Supporting information

Machine learning approach toward generating the focused molecule library targeting CAG repeat DNA

Qingwen Chen,[a] Takeshi Yamada,[a,c] Asako Murata,[a,d] Ayako Sugai,[a] Yasuyuki Matsushita,[b] and Kazuhiko Nakatani[a*]

[a]SANKEN (The Institute of Scientific and Industrial Research), Osaka University, 8-1 Mihogaoka, Ibaraki 567-0047, Japan. E-mail: nakatani@sanken.osaka-u.ac.jp
[b]Graduate School of Information Science and Technology, Osaka University, 1-5 Yamadaoka, Suita 565-0871, Japan. E-mail: yasumat@ist.osaka-u.ac.jp
[c]Current address: Nucleotide and Peptide Drug Discovery Center, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8519, Japan.
[d]Current address: Department of Material Sciences, Faculty of Engineering Sciences, Kyushu University, 6-1 Kasuga-koen, Kasuga, Fukuoka 816-8580, Japan.

Contents

Supporting Figures

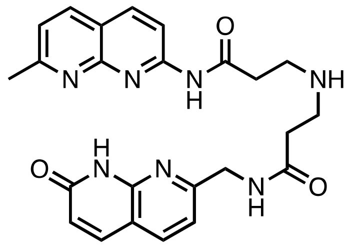


Fig. S1 The structure of **NA.**[1]

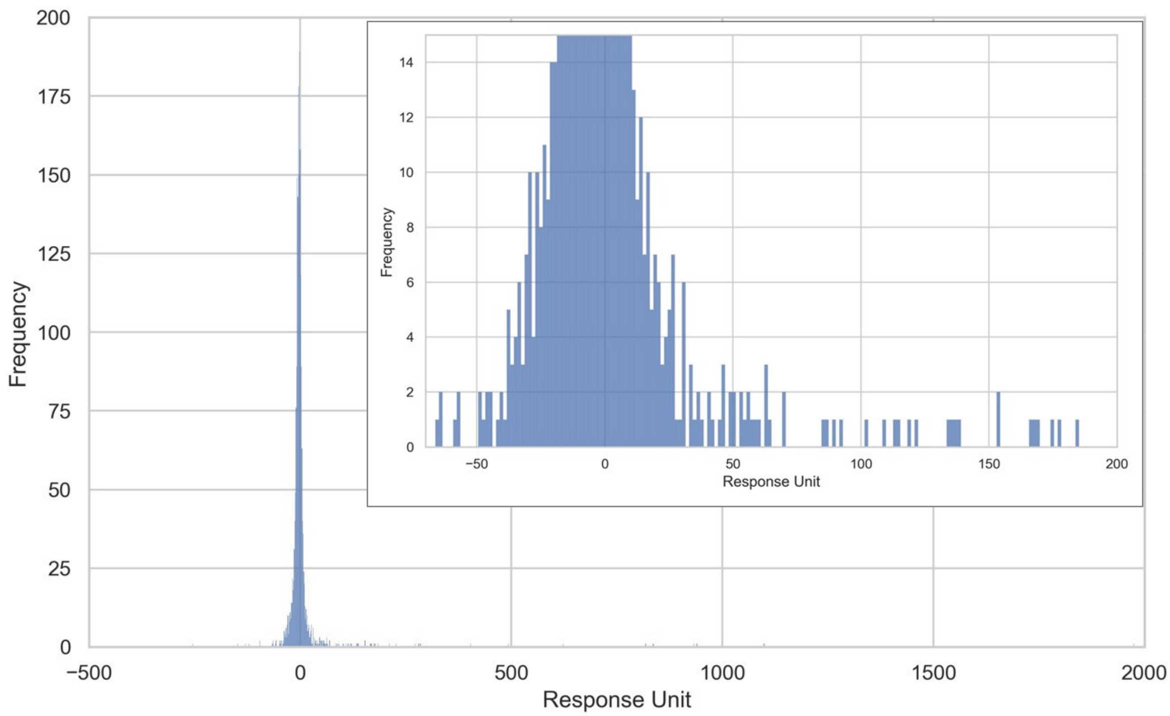


Fig. S2 A skew distribution of SPR response units for 2000 small molecules. Response strength for most molecules is near zero, suggesting these are likely non-hits. An enlarged plot is displayed in the top right for detailed viewing.

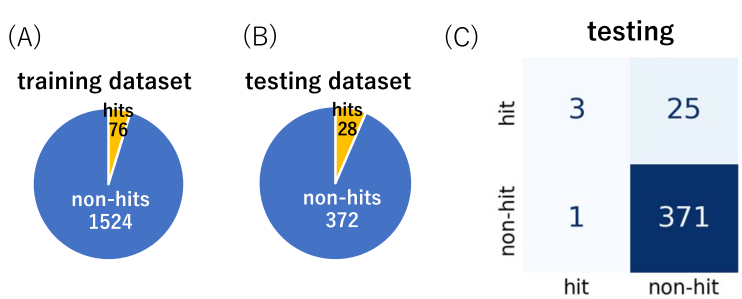


Fig. S3 The binary classification confusion matrixes in the initial classification. The data proportion of (A) training dataset and (B) testing dataset. (C) The confusion matrix of the classification results of the testing dataset.

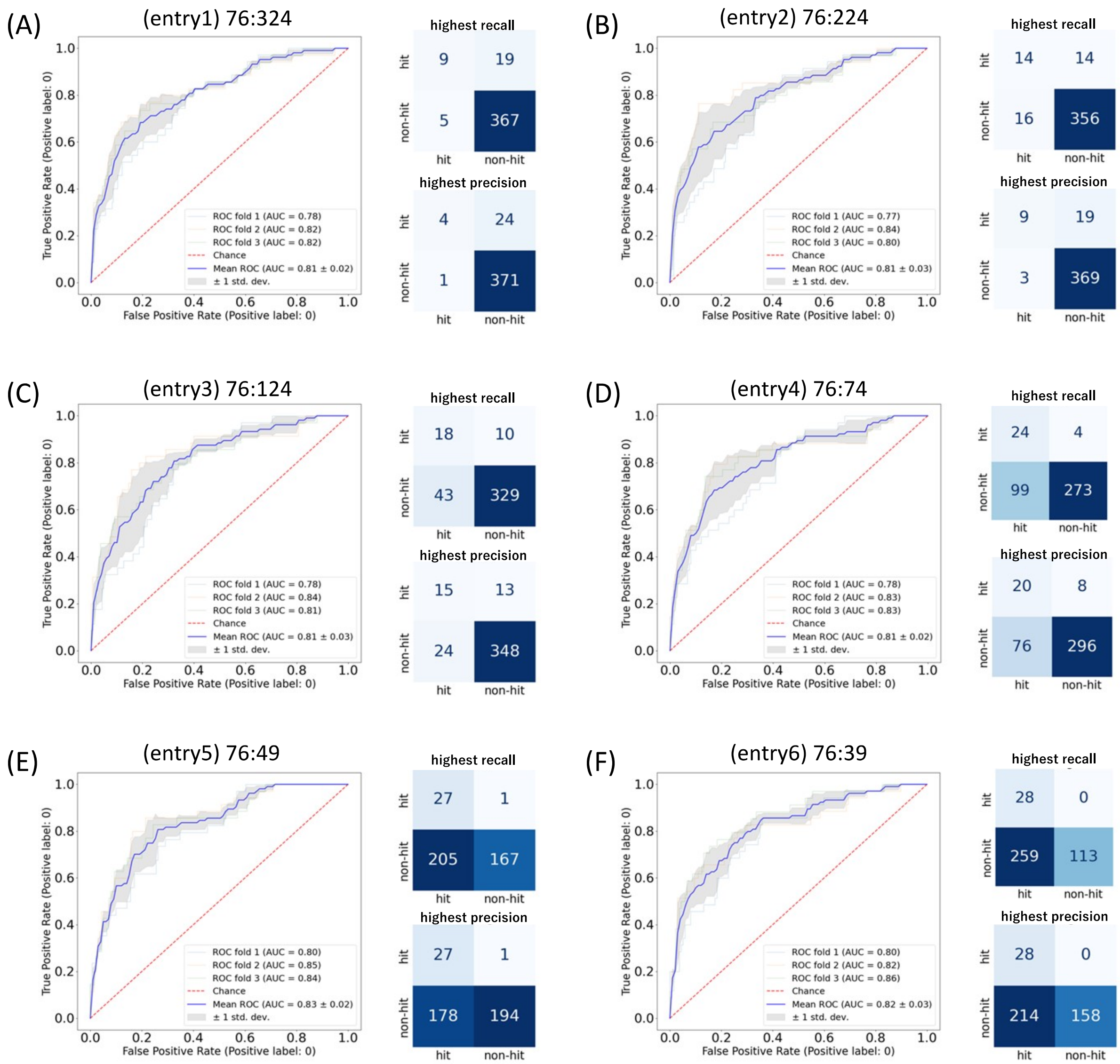


Fig. S4 (A-F) The binary classification results in 6 different hits/non-hits ratios (figure title on the left) in the training dataset with down-sampling, the same as Table S3. In each classification result, the ROC curve was shown on the left. The confusion matrix with the highest recall was shown in the upper right, and the highest precision in the bottom right.

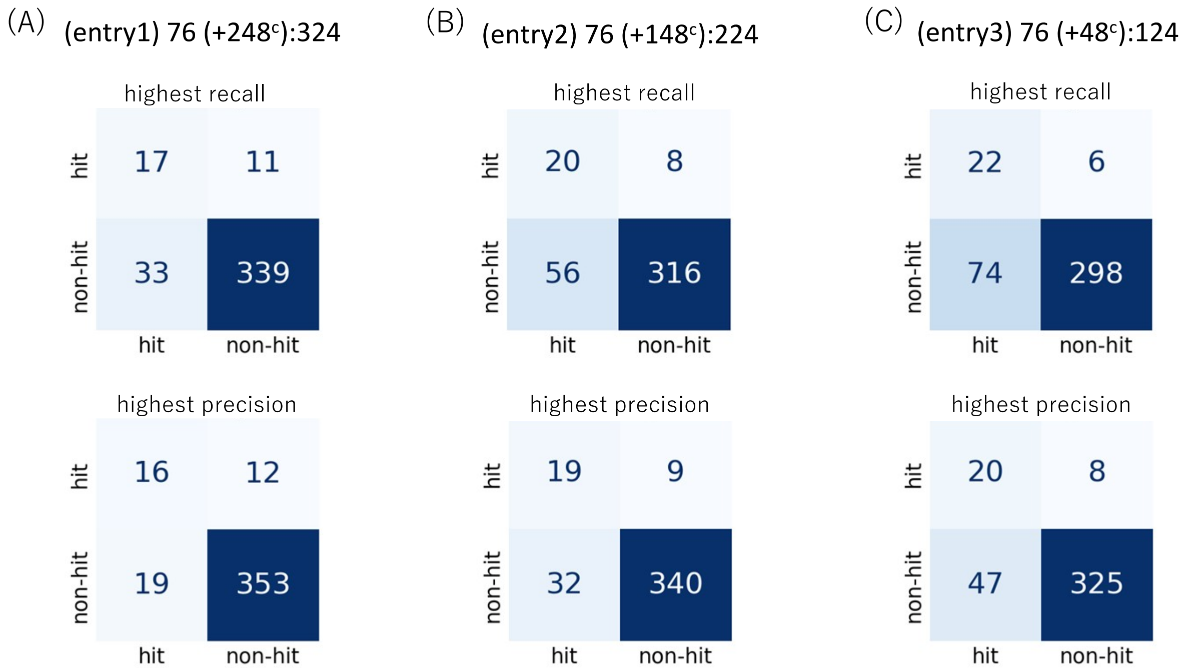


Fig. S5 (A-C) The binary classification results in 3 different hits/non-hits ratios in the training dataset with oversampling, (the "+" indicates oversampled hit counts. [c]Non-hits dropped in the whole dataset.) the same as Table S4. In each classification result, the confusion matrix with the highest recall was shown in the upper part, and the highest precision in the bottom.

