.3	0.0002
O__C.2	0.0296	CV__N.2	0.0017	CW__O.co2	0.0007	OH__Cl	0.0002
C__C.2x	0.0284	CV__C.3x	0.0017	O2__S.3	0.0007	CV__S.3	0.0002
CA__C.2	0.0263	NB__O.co2	0.0017	C__P	0.0007	CC__P	0.0002
C__C.2	0.0244	CC__C.2	0.0017	CC__N.3h	0.0007	Cs__S.3	0.0002
N__C.2	0.0226	NA__O.3	0.0017	CC__C.arx	0.0007	Cs__Br	0.0001
O2__C.3x	0.0208	CV__C.2	0.0017	SH__C.2	0.0007	Cs__N.3h	0.0001
CA__C.3	0.0194	O2__O.co2	0.0016	NB__O.2v	0.0007	CC__S.o	0.0001
CT__C.3	0.0194	O__O.3p	0.0016	C__S.3	0.0007	CB__Cl	0.0001
CA__O.3	0.0162	Cs__C.3	0.0016	OH__S.3	0.0006	NA__P	0.0001
N__C.2x	0.0161	C__O.3p	0.0016	SH__C.3x	0.0006	S__S.o	0.0001
CT__N.2	0.0159	NA__C.3	0.0016	CW__C.2x	0.0006	CC__S.3	0.0001
O__C.2x	0.0158	N__O.3	0.0016	CN__O.3p	0.0006	NA__Cl	0.0001
CA__C.2x	0.0155	S__C.2x	0.0016	C__F	0.0006	CR__Cl	0.0001
CT__C.3x	0.0152	C__N.2	0.0016	N2__C.arx	0.0006	CB__Br	0.0001
C__C.3x	0.015	N3__C.ar	0.0015	CC__O.2v	0.0006	CT__Br	0.0001
O2__O.3	0.0122	SH__C.ar	0.0015	C__Cl	0.0006	CV__Cl	0.0001
O2__N.2h	0.0118	CR__O.3p	0.0015	N2__S.3	0.0006	CB__N.3h	0.0001
O2__C.2	0.0114	OH__N.1h	0.0015	NA__N.2h	0.0006	CN__F	0.0001
OH__C.2	0.0113	N2__C.3x	0.0015	CN__N.1h	0.0006	CV__P	0.0001
CT__O.3	0.011	N3__S.o	0.0015	Cs__C.ar	0.0006	CA__Br	0.0001
CT__N.1h	0.0101	N3__C.arx	0.0015	CA__O.2v	0.0006	OH__Br	0.0001
CA__C.3x	0.0097	NB__C.ar	0.0015	CN__N.2	0.0006	N2__Cl	0.0001
N__C.ar	0.0091	CW__C.3	0.0015	CC__N.1h	0.0005	CN__S.o	0.0001
O2__C.2x	0.0089	CT__N.3h	0.0015	Cs__C.2	0.0005	SH__Cl	0.0001
N2__O.co2	0.0088	CA__N.3h	0.0014	O__F	0.0005	CN__Cl	0.0001
CT__C.ar	0.0088	OH__N.3h	0.0014	NA__O.2v	0.0005	CC__Cl	0.0001
CT__O.co2	0.0088	N__O.3p	0.0014	SH__N.2h	0.0005	OH__C.cat	0.0001
OH__C.2x	0.0079	S__C.3	0.0014	CB__O.3p	0.0005	NB__Cl	0.0001
NA__C.3x	0.0079	CV__O.co2	0.0014	N3__N.2h	0.0005	CB__F	0.0001
N__C.3x	0.0078	CV__C.3	0.0014	O2__O.2v	0.0005	N3__Cl	0.0001
CT__N.2h	0.007	O2__N.1h	0.0014	CW__O.3p	0.0005	CW__Cl	0.0001
CT__C.arx	0.0068	N3__C.2x	0.0013	O__C.cat	0.0005	Cs__Cl	0.0001
CT__O.3p	0.0067	N2__O.3p	0.0013	CV__O.2v	0.0005	O2__C.cat	0.0001
NA__C.2x	0.0064	CC__N.2	0.0013	Cs__O.2	0.0005	CW__N.3h	0.0001
C__C.3	0.0062	O2__O.3p	0.0013	S__N.1h	0.0005	N__I	0.0001

