Supporting Information

In Silico Prediction of Chemical Ames Mutagenicity

Congying Xu[†], Feixiong Cheng[†], Lei Chen[†], Zheng Du[†], Weihua Li[†], Guixia Liu[*,†,‡], Philip W. Lee[†] and Yun Tang[*,†]

[†]Shanghai Key Laboratory of New Drug Design, School of Pharmacy, East China University of Science and Technology, 130 Meilong Road, Shanghai 200237, China

[‡] State key Laboratory of Drug Research, Shanghai Institute of Materia Medica, Chinese Academy of Sciences, 555 Zu Chongzhi Road, Shanghai 201203, China

* To whom correspondence should be addressed.

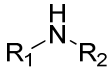
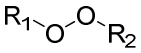
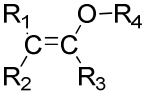
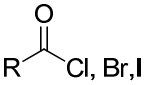
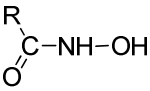
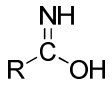
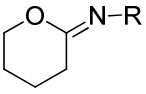
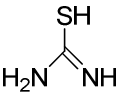
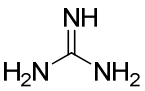
Tel: +86-21-64250811; Fax: +86-21-64253651

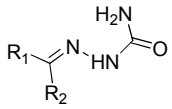
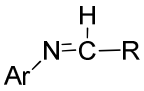
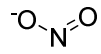
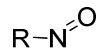
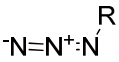
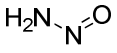
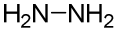
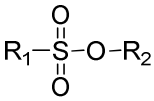
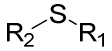
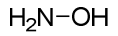
Email: gxliu@ecust.edu.cn (G. Liu), ytang234@ecust.edu.cn (Y. Tang)

Table S2. The performance of the balanced external validation set

Model Name	Q	SE	SP	AUC
PubChem-*k*NN	0.944	0.983	0.906	0.969
MACCS-*k*NN	0.906	0.992	0.821	0.958
PubChem-SVM	0.850	0.974	0.726	0.934

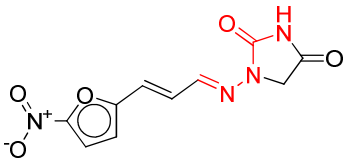
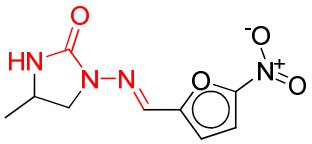
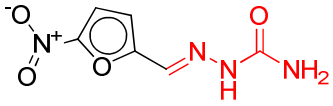
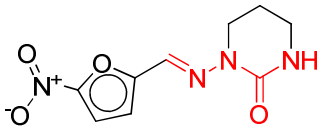
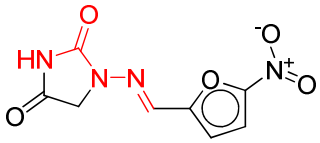
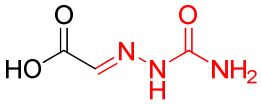
Table S3. The "Frequency of a fragment" enrichment factor of 25 representative substructure fragments for mutagens using FP4.