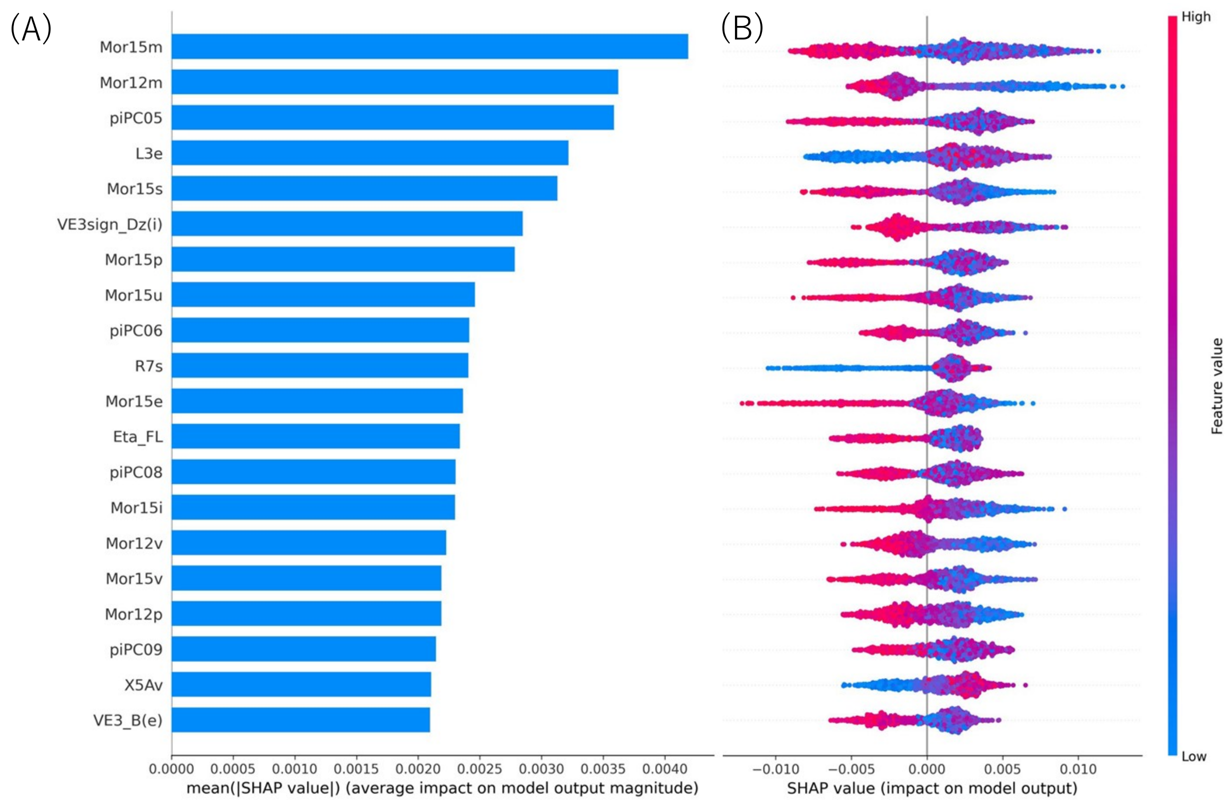


Fig. S6 The classification is the same as Table S3, entry 4. The top 20 features rank on (A) global SHAP value and (B) local beeswarm plot. Each instance (here is the molecule) of the dataset appears as its point for each variable. The vertical location is related to the feature's name, as shown in the y-axis. The red and blue colors correspond to each molecule's original descriptor values. Horizontal location indicates a negative or positive prediction with the SHAP value. And the vertical width will increase when data points have a high-density SHAP value.

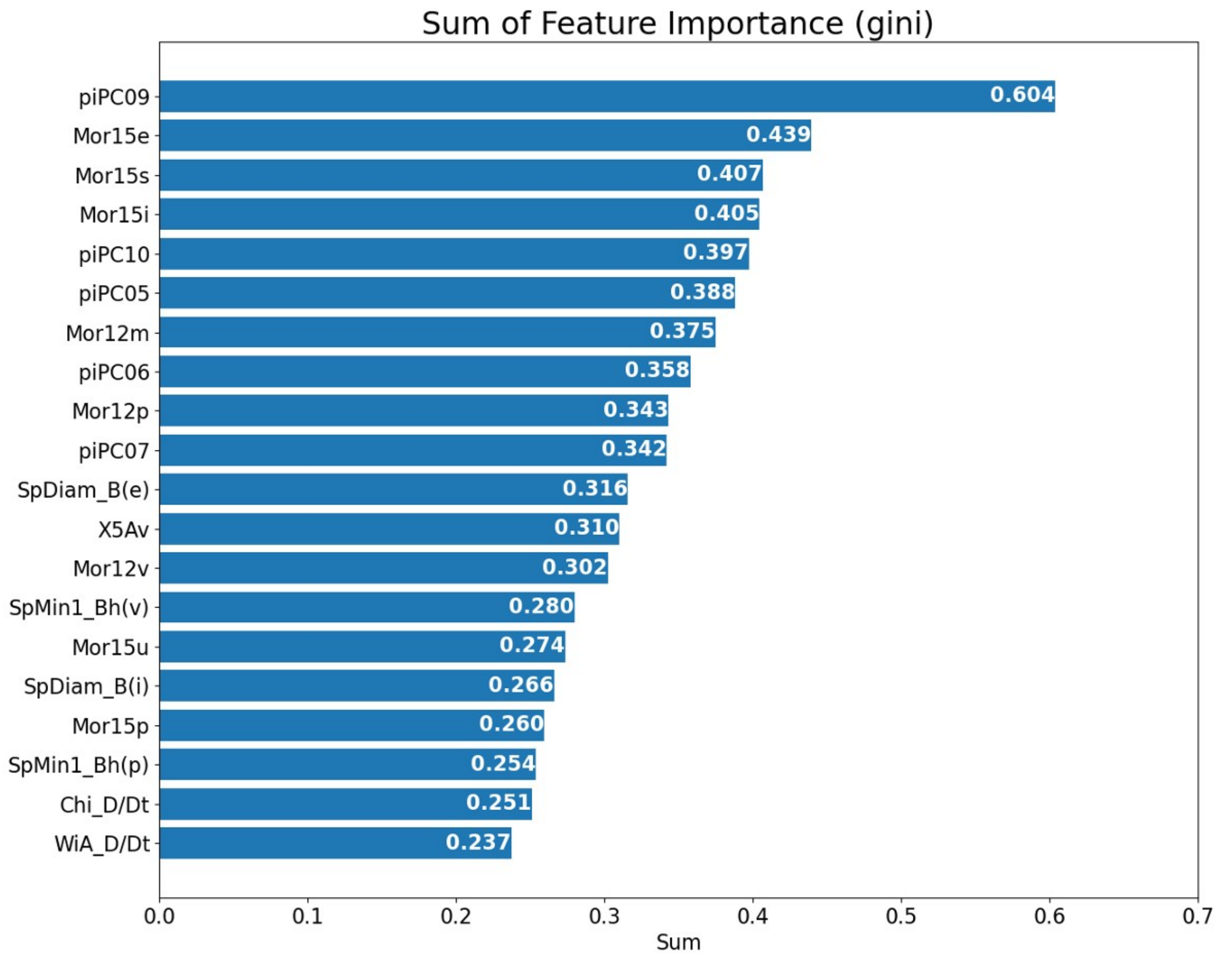


Fig. S7 The top 20 features ranked by feature importance based on the Gini index in RF algorithm. The Gini index of each feature is the sum of 100 independently experiments with the hits/non-hit ratio 76:74 shown in Fig. 3A.

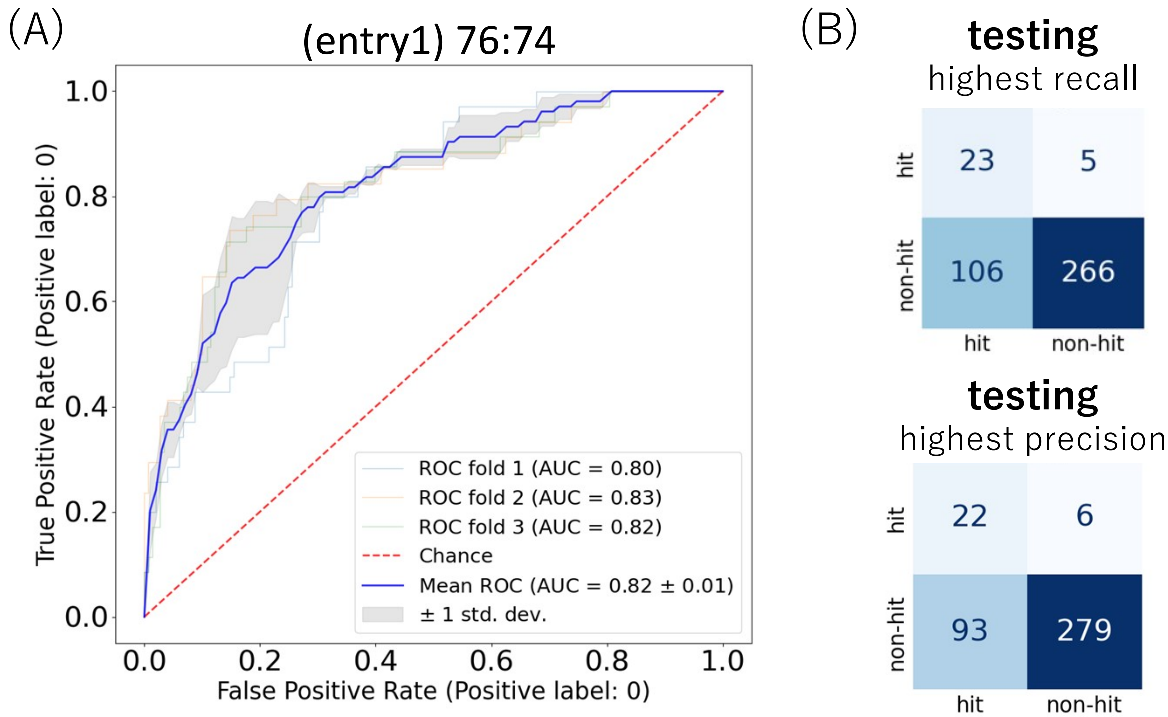


Fig. S8 The binary classification results where the top 10 descriptors in Fig. 4 were removed, the same as Table S5, entry 1. (A) receiver operating characteristic (ROC) curve with 3-fold cross-validation of the RF model performance. (B) The confusion matrix on the testing dataset. The confusion matrix has the highest recall (upper) and precision (bottom).