N__O.2	0.0058	CC__C.3x	0.0013	N2__P	0.0005	CN__Br	0.0001
CN__C.3x	0.0057	CA__N.2h	0.0013	CN__O.2	0.0005	CW__Br	0.0001
O2__N.3h	0.0057	CR__O.2v	0.0013	NA__N.3h	0.0005	CW__F	0.0001
NB__C.3x	0.0054	Cs__O.3	0.0013	OH__O.2v	0.0005	CN__N.3h	0.0001
CV__O.3	0.0052	NA__N.1h	0.0013	CB__N.2	0.0005	N2__C.cat	0.0001
CR__C.3x	0.005	CT__Cl	0.0013	C__S.o	0.0005	S__Br	0.0001
O__C.3	0.0048	CC__O.co2	0.0013	CB__N.1h	0.0004	NA__Br	0.0001
O2__C.ar	0.0048	CC__C.3	0.0013	CW__N.1h	0.0004	Cs__F	0.0001
O__C.ar	0.0047	N__P	0.0013	CW__C.ar	0.0004	CA__C.cat	0.0001
C__C.ar	0.0047	N3__N.2	0.0013	O__S.o	0.0004	N2__Br	0.0001
CR__O.3	0.0044	N2__C.3	0.0012	OH__F	0.0004	SH__S.3	0.0001
NA__C.2	0.0043	CC__C.ar	0.0012	S__N.3h	0.0004	C__I	0
NB__O.3	0.004	O2__O.2	0.0012	CT__P	0.0004	SH__C.cat	0
CT__O.2	0.004	CR__C.3	0.0012	N__S.o	0.0004	CT__I	0
C__N.1h	0.0039	CN__C.2	0.0012	CN__O.co2	0.0004	O__I	0
N3__C.3x	0.0037	N3__O.3p	0.0012	N__F	0.0004	CB__O.2v	0
N3__O.2v	0.0037	CR__C.ar	0.0012	CA__P	0.0004	SH__I	0
CA__O.co2	0.0035	OH__N.2h	0.0012	CW__O.2	0.0004	Cs__O.2v	0
CC__C.2x	0.0035	CV__C.ar	0.0012	N3__S.3	0.0004	Cs__S.o	0
O__C.3x	0.0035	N2__N.2h	0.0012	CB__C.arx	0.0004	CR__C.cat	0
N3__O.co2	0.0035	N2__N.1h	0.0012	SH__N.3h	0.0004	CB__S.o	0
C__O.co2	0.0034	OH__O.2	0.0012	Cs__O.co2	0.0004	NA__C.cat	0
C__O.3	0.0033	O__N.2h	0.0012	CN__S.3	0.0004	SH__O.2v	0
CA__N.2	0.0033	S__O.3	0.0011	Cs__N.1h	0.0004	S__P	0
O__O.2	0.0033	CC__O.3p	0.0011	CW__C.2	0.0004	CC__C.cat	0
CA__O.2	0.0033	S__C.3x	0.0011	OH__P	0.0004	CV__C.cat	0
NB__C.2	0.0032	N3__C.2	0.0011	N__Cl	0.0004	OH__I	0
C__O.2	0.0031	N2__O.3	0.0011	CN__C.ar	0.0004	SH__Br	0
C__C.arx	0.0031	S__O.2	0.0011	CW__N.2	0.0003	NB__C.cat	0
CB__C.2x	0.0031	NA__N.2	0.0011	O2__P	0.0003	CW__O.2v	0
CW__C.3x	0.003	CV__N.3h	0.0011	Cs__N.2h	0.0003	N3__C.cat	0
CA__C.ar	0.0029	NB__O.3p	0.0011	SH__F	0.0003	SH__S.o	0
N2__C.2	0.0028	CT__S.o	0.0011	CR__S.3	0.0003	N3__Br	0
NB__C.2x	0.0028	NB__N.1h	0.0011	N__C.cat	0.0003	NB__I	0
O__C.arx	0.0028	NB__N.2	0.0011	NB__F	0.0003	NA__I	0
N__C.arx	0.0028	SH__C.2x	0.0011	Cs__C.arx	0.0003	CR__I	0
NA__O.co2	0.0027	N2__N.2	0.0011	CB__C.ar	0.0003	O2__I	0
N__C.3	0.0027	SH__N.2	0.001	CR__N.3h	0.0003	CV__Br	0
OH__C.ar	0.0027	S__N.2	0.001	O__Cl	0.0003	Cs__C.cat	0
CV__C.2x	0.0027	CV__O.3p	0.001	N2__O.2v	0.0003	CV__I	0
N__N.2h	0.0026	N__O.2v	0.001	NA__S.3	0.0003	CC__Br	0
CC__O.3	0.0026	S__O.co2	0.001	Cs__N.2	0.0003	Cs__I	0
CT__O.2v	0.0025	SH__O.3p	0.001	SH__O.3	0.0003	CA__I	0
CA__O.3p	0.0025	CC__N.2h	0.001	CB__O.co2	0.0003	SH__P	0
O__N.1h	0.0025	CA__S.3	0.001	C__C.cat	0.0003	NB__Br	0
OH__O.co2	0.0025	CR__N.1h	0.001	SH__O.co2	0.0003	CB__C.cat	0
S__C.2	0.0025	S__O.3p	0.001	O2__F	0.0003	CW__S.o	0
N__N.1h	0.0024	CW__O.3	0.001	CN__N.2h	0.0003	O2__Br	0
OH__O.3	0.0024	NA__O.3p	0.001	N3__F	0.0003	CR__Br	0
CT__F	0.0024	N3__C.3	0.0009	O2__S.o	0.0003	Cs__P	0
Cs__C.3x	0.0024	CB__C.2	0.0009	CR__F	0.0003	CN__P	0
O2__C.3	0.0023	NA__C.ar	0.0009	CC__F	0.0003	CW__P	0
OH__C.3x	0.0023	NB__N.2h	0.0009	CW__S.3	0.0003	CW__C.cat	0
C__O.2v	0.0023	S__C.arx	0.0009	O__Br	0.0003	CB__P	0
CB__C.3	0.0022	N3__N.1h	0.0008	NB__S.3	0.0003	CC__I	0
N3__O.3	0.0022	CR__N.2h	0.0008	CB__S.3	0.0003	CN__C.cat	0