NO	SMARTS	Description	General tructure	IG	N_m	N_{Nm}	F_m
1	[IX1][CX4]	Alkyliodide		0.000	4866	12	1.287
2	[NX3H0+0,NX4H1+;!$([N][!c]);!$([N]*~[#7,#8,#15,#16])]	Tertiary_arom_amine		0.015	4866	453	1.439
3	[NX3H1+0,NX4H2+;$([N]([c])[C]);!$([N]*~[#7,#8,#15,#16])]	Secondary_mixed_amine		0.001	4866	97	1.290
4	[OX2H][OX2]	Hydroperoxide		0.001	4866	17	1.535
5	[OX2]([#6;!$(C=[N,O,S])])[CX3;$([H0][#6]),$([H1])]=[CX3]	Enolether		0.001	4866	74	1.309
6	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[ClX1]	Acylchloride		0.002	4866	29	1.605
7	[CX3;$([R0][#6]),$([H1R0])](=[OX1])[FX1,ClX1,BrX1,IX1]	Acylhalide		0.002	4866	29	1.605
8	[CX3;$([H0][#6]),$([H1])](=[OX1])[#7X3;$([H1]),$([H0][#6;!$(C=[O,N,S])])][$([OX2H]),$([OX1-])]	Hydroxamic_acid		0.001	4866	35	1.305
9	[CX3R0;$([H0][#6]),$([H1])](=[NX2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[$([OX2H]),$([OX1-])]	Imidoacid		0.002	4866	41	1.599
10	[#6R][#6X3R](=,:[#7X2;$([H1]),$([H0][#6;!$(C=[O,N,S])])])[OX2][#6;!$(C=[O,N,S])]	Imidolactone		0.000	4866	26	1.312
11	[#7X2;!$([#7][!#6])]=,:[#6X3]([#16X2&!$([#16][!#6]),SX1-])[#7X3;!$([#7][!#6])]	Isothiourea		0.002	4866	77	1.420
12	[N;v3X3,v4X4+][CX3](=[N;v3X2,v4X3+])[N;v3X3,v4X4+]	Guanidine		0.000	4866	29	1.309

13	[#7X2](=[#6])[#7X3][#6X3]([#7X3;!$([#7][#7])])=[OX1]	Semicarbazone		0.001	4866	14	1.601
14	[c]=[NX2;$([H1]),$([H0][#6;!$([C]=[N,S,O])])]	Iminoarene		0.001	4866	24	1.372
15	[NX2](=[OX1])[O;$([X2]),$([X1-])]	Nitrite		0.001	4866	8	1.716
16	[NX2](=[OX1])[!#7;!#8]	Nitroso		0.002	4866	51	1.535
17	[NX1]~[NX2]~[NX2,NX1]	Azide		0.004	4866	53	1.653
18	[NX2](=[OX1])N-*=O	Nitrosamide		0.004	4866	54	1.684
19	[NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])] [NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])]	Hydrazine		0.001	4866	30	1.391
20	[SX4;$([H1]),$([H0][#6])](=[OX1])(=[OX1])[OX2][#6;!$(C=[O,N,S])]	Sulfonic_ester		0.001	4866	30	1.430
21	[OX2r3]1[#6r3][#6r3]1	Epoxide		0.006	4866	342	1.316
22	[NX3H1r3]1[#6r3][#6r3]1	NH_aziridine		0.003	4866	29	1.716
23	[sX2]	Hetero_S		0.005	4866	286	1.330
24	[$([NX3](=O)=O),$([NX3+](=O)[O-])][!#8]	Nitro		0.060	4866	1448	1.459
25	[NX3;$([H2]),$([H1][#6]),$([H0]([#6])[#6]);!$(NC=[O,N,S])] [OX2;$([H1]),$(O[#6;!$(C=[N,O,S])])]	Hydroxylamine		0.001	4866	51	1.250

N_{Nm} is the number of mutagenicity class with specified fragment. Nm is the number of mutagens. Fm is the "Frequency of a fragment" enrichment factor of a specified fragment mutagenicity class

Table S4. Semicarbazone substructural alert and representative structures. M in Label means mutagens, N in Label means non-mutagens. The red fragment stands for the semicarbazone.

SMILES	CAS	Structure	Label
[O-][N+](=O)c1oc(\C=C\C=N\N2CC(=O)NC2=O)cc1	1672-88-4		M
CC1CN(\N=C\c2oc(cc2)[N+](=O)[O-])C(=O)N1	21638-36-8		M
NC(=O)N\N=C\c1oc(cc1)[N+](=O)[O-]	59-87-0		M
[O-][N+](=O)c1oc(\C=N\N2CCCNC2=O)cc1	75888-03-8		M
ON(=O)c1ccc(C=NN2CC(=O)NC2=O)o1	67-20-9		M
NC(=O)N\N=C\C(=O)O	928-73-4		N