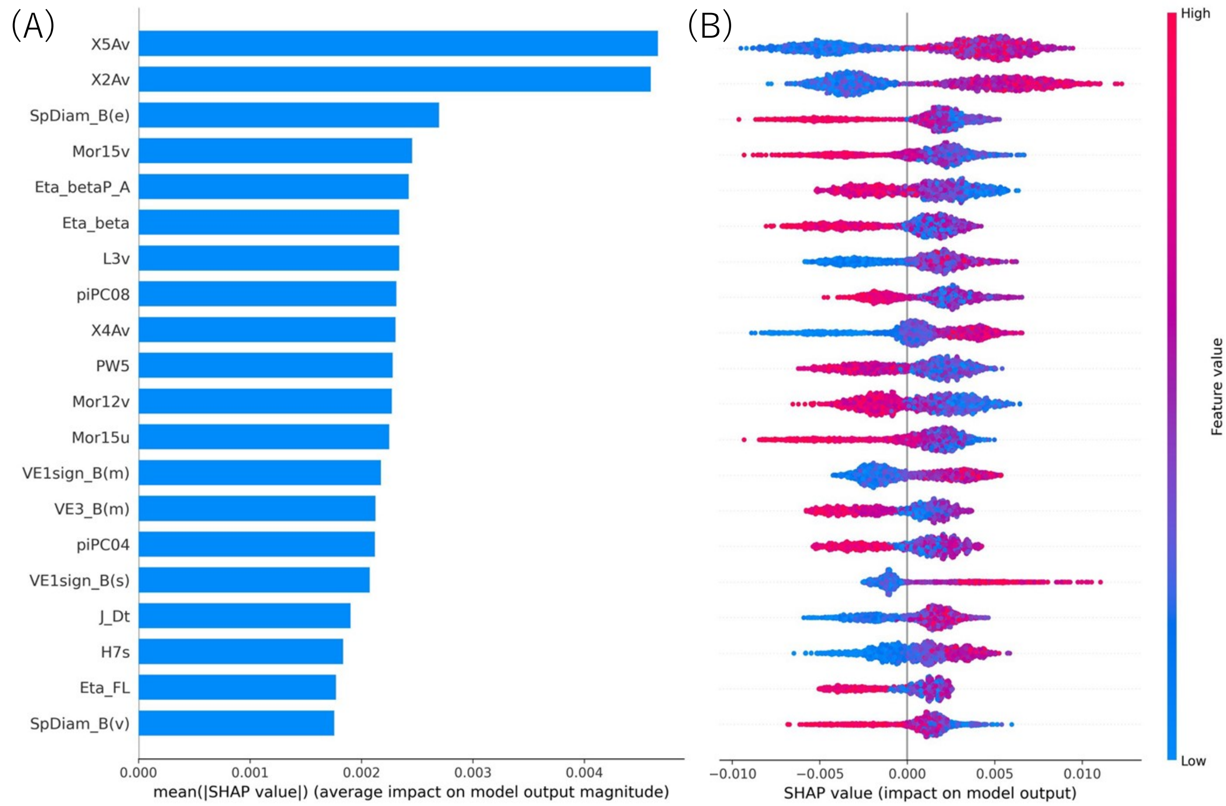
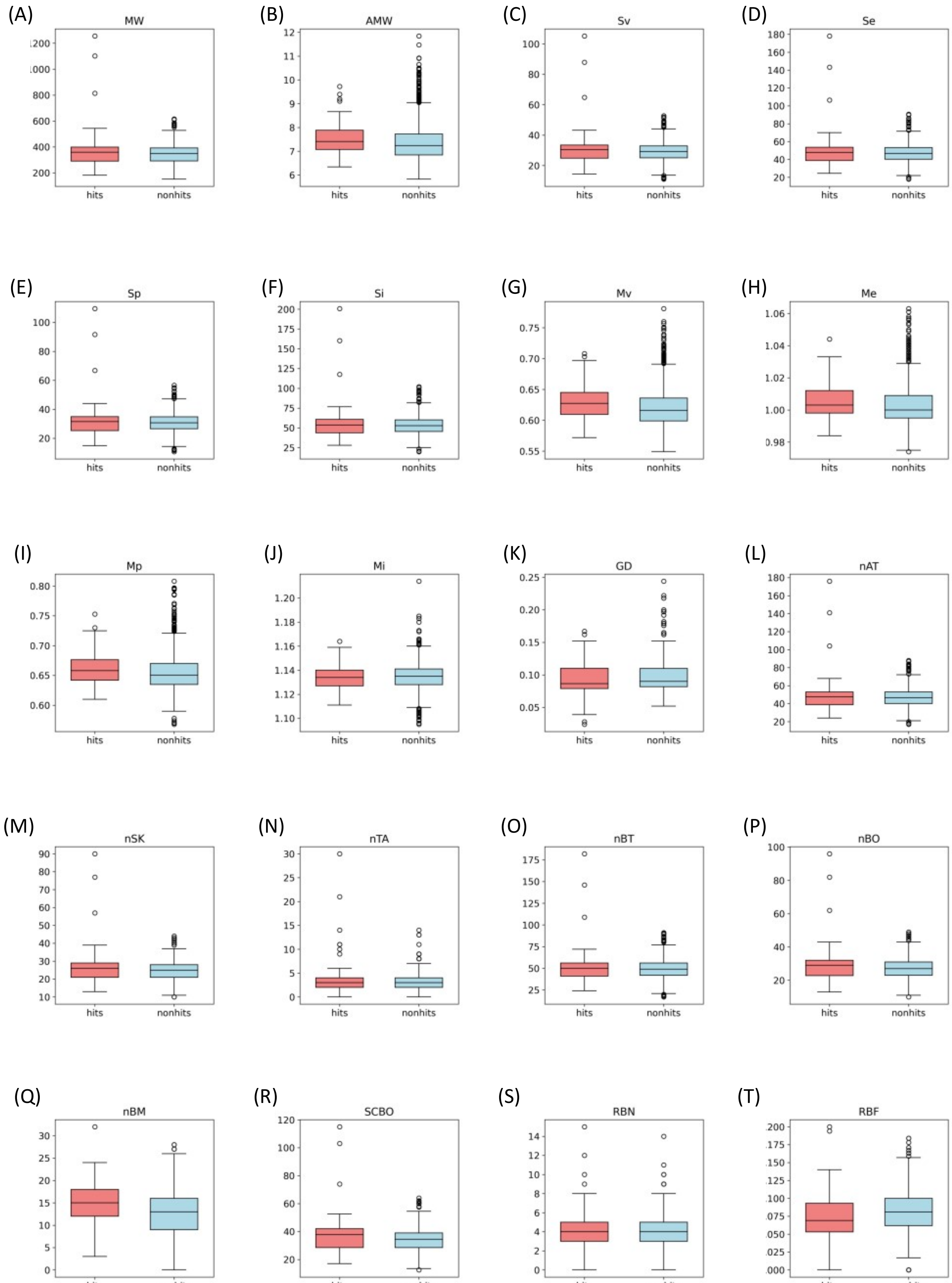


Fig. S9 The classification is the same as Table S5, entry 1, where the top 10 descriptors in Fig.4 were removed. The new top 20 features rank on (A) global SHAP value and (B) local beeswarm plot. Each instance (here is the molecule) of the dataset appears as its point for each variable. The vertical location is related to the feature's name, as shown in the y-axis. The red and blue colors correspond to each molecule's original descriptor values. Horizontal location indicates a negative or positive prediction with the SHAP value. And the vertical width will increase when data points have a high-density SHAP value.


(A) MW
(B) AMW
(C) Sv
(D) Se
(E) Sp
(F) Si
(G) Mv
(H) Me
(I) Mp
(J) Mi
(K) GD
(L) nAT
(M) nSK
(N) nTA
(O) nBT
(P) nBO
(Q) nBM
(R) SCBO
(S) RBN
(T) RBF
hits
nonhits

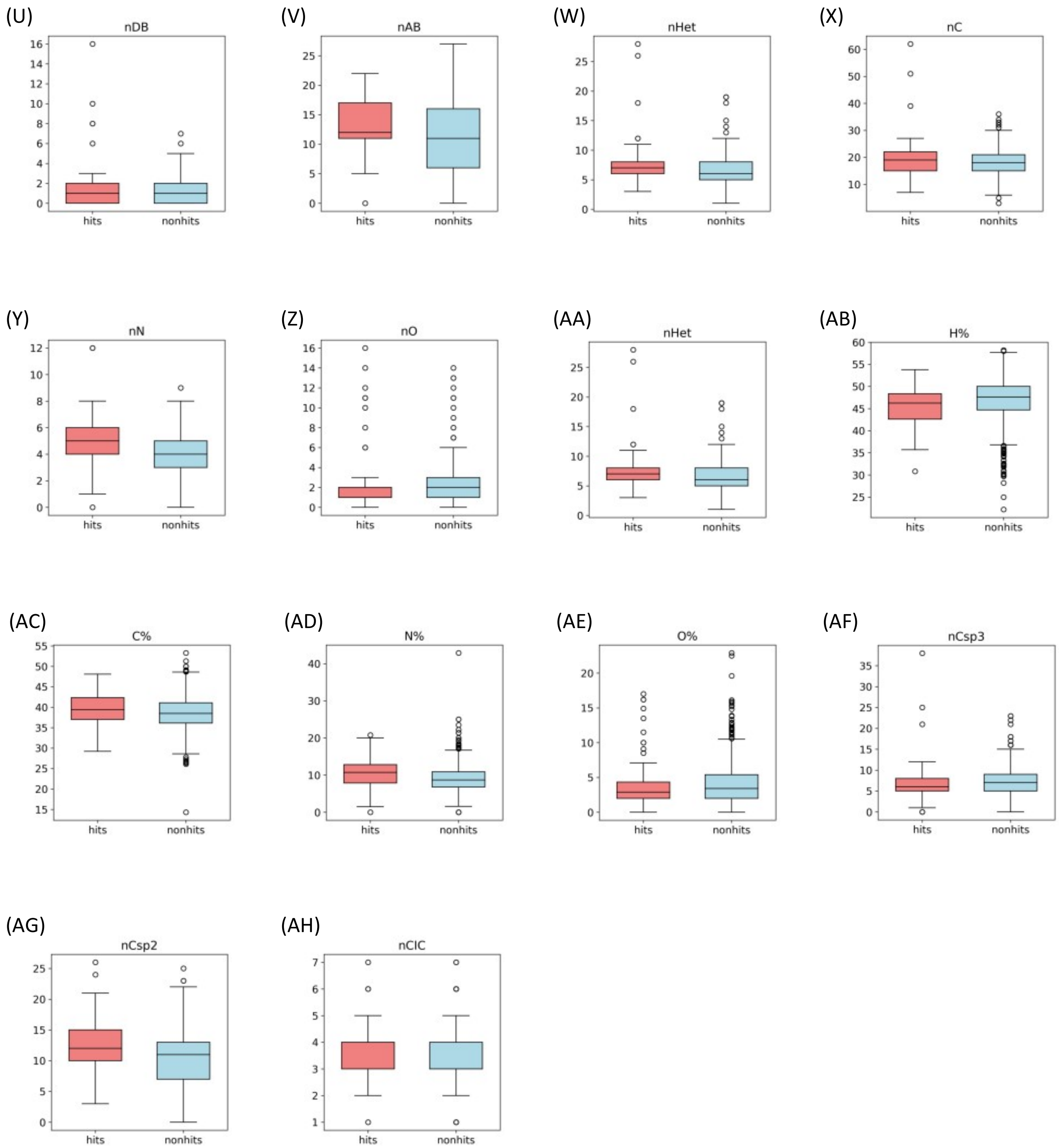


Fig. S10 Boxplot comparing constitutional descriptors of 104 identified hits (red) and 1896 non-hits (blue). (A) MW: molecular weight; (B) AMW: average molecular weight; (C) Sv: sum of atomic van der Waals volumes (scaled on Carbon atom); (D) Se: sum of atomic Sanderson electronegativities (scaled on Carbon atom); (E) Sp: sum of atomic polarizabilities (scaled on Carbon atom); (F) Si: sum of first ionization potentials (scaled on Carbon atom); (G) Mv: mean atomic van der Waals volume (scaled on Carbon atom); (H) Me: mean atomic Sanderson electronegativity (scaled on Carbon atom); (I) Mp: mean atomic polarizability (scaled on Carbon atom); (J) Mi: mean first ionization potential (scaled on Carbon atom); (K) GD: graph density; (L) nAT: number of atoms; (M) nSK: number of non-H atoms; (N) nTA: number of terminal atoms; (O) nBT: number of bonds; (P) nBO: number of non-H bonds; (Q) nBM: number of multiple bonds; (R) SCBO: sum of conventional bond orders (H-depleted); (S) RBN: number of rotatable bonds; (T) RBF: rotatable bond fraction; (U) nDB: number of double bonds; (V) nTB: number of triple bonds; (W) nAB: number of aromatic bonds; (X) nC: number of Carbon atoms; (Y) nN: number of Nitrogen atoms; (Z) nO: number of Oxygen atoms; (AA) nHet: number of heteroatoms; (AB) H%: percentage of H atoms; (AC) C%: percentage of C atoms; (AD) N%: percentage of N atoms; (AE) O%: percentage of O atoms; (AF) nCsp3: number of sp3 hybridized Carbon atoms; (AG) nCsp2: number of sp2 hybridized Carbon atoms; (AH) nCIC: number of rings (cyclomatic number)

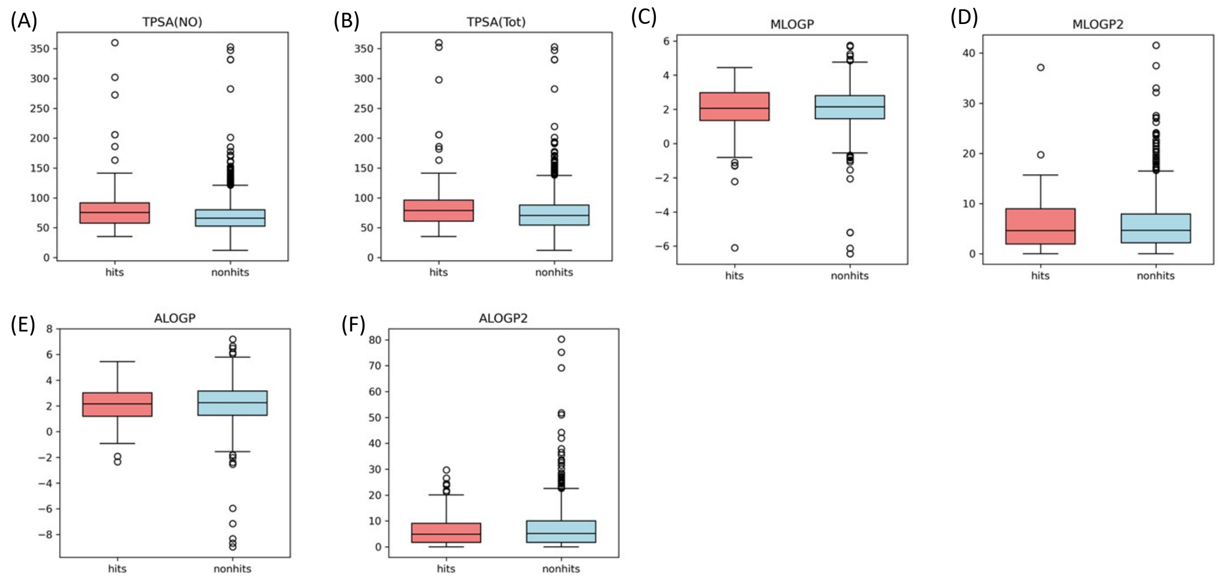


Fig. S11 Boxplot comparing other physico properties descriptors of 104 identified hits (red) and 1896 non-hits (blue). (A) TPSA (NO): topological polar surface area using N, O polar contributions; (B) TPSA (Tot): topological polar surface area using N, O, S, P polar contributions; (C) MLOGP: Moriguchi octanol-water partition coefficient (logP); (D) MLOGP2: squared Moriguchi octanol-water partition coefficient (logP^2); (E) ALOGP: Ghose-Crippen octanol-water partition coefficient (logP); (F) ALOGP2: squared Ghose-Crippen octanol-water partition coefficient (logP^2).