N__O.co2	0.0022	N2__N.3h	0.0008	CA__S.o	0.0003	CB__I	0
OH__O.3p	0.0021	N3__O.2	0.0008	CR__S.o	0.0003	N3__I	0
O__O.3	0.0021	N2__C.ar	0.0008	N3__P	0.0003	OH__C.1	0
O2__N.2	0.0021	CB__O.2	0.0008	NB__S.o	0.0003	O__C.1	0
N__N.2	0.0021	Cs__C.2x	0.0008	CN__C.arx	0.0003	N2__C.1	0
CN__O.3	0.0021	CA__F	0.0008	CW__C.arx	0.0003	S__I	0
N__N.3h	0.0021	CV__C.arx	0.0008	S__Cl	0.0003	CB__C.1	0
OH__C.3	0.0021	CV__O.2	0.0008	CR__P	0.0003	N2__I	0
CA__C.arx	0.0021	N__S.3	0.0008	S__F	0.0002	CV__C.1	0
SH__C.arx	0.002	NB__O.2	0.0008	N3__N.3h	0.0002	CC__C.1	0
C__N.3h	0.002	CR__C.arx	0.0008	NB__P	0.0002	S__C.1	0
CA__N.1h	0.002	CA__Cl	0.0008	NA__F	0.0002	C__C.1	0
CT__S.3	0.002	CV__N.2h	0.0008	CW__N.2h	0.0002	CW__I	0
CR__O.co2	0.002	O__O.2v	0.0008	CB__N.2h	0.0002	N3__C.1	0
OH__C.arx	0.002	CN__C.2x	0.0008	OH__S.o	0.0002	CN__I	0
O__N.2	0.0019	O__P	0.0008	CN__O.2v	0.0002	S__C.cat	0
O2__C.arx	0.0019	NB__C.arx	0.0008	O2__Cl	0.0002	Cs__C.1	0
C__N.2h	0.0019	N2__O.2	0.0008	CV__F	0.0002	CR__C.1	0
CN__C.3	0.0019	SH__N.1h	0.0008	N__Br	0.0002	CA__C.1	0
N2__C.2x	0.0018	CC__O.2	0.0008	CT__C.cat	0.0002	CT__C.1	0
S__C.ar	0.0018	NA__O.2	0.0008	S__N.2h	0.0002	NA__C.1	0
CB__O.3	0.0018	O__S.3	0.0008	N2__F	0.0002	CN__C.1	0
OH__N.2	0.0018	CR__O.2	0.0007	CV__S.o	0.0002	SH__C.1	0
O__N.3h	0.0018	Cs__O.3p	0.0007	S__O.2v	0.0002	N__C.1	0
CR__C.2x	0.0018	NB__N.3h	0.0007	NA__S.o	0.0002	O2__C.1	0
O__O.co2	0.0017	CR__N.2	0.0007	C__Br	0.0002	NB__C.1	0
NB__C.3	0.0017	SH__C.3	0.0007	N2__S.o	0.0002	CW__C.1	0

References

(1) Zheng, Z.; Pei, J.; Bansal, N.; Liu, H.; Song, L. F.; Merz, K. M. Generation of Pairwise Potentials Using Multidimensional Data Mining. *J. Chem. Theory Comput.* **2018,** *14(10),* 5045–5067.

(2) Li, Y.; Liu, Z.; Li, J.; Han, L.; Liu, J.; Zhao, Z.; Wang, R. Comparative assessment of scoring functions on an updated benchmark: 1. compilation of the test set. *J. Chem. Inf. Model.* **2014,** *54(6)*, 1700–1716.