Supporting Tables

Table S1 The 30 descriptor categories[2] in this study.

Entry	category name	Number of descriptors
1	Constitutional	47
2	Ring descriptors	32
3	Topological indices	75
4	Walk and path counts	46
5	Connectivity indices	37
6	Information indices	50
7	2D matrix-based descriptors	607
8	2D autocorrelations	213
9	Burden eigenvalues	96
10	P-VSA-like descriptors	55
11	ETA indices	23
12	Edge adjacency indices	324
13	Geometrical descriptors	38
14	3D matrix-based descriptors	99
15	3D autocorrelations	80
16	RDF descriptors	210
17	3D-MoRSE descriptors	224
18	WHIM descriptors	114
19	GETAWAY descriptors	273
20	Randic molecular profiles	41
21	Functional groups count	154
22	Atom-centred fragments	115
23	Atom-type E-state indices	172
24	CATS 2D	150
25	2D Atom Pairs	1596
26	3D Atom Pairs	36
27	Charge descriptors	15
28	Molecular properties	20
29	Drug-like indices	28
30	CATS 3D	300

Table S2 The scores for the binary classification of the original imbalanced data.

	Recall	**Precision**	**F1**
Classification scores of hits	0.11	0.75	0.19

Table S3 The evaluation scores in applying down-sampling. (The scores in this table are for the hit class)

entry	average recall[a]	highest recall[b]	average precision[a]	highest precision[b]	average F1[a]	highest F1[b]	hits:non-hits in training	dropped non-hits[c]
1	0.17	0.32	0.51	0.8	0.26	0.43	76:324	1200
2	0.31	0.5	0.41	0.75	0.36	0.48	76:224	1300
3	0.54	0.64	0.28	0.38	0.37	0.45	76:124	1400
4	0.75	0.86	0.16	0.21	0.26	0.33	76:74	1450
5	0.89	0.96	0.11	0.13	0.20	0.23	76:49	1475
6	0.94	1	0.09	0.12	0.18	0.21	76:39	1485

[a]Average scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed. The five top and bottom values were excluded from the calculation.
[b]Highest scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed.
[c]Non-hits dropped in the whole dataset.

Table S4 The evaluation scores in applying over-sampling. (The scores in this table are for the hit class)

entry	avgerage recall[a]	highest recall[b]	average precision[a]	highest precision[b]	average F1[a]	highest F1[b]	hits:non-hits in training	dropped nonhits[d]
1	0.52	0.61	0.42	0.51	0.35	0.46	76 (+248[c]):324	1200
2	0.59	0.71	0.29	0.37	0.39	0.48	76 (+148[c]):224	1300
3	0.67	0.79	0.21	0.3	0.32	0.42	76 (+48[c]):124	1400

[a]Average scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed. The five top and bottom values were excluded from the calculation.
[b]Highest scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed.
[c]Hits oversampled in the training dataset.
[d]Non-hits dropped in the whole dataset.

Table S5 The prediction scores without the top 10 and 20 descriptors in Fig.4. (The scores in this table are for the hit class)

entry	removed features	average recall[a]	highest recall[b]	average precision[a]	highest precision[b]	average F1[a]	highest F1[b]	hits:non-hits in training	dropped nonhit[c]
1	top 10	0.74	0.82	0.15	0.19	0.25	0.31	76:74	1450
2	top 20	0.74	0.86	0.15	0.19	0.25	0.32	76:74	1450

[a]Average scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed. The five top and bottom values were excluded from the calculation.
[b]Highest scores obtained from 100 recorded prediction scores where the non-hits removed in each replicate experiment differed.
[c]Non-hits dropped in the whole dataset.

MATERIALS AND METHODS

1. Screening of a 2000-compound library using a Biacore T200

SPR measurement were carried out at 25 ℃ using a Biacore T200. 5' biotinylated DNA oligonucleotides (d(CAG)$_{40}$, d(GAA)$_{40}$ and dCont) were immobilized on a streptavidin (SA)-coated SPR sensor chip (Series S Sensor Chip SA, Cytiva) by biotin-SA interaction. A partial sequence of the Htt (huntingtin) intron 1 sequences was selected as the control dCont (AAG TGT TGA CAT TTT TAT TTT ATT TTG TTT TGT TTT GTT TTT TTT GAG ACA GTT CTT GCT CTA TCA GCC AGG CTG GAG TGC ACT AGT GTG ATC TTG GCT CAC TGC AAC CTC TGC CTC TTG). After washing the surface of a SA sensor chip with 50 mM NaOH-1 M NaCl solution, each DNA solution (0.25 µM) in 10 mM HEPES-500 mM NaCl buffer was flowed onto the SA chip at 5 µL/min at 25 ℃. The immobilized amount of DNA oligonucleotides was summarized in the table below. A 2000-compound library, which contains 201 compounds predicted to bind to CAG repeat DNAs, 1472 compounds predicted to bind to DNAs/RNAs, and 327 compounds with diverse structures, was screened against the DNA-immobilized sensor chip. The library compounds were diluted to 50 µM in the running buffer (1×HBS-EP+ buffer containing 5% (v/v) DMSO) and flowed onto the DNA-immobilized sensor surface (A contact time of 15 s with a 15-s dissociation time). Naphthyridine-Azaquinolone (**NA**) was used as a positive control, and a non-binder (Np1-N012) from our in-house chemical library was used as a negative control. Solvent correction was performed with 8-point samples at from 2–4% DMSO. Regeneration of the sensor surface was achieved by flowing the regeneration solution (1.2 mM NaOH, 0.2 M NaCl, and 0.1 mM EDTA) for 600 sec every 40 compounds. Injections of the DMSO containing solutions for solvent correction, the control samples, and the regeneration solution were included every 40 cycles in this order.

The sensorgrams obtained were analyzed using the BIAevaluation software. The row sensorgrams were solvent-corrected and referenced by subtracting the signals obtained from the reference cell (Fc1). The binding response values were calculated 4 seconds before the end of each injection and averaged over a 4-sec window. The binding response was normalized to those of control samples and molecular weight. From the screening process, compounds that showed a relative binding response unit (RU) value of >20 against d(CAG)$_{40}$ were selected, resulting in 104 hits.

Immobilized DNA and its amount on a SPR sensor surface

Flow cell	DNA	Immobilized amount (in RU)
Fc1	none (reference cell)	20.2
Fc2	Biotin-TEG-d(CAG)$_{40}$	840.8
Fc3	Biotin-TEG-d(GAA)$_{40}$	872.9
Fc4	Biotin-TEG-dCont	833.6

2. Machine learning

Molecular descriptors for the 2000 compounds were generated by transforming the chemical structures to numerical values using Dragon 7.0 software, providing 5270 descriptors classified into 30 categories (Table S1) including atom types, functional groups, geometrical descriptions, and properties. The machine learning predictive analysis was conducted utilizing the PyCharm professional (version 2021.3.2) Integrated Development Environment (IDE). PyCharm was set up with the relevant Python interpreter. A new project was then initiated within this environment. Essential libraries (shown in the requirements below) were integrated by employing PyCharm's package manager with the compatible version. Subsequently, the algorithmic code was developed, beginning with the importation and preprocessing of the dataset via pandas and numpy. The scikit-learn library facilitated the implementation of a random forest and XGBoost classifiers. The dataset of 2000 compounds were partitioned into training and testing subsets (8:2 ratio), then, with the model being trained on the former and its performance was evaluated against the latter. Model parameters (such as the tree number: n_estimators, and size of each tree: max_depth were fine-tuned to enhance predictive accuracy, but the grid search process is not involved in the provided code.)

2000 small molecule dataset

dataset/Dragon_2000cpds.csv

Scripts

RFcls_main_latest.py (Classification and evaluation.)
categorical_index.py (A required module for over-sampling)

Dataset and code are available in GitHub:
https://github.com/chen26sanken/RF_Generation-of-a-focused-molecule-library-by-machine-learning-targeting-CAG-repeat-DNA.git.

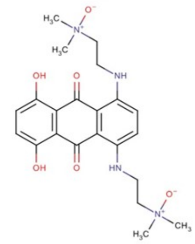
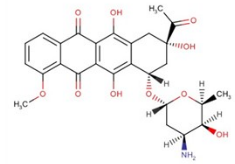
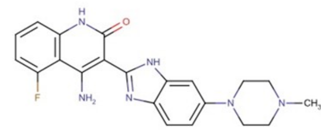
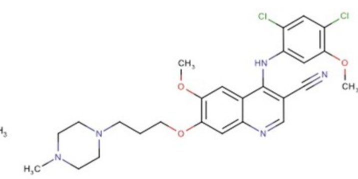
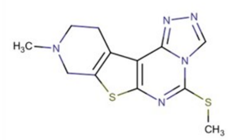
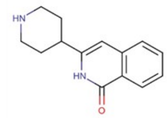
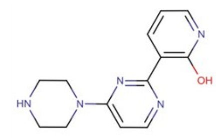
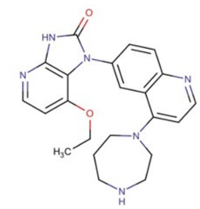
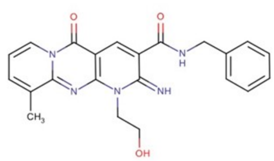
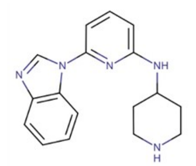
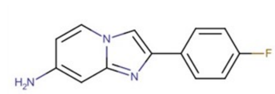
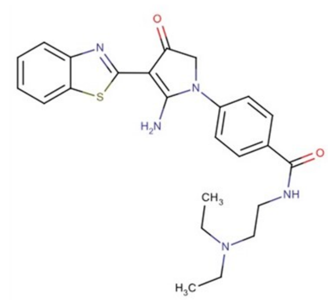
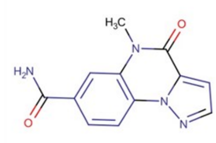
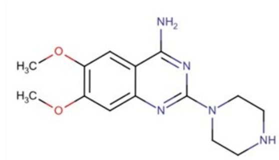
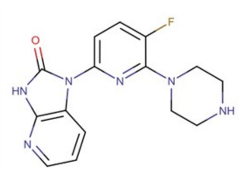
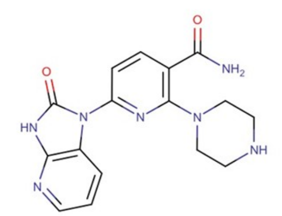
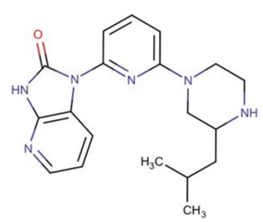
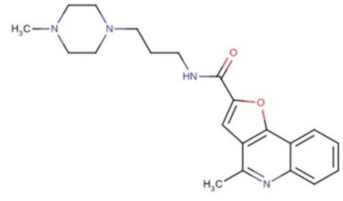
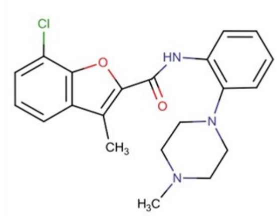
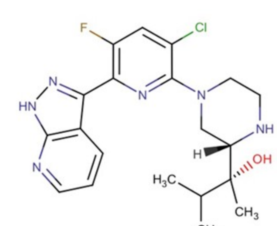
Requirements for creating the virtual environment:

matplotlib~=3.5.1
scikit-learn~=1.1.1
numpy~=1.21.6
pandas~=1.4.3
seaborn~=0.11.2
shap~=0.41.0
ipython~=8.5.0
xgboost~=1.6.2
tqdm~=4.64.1
~=: compatible version

References

1. K. Nakatani, S. Hagihara, Y. Goto, A. Kobori, M. Hagihara, G. Hayashi, M. Kyo, M. Nomura, M. Mishima and C. Kojima, *Nat. Chem. Biol.*, 2005, **1**, 39–43.
2. I. Tetko, J. Gasteiger, R. Todeschini, A. Mauri, D. Livingstone, P. Ertl, V. Palyulin, E. Radchenko, N. Zefirov, A. Makarenko, V. Tanchuk and V. Prokopenko, *J. Comput. Aid. Mol. Des.,* 2005, **19**, 453–463.

Appendix A

Appendix B

The categorical features, which used SMOTE-NC function to oversample in this study.

nAT	nSK	nTA	nBT	nBO
nBM	RBN	nDB	nTB	nAB
nH	nC	nN	nO	nP
nS	nF	nCL	nBR	nI
nB	nHM	nHet	nX	nCsp3
nCsp2	nCsp	nStructures	totalcharge	nCIC
nCIR	TRS	Rperim	Rbrid	NRS
nR03	nR04	nR05	nR06	nR07
nR08	nR09	nR10	nR11	nR12
nBnz	ZM1	ZM1V	ZM2	ZM2V
Qindex	BBI	Ram	Pol	ECC
UNIP	CENT	VAR	SMTI	SMTIV
GMTI	GMTIV	CSI	Wap	Psi_e_t
BAC	Ges	Wi_D	SpPos_L	SpMax_X
Wi_Dt	P_VSA_v_4	Mor01u	H0u	HATSu
nCp	nCs	nCt	nCq	nCrs
nCrt	nCrq	nCar	nCbH	nCb-
nCconj	nR=Cp	nR=Cs	nR=Ct	n=C=
nR#CH/X	nR#C-	nROCN	nArOCN	nRNCO
nArNCO	nRSCN	nArSCN	nRNCS	nArNCS
nRCOOH	nArCOOH	nRCOOR	nArCOOR	nRCONH2
nArCONH2	nRCONHR	nArCONHR	nRCONR2	nArCONR2
nROCON	nArOCON	nRCOX	nArCOX	nRCSOH
nArCSOH	nRCSSH	nArCSSH	nRCOSR	nArCOSR
nRCSSR	nArCSSR	nRCHO	nArCHO	nRCO
nArCO	nCONN	nC=O(O)2	nN=C-N<	nC(=N)N2
nRC=N	nArC=N	nRCNO	nArCNO	nRNH2
nArNH2	nRNHR	nArNHR	nRNR2	nArNR2
nN-N	nN=N	nRCN	nArCN	nN+
nNq	nRNHO	nArNHO	nRNNOx	nArNNOx
nRNO	nArNO	nRNO2	nArNO2	nN(CO)2
nC=N-N<	nROH	nArOH	nOHp	nOHs
nOHt	nROR	nArOR	nROX	nArOX
nO(C=O)2	nH2O	nSH	nC=S	nRSR
nRSSR	nSO	nS(=O)2	nSOH	nSOOH
nSO2OH	nSO3OH	nSO2	nSO3	nSO4
nSO2N	nPO3	nPO4	nPR3	nP(=O)O2R
nP(=O)R3/nPR5	nCH2RX	nCHR2X	nCR3X	nR=CHX
nR=CRX	nR#CX	nCHRX2	nCR2X2	nR=CX2
nCRX3	nArX	nCXr	nCXr=	nCconjX
nAziridines	nOxiranes	nThiranes	nAzetidines	nOxetanes

nThioethanes	nBeta-Lactams	nPyrrolidines	nOxolanes	ntH-Thiophenes
nPyrroles	nPyrazoles	nImidazoles	nFuranes	nThiophenes
nOxazoles	nIsoxazoles	nThiazoles	nIsothiazoles	nTriazoles
nPyridines	nPyridazines	nPyrimidines	nPyrazines	n135-Triazines
n124-Triazines	nHDon	nHAcc	nHBonds	C-001
C-002	C-003	C-004	C-005	C-006
C-007	C-008	C-009	C-010	C-011
C-012	C-013	C-014	C-015	C-016
C-017	C-018	C-019	C-020	C-021
C-022	C-023	C-024	C-025	C-026
C-027	C-028	C-029	C-030	C-031
C-032	C-033	C-034	C-035	C-036
C-037	C-038	C-039	C-040	C-041
C-042	C-043	C-044	H-046	H-047
H-048	H-049	H-050	H-051	H-052
H-053	H-054	H-055	O-056	O-057
O-058	O-059	O-060	O-061	O-062
O-063	Se-064	Se-065	N-066	N-067
N-068	N-069	N-070	N-071	N-072
N-073	N-074	N-075	N-076	N-077
N-078	N-079	F-081	F-082	F-083
F-084	F-085	Cl-086	Cl-087	Cl-088
Cl-089	Cl-090	Br-091	Br-092	Br-093
Br-094	Br-095	I-096	I-097	I-098
I-099	I-100	F-101	Cl-102	Br-103
I-104	S-106	S-107	S-108	S-109
S-110	Si-111	B-112	P-115	P-116
P-117	P-118	P-119	P-120	SddC
SsNH3+	SssNH2+	SdNH2+	SsssNH+	SsPH2
SssPH	SsssP	SddsP	SsssssP	SsSH
SsssssS	SsI	SsLi	SssBe	SssssBe-
SsBH2	SssBH	SsssB	SssssB-	SsGeH3
SssGeH2	SsssGeH	SssssGe	SsAsH2	SssAsH
SsssAs	SsssssAs	SdsssAs	SddsAs	SsSeH
SdSe	SssSe	SaaSe	SdssSe	SssssssSe
SddssSe	SsSnH3	SssSnH2	SsssSnH	SssssSn
SsPbH3	SssPbH2	SsssPbH	SssssPb	SsSiH3
SssSiH2	SsssSiH	SssssSi	NsCH3	NdCH2
NssCH2	NtCH	NdsCH	NaaCH	NsssCH
NddC	NtsC	NdssC	NaasC	NaaaC
NssssC	NsNH2	NssNH	NdNH	NsssN
NdsN	NaaN	NtN	NsNH3+	NssNH2+
NdNH2+	NsssNH+	NssssN+	NddsN	NaadN
NaasN	NaaNH	NsOH	NdO	NssO

NaaO	NsPH2	NssPH	NsssP	NdsssP
NddsP	NsssssP	NsSH	NdS	NssS
NaaS	NdssS	NddssS	NssssssS	NsF
NsCl	NsBr	NsI	NsLi	NssBe
NssssBe-	NsBH2	NssBH	NsssB	NssssB-
NsGeH3	NssGeH2	NsssGeH	NssssGe	NsAsH2
NssAsH	NsssAs	NsssssAs	NdsssAs	NddsAs
NsSeH	NdSe	NssSe	NaaSe	NdssSe
NssssssSe	NddssSe	NsSnH3	NssSnH2	NsssSnH
NssssSn	NsPbH3	NssPbH2	NsssPbH	NssssPb
NsSiH3	NssSiH2	NsssSiH	NssssSi	CATS2D_00_DD
CATS2D_01_DD	CATS2D_02_DD	CATS2D_03_DD	CATS2D_04_DD	CATS2D_05_DD
CATS2D_06_DD	CATS2D_07_DD	CATS2D_08_DD	CATS2D_09_DD	CATS2D_00_DA
CATS2D_01_DA	CATS2D_02_DA	CATS2D_03_DA	CATS2D_04_DA	CATS2D_05_DA
CATS2D_06_DA	CATS2D_07_DA	CATS2D_08_DA	CATS2D_09_DA	CATS2D_00_DP
CATS2D_01_DP	CATS2D_02_DP	CATS2D_03_DP	CATS2D_04_DP	CATS2D_05_DP
CATS2D_06_DP	CATS2D_07_DP	CATS2D_08_DP	CATS2D_09_DP	CATS2D_00_DN
CATS2D_01_DN	CATS2D_02_DN	CATS2D_03_DN	CATS2D_04_DN	CATS2D_05_DN
CATS2D_06_DN	CATS2D_07_DN	CATS2D_08_DN	CATS2D_09_DN	CATS2D_00_DL
CATS2D_01_DL	CATS2D_02_DL	CATS2D_03_DL	CATS2D_04_DL	CATS2D_05_DL
CATS2D_06_DL	CATS2D_07_DL	CATS2D_08_DL	CATS2D_09_DL	CATS2D_00_AA
CATS2D_01_AA	CATS2D_02_AA	CATS2D_03_AA	CATS2D_04_AA	CATS2D_05_AA
CATS2D_06_AA	CATS2D_07_AA	CATS2D_08_AA	CATS2D_09_AA	CATS2D_00_AP
CATS2D_01_AP	CATS2D_02_AP	CATS2D_03_AP	CATS2D_04_AP	CATS2D_05_AP
CATS2D_06_AP	CATS2D_07_AP	CATS2D_08_AP	CATS2D_09_AP	CATS2D_00_AN
CATS2D_01_AN	CATS2D_02_AN	CATS2D_03_AN	CATS2D_04_AN	CATS2D_05_AN
CATS2D_06_AN	CATS2D_07_AN	CATS2D_08_AN	CATS2D_09_AN	CATS2D_00_AL
CATS2D_01_AL	CATS2D_02_AL	CATS2D_03_AL	CATS2D_04_AL	CATS2D_05_AL
CATS2D_06_AL	CATS2D_07_AL	CATS2D_08_AL	CATS2D_09_AL	CATS2D_00_PP
CATS2D_01_PP	CATS2D_02_PP	CATS2D_03_PP	CATS2D_04_PP	CATS2D_05_PP
CATS2D_06_PP	CATS2D_07_PP	CATS2D_08_PP	CATS2D_09_PP	CATS2D_00_PN
CATS2D_01_PN	CATS2D_02_PN	CATS2D_03_PN	CATS2D_04_PN	CATS2D_05_PN
CATS2D_06_PN	CATS2D_07_PN	CATS2D_08_PN	CATS2D_09_PN	CATS2D_00_PL
CATS2D_01_PL	CATS2D_02_PL	CATS2D_03_PL	CATS2D_04_PL	CATS2D_05_PL
CATS2D_06_PL	CATS2D_07_PL	CATS2D_08_PL	CATS2D_09_PL	CATS2D_00_NN
CATS2D_01_NN	CATS2D_02_NN	CATS2D_03_NN	CATS2D_04_NN	CATS2D_05_NN
CATS2D_06_NN	CATS2D_07_NN	CATS2D_08_NN	CATS2D_09_NN	CATS2D_00_NL
CATS2D_01_NL	CATS2D_02_NL	CATS2D_03_NL	CATS2D_04_NL	CATS2D_05_NL
CATS2D_06_NL	CATS2D_07_NL	CATS2D_08_NL	CATS2D_09_NL	CATS2D_00_LL
CATS2D_01_LL	CATS2D_02_LL	CATS2D_03_LL	CATS2D_04_LL	CATS2D_05_LL
CATS2D_06_LL	CATS2D_07_LL	CATS2D_08_LL	CATS2D_09_LL	T(N..N)
T(N..O)	T(N..S)	T(N..P)	T(N..F)	T(N..Cl)
T(N..Br)	T(N..I)	T(O..O)	T(O..S)	T(O..P)
T(O..F)	T(O..Cl)	T(O..Br)	T(O..I)	T(S..S)

T(S..P)	T(S..F)	T(S..Cl)	T(S..Br)	T(S..I)
T(P..P)	T(P..F)	T(P..Cl)	T(P..Br)	T(P..I)
T(F..F)	T(F..Cl)	T(F..Br)	T(F..I)	T(Cl..Cl)
T(Cl..Br)	T(Cl..I)	T(Br..Br)	T(Br..I)	T(I..I)
B01[C-C]	B01[C-N]	B01[C-O]	B01[C-S]	B01[C-P]
B01[C-F]	B01[C-Cl]	B01[C-Br]	B01[C-I]	B01[C-B]
B01[C-Si]	B01[C-X]	B01[N-N]	B01[N-O]	B01[N-S]
B01[N-P]	B01[N-F]	B01[N-Cl]	B01[N-Br]	B01[N-I]
B01[N-B]	B01[N-Si]	B01[N-X]	B01[O-O]	B01[O-S]
B01[O-P]	B01[O-F]	B01[O-Cl]	B01[O-Br]	B01[O-I]
B01[O-B]	B01[O-Si]	B01[O-X]	B01[S-S]	B01[S-P]
B01[S-F]	B01[S-Cl]	B01[S-Br]	B01[S-I]	B01[S-B]
B01[S-Si]	B01[S-X]	B01[P-P]	B01[P-F]	B01[P-Cl]
B01[P-Br]	B01[P-I]	B01[P-B]	B01[P-Si]	B01[P-X]
B01[F-F]	B01[F-Cl]	B01[F-Br]	B01[F-I]	B01[F-B]
B01[F-Si]	B01[F-X]	B01[Cl-Cl]	B01[Cl-Br]	B01[Cl-I]
B01[Cl-B]	B01[Cl-Si]	B01[Cl-X]	B01[Br-Br]	B01[Br-I]
B01[Br-B]	B01[Br-Si]	B01[Br-X]	B01[I-I]	B01[I-B]
B01[I-Si]	B01[I-X]	B01[B-B]	B01[B-Si]	B01[B-X]
B01[Si-Si]	B01[Si-X]	B01[X-X]	B02[C-C]	B02[C-N]
B02[C-O]	B02[C-S]	B02[C-P]	B02[C-F]	B02[C-Cl]
B02[C-Br]	B02[C-I]	B02[C-B]	B02[C-Si]	B02[C-X]
B02[N-N]	B02[N-O]	B02[N-S]	B02[N-P]	B02[N-F]
B02[N-Cl]	B02[N-Br]	B02[N-I]	B02[N-B]	B02[N-Si]
B02[N-X]	B02[O-O]	B02[O-S]	B02[O-P]	B02[O-F]
B02[O-Cl]	B02[O-Br]	B02[O-I]	B02[O-B]	B02[O-Si]
B02[O-X]	B02[S-S]	B02[S-P]	B02[S-F]	B02[S-Cl]
B02[S-Br]	B02[S-I]	B02[S-B]	B02[S-Si]	B02[S-X]
B02[P-P]	B02[P-F]	B02[P-Cl]	B02[P-Br]	B02[P-I]
B02[P-B]	B02[P-Si]	B02[P-X]	B02[F-F]	B02[F-Cl]
B02[F-Br]	B02[F-I]	B02[F-B]	B02[F-Si]	B02[F-X]
B02[Cl-Cl]	B02[Cl-Br]	B02[Cl-I]	B02[Cl-B]	B02[Cl-Si]
B02[Cl-X]	B02[Br-Br]	B02[Br-I]	B02[Br-B]	B02[Br-Si]
B02[Br-X]	B02[I-I]	B02[I-B]	B02[I-Si]	B02[I-X]
B02[B-B]	B02[B-Si]	B02[B-X]	B02[Si-Si]	B02[Si-X]
B02[X-X]	B03[C-C]	B03[C-N]	B03[C-O]	B03[C-S]
B03[C-P]	B03[C-F]	B03[C-Cl]	B03[C-Br]	B03[C-I]
B03[C-B]	B03[C-Si]	B03[C-X]	B03[N-N]	B03[N-O]
B03[N-S]	B03[N-P]	B03[N-F]	B03[N-Cl]	B03[N-Br]
B03[N-I]	B03[N-B]	B03[N-Si]	B03[N-X]	B03[O-O]
B03[O-S]	B03[O-P]	B03[O-F]	B03[O-Cl]	B03[O-Br]
B03[O-I]	B03[O-B]	B03[O-Si]	B03[O-X]	B03[S-S]
B03[S-P]	B03[S-F]	B03[S-Cl]	B03[S-Br]	B03[S-I]
B03[S-B]	B03[S-Si]	B03[S-X]	B03[P-P]	B03[P-F]

B03[P-Cl]	B03[P-Br]	B03[P-I]	B03[P-B]	B03[P-Si]
B03[P-X]	B03[F-F]	B03[F-Cl]	B03[F-Br]	B03[F-I]
B03[F-B]	B03[F-Si]	B03[F-X]	B03[Cl-Cl]	B03[Cl-Br]
B03[Cl-I]	B03[Cl-B]	B03[Cl-Si]	B03[Cl-X]	B03[Br-Br]
B03[Br-I]	B03[Br-B]	B03[Br-Si]	B03[Br-X]	B03[I-I]
B03[I-B]	B03[I-Si]	B03[I-X]	B03[B-B]	B03[B-Si]
B03[B-X]	B03[Si-Si]	B03[Si-X]	B03[X-X]	B04[C-C]
B04[C-N]	B04[C-O]	B04[C-S]	B04[C-P]	B04[C-F]
B04[C-Cl]	B04[C-Br]	B04[C-I]	B04[C-B]	B04[C-Si]
B04[C-X]	B04[N-N]	B04[N-O]	B04[N-S]	B04[N-P]
B04[N-F]	B04[N-Cl]	B04[N-Br]	B04[N-I]	B04[N-B]
B04[N-Si]	B04[N-X]	B04[O-O]	B04[O-S]	B04[O-P]
B04[O-F]	B04[O-Cl]	B04[O-Br]	B04[O-I]	B04[O-B]
B04[O-Si]	B04[O-X]	B04[S-S]	B04[S-P]	B04[S-F]
B04[S-Cl]	B04[S-Br]	B04[S-I]	B04[S-B]	B04[S-Si]
B04[S-X]	B04[P-P]	B04[P-F]	B04[P-Cl]	B04[P-Br]
B04[P-I]	B04[P-B]	B04[P-Si]	B04[P-X]	B04[F-F]
B04[F-Cl]	B04[F-Br]	B04[F-I]	B04[F-B]	B04[F-Si]
B04[F-X]	B04[Cl-Cl]	B04[Cl-Br]	B04[Cl-I]	B04[Cl-B]
B04[Cl-Si]	B04[Cl-X]	B04[Br-Br]	B04[Br-I]	B04[Br-B]
B04[Br-Si]	B04[Br-X]	B04[I-I]	B04[I-B]	B04[I-Si]
B04[I-X]	B04[B-B]	B04[B-Si]	B04[B-X]	B04[Si-Si]
B04[Si-X]	B04[X-X]	B05[C-C]	B05[C-N]	B05[C-O]
B05[C-S]	B05[C-P]	B05[C-F]	B05[C-Cl]	B05[C-Br]
B05[C-I]	B05[C-B]	B05[C-Si]	B05[C-X]	B05[N-N]
B05[N-O]	B05[N-S]	B05[N-P]	B05[N-F]	B05[N-Cl]
B05[N-Br]	B05[N-I]	B05[N-B]	B05[N-Si]	B05[N-X]
B05[O-O]	B05[O-S]	B05[O-P]	B05[O-F]	B05[O-Cl]
B05[O-Br]	B05[O-I]	B05[O-B]	B05[O-Si]	B05[O-X]
B05[S-S]	B05[S-P]	B05[S-F]	B05[S-Cl]	B05[S-Br]
B05[S-I]	B05[S-B]	B05[S-Si]	B05[S-X]	B05[P-P]
B05[P-F]	B05[P-Cl]	B05[P-Br]	B05[P-I]	B05[P-B]
B05[P-Si]	B05[P-X]	B05[F-F]	B05[F-Cl]	B05[F-Br]
B05[F-I]	B05[F-B]	B05[F-Si]	B05[F-X]	B05[Cl-Cl]
B05[Cl-Br]	B05[Cl-I]	B05[Cl-B]	B05[Cl-Si]	B05[Cl-X]
B05[Br-Br]	B05[Br-I]	B05[Br-B]	B05[Br-Si]	B05[Br-X]
B05[I-I]	B05[I-B]	B05[I-Si]	B05[I-X]	B05[B-B]
B05[B-Si]	B05[B-X]	B05[Si-Si]	B05[Si-X]	B05[X-X]
B06[C-C]	B06[C-N]	B06[C-O]	B06[C-S]	B06[C-P]
B06[C-F]	B06[C-Cl]	B06[C-Br]	B06[C-I]	B06[C-B]
B06[C-Si]	B06[C-X]	B06[N-N]	B06[N-O]	B06[N-S]
B06[N-P]	B06[N-F]	B06[N-Cl]	B06[N-Br]	B06[N-I]
B06[N-B]	B06[N-Si]	B06[N-X]	B06[O-O]	B06[O-S]
B06[O-P]	B06[O-F]	B06[O-Cl]	B06[O-Br]	B06[O-I]

B06[O-B]	B06[O-Si]	B06[O-X]	B06[S-S]	B06[S-P]
B06[S-F]	B06[S-Cl]	B06[S-Br]	B06[S-I]	B06[S-B]
B06[S-Si]	B06[S-X]	B06[P-P]	B06[P-F]	B06[P-Cl]
B06[P-Br]	B06[P-I]	B06[P-B]	B06[P-Si]	B06[P-X]
B06[F-F]	B06[F-Cl]	B06[F-Br]	B06[F-I]	B06[F-B]
B06[F-Si]	B06[F-X]	B06[Cl-Cl]	B06[Cl-Br]	B06[Cl-I]
B06[Cl-B]	B06[Cl-Si]	B06[Cl-X]	B06[Br-Br]	B06[Br-I]
B06[Br-B]	B06[Br-Si]	B06[Br-X]	B06[I-I]	B06[I-B]
B06[I-Si]	B06[I-X]	B06[B-B]	B06[B-Si]	B06[B-X]
B06[Si-Si]	B06[Si-X]	B06[X-X]	B07[C-C]	B07[C-N]
B07[C-O]	B07[C-S]	B07[C-P]	B07[C-F]	B07[C-Cl]
B07[C-Br]	B07[C-I]	B07[C-B]	B07[C-Si]	B07[C-X]
B07[N-N]	B07[N-O]	B07[N-S]	B07[N-P]	B07[N-F]
B07[N-Cl]	B07[N-Br]	B07[N-I]	B07[N-B]	B07[N-Si]
B07[N-X]	B07[O-O]	B07[O-S]	B07[O-P]	B07[O-F]
B07[O-Cl]	B07[O-Br]	B07[O-I]	B07[O-B]	B07[O-Si]
B07[O-X]	B07[S-S]	B07[S-P]	B07[S-F]	B07[S-Cl]
B07[S-Br]	B07[S-I]	B07[S-B]	B07[S-Si]	B07[S-X]
B07[P-P]	B07[P-F]	B07[P-Cl]	B07[P-Br]	B07[P-I]
B07[P-B]	B07[P-Si]	B07[P-X]	B07[F-F]	B07[F-Cl]
B07[F-Br]	B07[F-I]	B07[F-B]	B07[F-Si]	B07[F-X]
B07[Cl-Cl]	B07[Cl-Br]	B07[Cl-I]	B07[Cl-B]	B07[Cl-Si]
B07[Cl-X]	B07[Br-Br]	B07[Br-I]	B07[Br-B]	B07[Br-Si]
B07[Br-X]	B07[I-I]	B07[I-B]	B07[I-Si]	B07[I-X]
B07[B-B]	B07[B-Si]	B07[B-X]	B07[Si-Si]	B07[Si-X]
B07[X-X]	B08[C-C]	B08[C-N]	B08[C-O]	B08[C-S]
B08[C-P]	B08[C-F]	B08[C-Cl]	B08[C-Br]	B08[C-I]
B08[C-B]	B08[C-Si]	B08[C-X]	B08[N-N]	B08[N-O]
B08[N-S]	B08[N-P]	B08[N-F]	B08[N-Cl]	B08[N-Br]
B08[N-I]	B08[N-B]	B08[N-Si]	B08[N-X]	B08[O-O]
B08[O-S]	B08[O-P]	B08[O-F]	B08[O-Cl]	B08[O-Br]
B08[O-I]	B08[O-B]	B08[O-Si]	B08[O-X]	B08[S-S]
B08[S-P]	B08[S-F]	B08[S-Cl]	B08[S-Br]	B08[S-I]
B08[S-B]	B08[S-Si]	B08[S-X]	B08[P-P]	B08[P-F]
B08[P-Cl]	B08[P-Br]	B08[P-I]	B08[P-B]	B08[P-Si]
B08[P-X]	B08[F-F]	B08[F-Cl]	B08[F-Br]	B08[F-I]
B08[F-B]	B08[F-Si]	B08[F-X]	B08[Cl-Cl]	B08[Cl-Br]
B08[Cl-I]	B08[Cl-B]	B08[Cl-Si]	B08[Cl-X]	B08[Br-Br]
B08[Br-I]	B08[Br-B]	B08[Br-Si]	B08[Br-X]	B08[I-I]
B08[I-B]	B08[I-Si]	B08[I-X]	B08[B-B]	B08[B-Si]
B08[B-X]	B08[Si-Si]	B08[Si-X]	B08[X-X]	B09[C-C]
B09[C-N]	B09[C-O]	B09[C-S]	B09[C-P]	B09[C-F]
B09[C-Cl]	B09[C-Br]	B09[C-I]	B09[C-B]	B09[C-Si]
B09[C-X]	B09[N-N]	B09[N-O]	B09[N-S]	B09[N-P]

B09[N-F]	B09[N-Cl]	B09[N-Br]	B09[N-I]	B09[N-B]
B09[N-Si]	B09[N-X]	B09[O-O]	B09[O-S]	B09[O-P]
B09[O-F]	B09[O-Cl]	B09[O-Br]	B09[O-I]	B09[O-B]
B09[O-Si]	B09[O-X]	B09[S-S]	B09[S-P]	B09[S-F]
B09[S-Cl]	B09[S-Br]	B09[S-I]	B09[S-B]	B09[S-Si]
B09[S-X]	B09[P-P]	B09[P-F]	B09[P-Cl]	B09[P-Br]
B09[P-I]	B09[P-B]	B09[P-Si]	B09[P-X]	B09[F-F]
B09[F-Cl]	B09[F-Br]	B09[F-I]	B09[F-B]	B09[F-Si]
B09[F-X]	B09[Cl-Cl]	B09[Cl-Br]	B09[Cl-I]	B09[Cl-B]
B09[Cl-Si]	B09[Cl-X]	B09[Br-Br]	B09[Br-I]	B09[Br-B]
B09[Br-Si]	B09[Br-X]	B09[I-I]	B09[I-B]	B09[I-Si]
B09[I-X]	B09[B-B]	B09[B-Si]	B09[B-X]	B09[Si-Si]
B09[Si-X]	B09[X-X]	B10[C-C]	B10[C-N]	B10[C-O]
B10[C-S]	B10[C-P]	B10[C-F]	B10[C-Cl]	B10[C-Br]
B10[C-I]	B10[C-B]	B10[C-Si]	B10[C-X]	B10[N-N]
B10[N-O]	B10[N-S]	B10[N-P]	B10[N-F]	B10[N-Cl]
B10[N-Br]	B10[N-I]	B10[N-B]	B10[N-Si]	B10[N-X]
B10[O-O]	B10[O-S]	B10[O-P]	B10[O-F]	B10[O-Cl]
B10[O-Br]	B10[O-I]	B10[O-B]	B10[O-Si]	B10[O-X]
B10[S-S]	B10[S-P]	B10[S-F]	B10[S-Cl]	B10[S-Br]
B10[S-I]	B10[S-B]	B10[S-Si]	B10[S-X]	B10[P-P]
B10[P-F]	B10[P-Cl]	B10[P-Br]	B10[P-I]	B10[P-B]
B10[P-Si]	B10[P-X]	B10[F-F]	B10[F-Cl]	B10[F-Br]
B10[F-I]	B10[F-B]	B10[F-Si]	B10[F-X]	B10[Cl-Cl]
B10[Cl-Br]	B10[Cl-I]	B10[Cl-B]	B10[Cl-Si]	B10[Cl-X]
B10[Br-Br]	B10[Br-I]	B10[Br-B]	B10[Br-Si]	B10[Br-X]
B10[I-I]	B10[I-B]	B10[I-Si]	B10[I-X]	B10[B-B]
B10[B-Si]	B10[B-X]	B10[Si-Si]	B10[Si-X]	B10[X-X]
F01[C-C]	F01[C-N]	F01[C-O]	F01[C-S]	F01[C-P]
F01[C-F]	F01[C-Cl]	F01[C-Br]	F01[C-I]	F01[C-B]
F01[C-Si]	F01[C-X]	F01[N-N]	F01[N-O]	F01[N-S]
F01[N-P]	F01[N-F]	F01[N-Cl]	F01[N-Br]	F01[N-I]
F01[N-B]	F01[N-Si]	F01[N-X]	F01[O-O]	F01[O-S]
F01[O-P]	F01[O-F]	F01[O-Cl]	F01[O-Br]	F01[O-I]
F01[O-B]	F01[O-Si]	F01[O-X]	F01[S-S]	F01[S-P]
F01[S-F]	F01[S-Cl]	F01[S-Br]	F01[S-I]	F01[S-B]
F01[S-Si]	F01[S-X]	F01[P-P]	F01[P-F]	F01[P-Cl]
F01[P-Br]	F01[P-I]	F01[P-B]	F01[P-Si]	F01[P-X]
F01[F-F]	F01[F-Cl]	F01[F-Br]	F01[F-I]	F01[F-B]
F01[F-Si]	F01[F-X]	F01[Cl-Cl]	F01[Cl-Br]	F01[Cl-I]
F01[Cl-B]	F01[Cl-Si]	F01[Cl-X]	F01[Br-Br]	F01[Br-I]
F01[Br-B]	F01[Br-Si]	F01[Br-X]	F01[I-I]	F01[I-B]
F01[I-Si]	F01[I-X]	F01[B-B]	F01[B-Si]	F01[B-X]
F01[Si-Si]	F01[Si-X]	F01[X-X]	F02[C-C]	F02[C-N]

F02[C-O]	F02[C-S]	F02[C-P]	F02[C-F]	F02[C-Cl]
F02[C-Br]	F02[C-I]	F02[C-B]	F02[C-Si]	F02[C-X]
F02[N-N]	F02[N-O]	F02[N-S]	F02[N-P]	F02[N-F]
F02[N-Cl]	F02[N-Br]	F02[N-I]	F02[N-B]	F02[N-Si]
F02[N-X]	F02[O-O]	F02[O-S]	F02[O-P]	F02[O-F]
F02[O-Cl]	F02[O-Br]	F02[O-I]	F02[O-B]	F02[O-Si]
F02[O-X]	F02[S-S]	F02[S-P]	F02[S-F]	F02[S-Cl]
F02[S-Br]	F02[S-I]	F02[S-B]	F02[S-Si]	F02[S-X]
F02[P-P]	F02[P-F]	F02[P-Cl]	F02[P-Br]	F02[P-I]
F02[P-B]	F02[P-Si]	F02[P-X]	F02[F-F]	F02[F-Cl]
F02[F-Br]	F02[F-I]	F02[F-B]	F02[F-Si]	F02[F-X]
F02[Cl-Cl]	F02[Cl-Br]	F02[Cl-I]	F02[Cl-B]	F02[Cl-Si]
F02[Cl-X]	F02[Br-Br]	F02[Br-I]	F02[Br-B]	F02[Br-Si]
F02[Br-X]	F02[I-I]	F02[I-B]	F02[I-Si]	F02[I-X]
F02[B-B]	F02[B-Si]	F02[B-X]	F02[Si-Si]	F02[Si-X]
F02[X-X]	F03[C-C]	F03[C-N]	F03[C-O]	F03[C-S]
F03[C-P]	F03[C-F]	F03[C-Cl]	F03[C-Br]	F03[C-I]
F03[C-B]	F03[C-Si]	F03[C-X]	F03[N-N]	F03[N-O]
F03[N-S]	F03[N-P]	F03[N-F]	F03[N-Cl]	F03[N-Br]
F03[N-I]	F03[N-B]	F03[N-Si]	F03[N-X]	F03[O-O]
F03[O-S]	F03[O-P]	F03[O-F]	F03[O-Cl]	F03[O-Br]
F03[O-I]	F03[O-B]	F03[O-Si]	F03[O-X]	F03[S-S]
F03[S-P]	F03[S-F]	F03[S-Cl]	F03[S-Br]	F03[S-I]
F03[S-B]	F03[S-Si]	F03[S-X]	F03[P-P]	F03[P-F]
F03[P-Cl]	F03[P-Br]	F03[P-I]	F03[P-B]	F03[P-Si]
F03[P-X]	F03[F-F]	F03[F-Cl]	F03[F-Br]	F03[F-I]
F03[F-B]	F03[F-Si]	F03[F-X]	F03[Cl-Cl]	F03[Cl-Br]
F03[Cl-I]	F03[Cl-B]	F03[Cl-Si]	F03[Cl-X]	F03[Br-Br]
F03[Br-I]	F03[Br-B]	F03[Br-Si]	F03[Br-X]	F03[I-I]
F03[I-B]	F03[I-Si]	F03[I-X]	F03[B-B]	F03[B-Si]
F03[B-X]	F03[Si-Si]	F03[Si-X]	F03[X-X]	F04[C-C]
F04[C-N]	F04[C-O]	F04[C-S]	F04[C-P]	F04[C-F]
F04[C-Cl]	F04[C-Br]	F04[C-I]	F04[C-B]	F04[C-Si]
F04[C-X]	F04[N-N]	F04[N-O]	F04[N-S]	F04[N-P]
F04[N-F]	F04[N-Cl]	F04[N-Br]	F04[N-I]	F04[N-B]
F04[N-Si]	F04[N-X]	F04[O-O]	F04[O-S]	F04[O-P]
F04[O-F]	F04[O-Cl]	F04[O-Br]	F04[O-I]	F04[O-B]
F04[O-Si]	F04[O-X]	F04[S-S]	F04[S-P]	F04[S-F]
F04[S-Cl]	F04[S-Br]	F04[S-I]	F04[S-B]	F04[S-Si]
F04[S-X]	F04[P-P]	F04[P-F]	F04[P-Cl]	F04[P-Br]
F04[P-I]	F04[P-B]	F04[P-Si]	F04[P-X]	F04[F-F]
F04[F-Cl]	F04[F-Br]	F04[F-I]	F04[F-B]	F04[F-Si]
F04[F-X]	F04[Cl-Cl]	F04[Cl-Br]	F04[Cl-I]	F04[Cl-B]
F04[Cl-Si]	F04[Cl-X]	F04[Br-Br]	F04[Br-I]	F04[Br-B]

F04[Br-Si]	F04[Br-X]	F04[I-I]	F04[I-B]	F04[I-Si]
F04[I-X]	F04[B-B]	F04[B-Si]	F04[B-X]	F04[Si-Si]
F04[Si-X]	F04[X-X]	F05[C-C]	F05[C-N]	F05[C-O]
F05[C-S]	F05[C-P]	F05[C-F]	F05[C-Cl]	F05[C-Br]
F05[C-I]	F05[C-B]	F05[C-Si]	F05[C-X]	F05[N-N]
F05[N-O]	F05[N-S]	F05[N-P]	F05[N-F]	F05[N-Cl]
F05[N-Br]	F05[N-I]	F05[N-B]	F05[N-Si]	F05[N-X]
F05[O-O]	F05[O-S]	F05[O-P]	F05[O-F]	F05[O-Cl]
F05[O-Br]	F05[O-I]	F05[O-B]	F05[O-Si]	F05[O-X]
F05[S-S]	F05[S-P]	F05[S-F]	F05[S-Cl]	F05[S-Br]
F05[S-I]	F05[S-B]	F05[S-Si]	F05[S-X]	F05[P-P]
F05[P-F]	F05[P-Cl]	F05[P-Br]	F05[P-I]	F05[P-B]
F05[P-Si]	F05[P-X]	F05[F-F]	F05[F-Cl]	F05[F-Br]
F05[F-I]	F05[F-B]	F05[F-Si]	F05[F-X]	F05[Cl-Cl]
F05[Cl-Br]	F05[Cl-I]	F05[Cl-B]	F05[Cl-Si]	F05[Cl-X]
F05[Br-Br]	F05[Br-I]	F05[Br-B]	F05[Br-Si]	F05[Br-X]
F05[I-I]	F05[I-B]	F05[I-Si]	F05[I-X]	F05[B-B]
F05[B-Si]	F05[B-X]	F05[Si-Si]	F05[Si-X]	F05[X-X]
F06[C-C]	F06[C-N]	F06[C-O]	F06[C-S]	F06[C-P]
F06[C-F]	F06[C-Cl]	F06[C-Br]	F06[C-I]	F06[C-B]
F06[C-Si]	F06[C-X]	F06[N-N]	F06[N-O]	F06[N-S]
F06[N-P]	F06[N-F]	F06[N-Cl]	F06[N-Br]	F06[N-I]
F06[N-B]	F06[N-Si]	F06[N-X]	F06[O-O]	F06[O-S]
F06[O-P]	F06[O-F]	F06[O-Cl]	F06[O-Br]	F06[O-I]
F06[O-B]	F06[O-Si]	F06[O-X]	F06[S-S]	F06[S-P]
F06[S-F]	F06[S-Cl]	F06[S-Br]	F06[S-I]	F06[S-B]
F06[S-Si]	F06[S-X]	F06[P-P]	F06[P-F]	F06[P-Cl]
F06[P-Br]	F06[P-I]	F06[P-B]	F06[P-Si]	F06[P-X]
F06[F-F]	F06[F-Cl]	F06[F-Br]	F06[F-I]	F06[F-B]
F06[F-Si]	F06[F-X]	F06[Cl-Cl]	F06[Cl-Br]	F06[Cl-I]
F06[Cl-B]	F06[Cl-Si]	F06[Cl-X]	F06[Br-Br]	F06[Br-I]
F06[Br-B]	F06[Br-Si]	F06[Br-X]	F06[I-I]	F06[I-B]
F06[I-Si]	F06[I-X]	F06[B-B]	F06[B-Si]	F06[B-X]
F06[Si-Si]	F06[Si-X]	F06[X-X]	F07[C-C]	F07[C-N]
F07[C-O]	F07[C-S]	F07[C-P]	F07[C-F]	F07[C-Cl]
F07[C-Br]	F07[C-I]	F07[C-B]	F07[C-Si]	F07[C-X]
F07[N-N]	F07[N-O]	F07[N-S]	F07[N-P]	F07[N-F]
F07[N-Cl]	F07[N-Br]	F07[N-I]	F07[N-B]	F07[N-Si]
F07[N-X]	F07[O-O]	F07[O-S]	F07[O-P]	F07[O-F]
F07[O-Cl]	F07[O-Br]	F07[O-I]	F07[O-B]	F07[O-Si]
F07[O-X]	F07[S-S]	F07[S-P]	F07[S-F]	F07[S-Cl]
F07[S-Br]	F07[S-I]	F07[S-B]	F07[S-Si]	F07[S-X]
F07[P-P]	F07[P-F]	F07[P-Cl]	F07[P-Br]	F07[P-I]
F07[P-B]	F07[P-Si]	F07[P-X]	F07[F-F]	F07[F-Cl]

F07[F-Br]	F07[F-I]	F07[F-B]	F07[F-Si]	F07[F-X]
F07[Cl-Cl]	F07[Cl-Br]	F07[Cl-I]	F07[Cl-B]	F07[Cl-Si]
F07[Cl-X]	F07[Br-Br]	F07[Br-I]	F07[Br-B]	F07[Br-Si]
F07[Br-X]	F07[I-I]	F07[I-B]	F07[I-Si]	F07[I-X]
F07[B-B]	F07[B-Si]	F07[B-X]	F07[Si-Si]	F07[Si-X]
F07[X-X]	F08[C-C]	F08[C-N]	F08[C-O]	F08[C-S]
F08[C-P]	F08[C-F]	F08[C-Cl]	F08[C-Br]	F08[C-I]
F08[C-B]	F08[C-Si]	F08[C-X]	F08[N-N]	F08[N-O]
F08[N-S]	F08[N-P]	F08[N-F]	F08[N-Cl]	F08[N-Br]
F08[N-I]	F08[N-B]	F08[N-Si]	F08[N-X]	F08[O-O]
F08[O-S]	F08[O-P]	F08[O-F]	F08[O-Cl]	F08[O-Br]
F08[O-I]	F08[O-B]	F08[O-Si]	F08[O-X]	F08[S-S]
F08[S-P]	F08[S-F]	F08[S-Cl]	F08[S-Br]	F08[S-I]
F08[S-B]	F08[S-Si]	F08[S-X]	F08[P-P]	F08[P-F]
F08[P-Cl]	F08[P-Br]	F08[P-I]	F08[P-B]	F08[P-Si]
F08[P-X]	F08[F-F]	F08[F-Cl]	F08[F-Br]	F08[F-I]
F08[F-B]	F08[F-Si]	F08[F-X]	F08[Cl-Cl]	F08[Cl-Br]
F08[Cl-I]	F08[Cl-B]	F08[Cl-Si]	F08[Cl-X]	F08[Br-Br]
F08[Br-I]	F08[Br-B]	F08[Br-Si]	F08[Br-X]	F08[I-I]
F08[I-B]	F08[I-Si]	F08[I-X]	F08[B-B]	F08[B-Si]
F08[B-X]	F08[Si-Si]	F08[Si-X]	F08[X-X]	F09[C-C]
F09[C-N]	F09[C-O]	F09[C-S]	F09[C-P]	F09[C-F]
F09[C-Cl]	F09[C-Br]	F09[C-I]	F09[C-B]	F09[C-Si]
F09[C-X]	F09[N-N]	F09[N-O]	F09[N-S]	F09[N-P]
F09[N-F]	F09[N-Cl]	F09[N-Br]	F09[N-I]	F09[N-B]
F09[N-Si]	F09[N-X]	F09[O-O]	F09[O-S]	F09[O-P]
F09[O-F]	F09[O-Cl]	F09[O-Br]	F09[O-I]	F09[O-B]
F09[O-Si]	F09[O-X]	F09[S-S]	F09[S-P]	F09[S-F]
F09[S-Cl]	F09[S-Br]	F09[S-I]	F09[S-B]	F09[S-Si]
F09[S-X]	F09[P-P]	F09[P-F]	F09[P-Cl]	F09[P-Br]
F09[P-I]	F09[P-B]	F09[P-Si]	F09[P-X]	F09[F-F]
F09[F-Cl]	F09[F-Br]	F09[F-I]	F09[F-B]	F09[F-Si]
F09[F-X]	F09[Cl-Cl]	F09[Cl-Br]	F09[Cl-I]	F09[Cl-B]
F09[Cl-Si]	F09[Cl-X]	F09[Br-Br]	F09[Br-I]	F09[Br-B]
F09[Br-Si]	F09[Br-X]	F09[I-I]	F09[I-B]	F09[I-Si]
F09[I-X]	F09[B-B]	F09[B-Si]	F09[B-X]	F09[Si-Si]
F09[Si-X]	F09[X-X]	F10[C-C]	F10[C-N]	F10[C-O]
F10[C-S]	F10[C-P]	F10[C-F]	F10[C-Cl]	F10[C-Br]
F10[C-I]	F10[C-B]	F10[C-Si]	F10[C-X]	F10[N-N]
F10[N-O]	F10[N-S]	F10[N-P]	F10[N-F]	F10[N-Cl]
F10[N-Br]	F10[N-I]	F10[N-B]	F10[N-Si]	F10[N-X]
F10[O-O]	F10[O-S]	F10[O-P]	F10[O-F]	F10[O-Cl]
F10[O-Br]	F10[O-I]	F10[O-B]	F10[O-Si]	F10[O-X]
F10[S-S]	F10[S-P]	F10[S-F]	F10[S-Cl]	F10[S-Br]

F10[S-I]	F10[S-B]	F10[S-Si]	F10[S-X]	F10[P-P]
F10[P-F]	F10[P-Cl]	F10[P-Br]	F10[P-I]	F10[P-B]
F10[P-Si]	F10[P-X]	F10[F-F]	F10[F-Cl]	F10[F-Br]
F10[F-I]	F10[F-B]	F10[F-Si]	F10[F-X]	F10[Cl-Cl]
F10[Cl-Br]	F10[Cl-I]	F10[Cl-B]	F10[Cl-Si]	F10[Cl-X]
F10[Br-Br]	F10[Br-I]	F10[Br-B]	F10[Br-Si]	F10[Br-X]
F10[I-I]	F10[I-B]	F10[I-Si]	F10[I-X]	F10[B-B]
F10[B-Si]	F10[B-X]	F10[Si-Si]	F10[Si-X]	F10[X-X]
G(N..I)	G(O..I)	G(S..I)	G(P..F)	G(P..Cl)
G(P..Br)	G(P..I)	G(F..Br)	G(F..I)	G(Cl..I)
G(Br..Br)	G(Br..I)	G(I..I)	Ro5	cRo5
CATS3D_00_DD	CATS3D_01_DD	CATS3D_02_DD	CATS3D_03_DD	CATS3D_04_DD
CATS3D_05_DD	CATS3D_06_DD	CATS3D_07_DD	CATS3D_08_DD	CATS3D_09_DD
CATS3D_10_DD	CATS3D_11_DD	CATS3D_12_DD	CATS3D_13_DD	CATS3D_14_DD
CATS3D_15_DD	CATS3D_16_DD	CATS3D_17_DD	CATS3D_18_DD	CATS3D_19_DD
CATS3D_00_DA	CATS3D_01_DA	CATS3D_02_DA	CATS3D_03_DA	CATS3D_04_DA
CATS3D_05_DA	CATS3D_06_DA	CATS3D_07_DA	CATS3D_08_DA	CATS3D_09_DA
CATS3D_10_DA	CATS3D_11_DA	CATS3D_12_DA	CATS3D_13_DA	CATS3D_14_DA
CATS3D_15_DA	CATS3D_16_DA	CATS3D_17_DA	CATS3D_18_DA	CATS3D_19_DA
CATS3D_00_DP	CATS3D_01_DP	CATS3D_02_DP	CATS3D_03_DP	CATS3D_04_DP
CATS3D_05_DP	CATS3D_06_DP	CATS3D_07_DP	CATS3D_08_DP	CATS3D_09_DP
CATS3D_10_DP	CATS3D_11_DP	CATS3D_12_DP	CATS3D_13_DP	CATS3D_14_DP
CATS3D_15_DP	CATS3D_16_DP	CATS3D_17_DP	CATS3D_18_DP	CATS3D_19_DP
CATS3D_00_DN	CATS3D_01_DN	CATS3D_02_DN	CATS3D_03_DN	CATS3D_04_DN
CATS3D_05_DN	CATS3D_06_DN	CATS3D_07_DN	CATS3D_08_DN	CATS3D_09_DN
CATS3D_10_DN	CATS3D_11_DN	CATS3D_12_DN	CATS3D_13_DN	CATS3D_14_DN
CATS3D_15_DN	CATS3D_16_DN	CATS3D_17_DN	CATS3D_18_DN	CATS3D_19_DN
CATS3D_00_DL	CATS3D_01_DL	CATS3D_02_DL	CATS3D_03_DL	CATS3D_04_DL
CATS3D_05_DL	CATS3D_06_DL	CATS3D_07_DL	CATS3D_08_DL	CATS3D_09_DL
CATS3D_10_DL	CATS3D_11_DL	CATS3D_12_DL	CATS3D_13_DL	CATS3D_14_DL
CATS3D_15_DL	CATS3D_16_DL	CATS3D_17_DL	CATS3D_18_DL	CATS3D_19_DL
CATS3D_00_AA	CATS3D_01_AA	CATS3D_02_AA	CATS3D_03_AA	CATS3D_04_AA
CATS3D_05_AA	CATS3D_06_AA	CATS3D_07_AA	CATS3D_08_AA	CATS3D_09_AA
CATS3D_10_AA	CATS3D_11_AA	CATS3D_12_AA	CATS3D_13_AA	CATS3D_14_AA
CATS3D_15_AA	CATS3D_16_AA	CATS3D_17_AA	CATS3D_18_AA	CATS3D_19_AA
CATS3D_00_AP	CATS3D_01_AP	CATS3D_02_AP	CATS3D_03_AP	CATS3D_04_AP
CATS3D_05_AP	CATS3D_06_AP	CATS3D_07_AP	CATS3D_08_AP	CATS3D_09_AP
CATS3D_10_AP	CATS3D_11_AP	CATS3D_12_AP	CATS3D_13_AP	CATS3D_14_AP
CATS3D_15_AP	CATS3D_16_AP	CATS3D_17_AP	CATS3D_18_AP	CATS3D_19_AP
CATS3D_00_AN	CATS3D_01_AN	CATS3D_02_AN	CATS3D_03_AN	CATS3D_04_AN
CATS3D_05_AN	CATS3D_06_AN	CATS3D_07_AN	CATS3D_08_AN	CATS3D_09_AN
CATS3D_10_AN	CATS3D_11_AN	CATS3D_12_AN	CATS3D_13_AN	CATS3D_14_AN
CATS3D_15_AN	CATS3D_16_AN	CATS3D_17_AN	CATS3D_18_AN	CATS3D_19_AN
CATS3D_00_AL	CATS3D_01_AL	CATS3D_02_AL	CATS3D_03_AL	CATS3D_04_AL

CATS3D_05_AL	CATS3D_06_AL	CATS3D_07_AL	CATS3D_08_AL	CATS3D_09_AL
CATS3D_10_AL	CATS3D_11_AL	CATS3D_12_AL	CATS3D_13_AL	CATS3D_14_AL
CATS3D_15_AL	CATS3D_16_AL	CATS3D_17_AL	CATS3D_18_AL	CATS3D_19_AL
CATS3D_00_PP	CATS3D_01_PP	CATS3D_02_PP	CATS3D_03_PP	CATS3D_04_PP
CATS3D_05_PP	CATS3D_06_PP	CATS3D_07_PP	CATS3D_08_PP	CATS3D_09_PP
CATS3D_10_PP	CATS3D_11_PP	CATS3D_12_PP	CATS3D_13_PP	CATS3D_14_PP
CATS3D_15_PP	CATS3D_16_PP	CATS3D_17_PP	CATS3D_18_PP	CATS3D_19_PP
CATS3D_00_PN	CATS3D_01_PN	CATS3D_02_PN	CATS3D_03_PN	CATS3D_04_PN
CATS3D_05_PN	CATS3D_06_PN	CATS3D_07_PN	CATS3D_08_PN	CATS3D_09_PN
CATS3D_10_PN	CATS3D_11_PN	CATS3D_12_PN	CATS3D_13_PN	CATS3D_14_PN
CATS3D_15_PN	CATS3D_16_PN	CATS3D_17_PN	CATS3D_18_PN	CATS3D_19_PN
CATS3D_00_PL	CATS3D_01_PL	CATS3D_02_PL	CATS3D_03_PL	CATS3D_04_PL
CATS3D_05_PL	CATS3D_06_PL	CATS3D_07_PL	CATS3D_08_PL	CATS3D_09_PL
CATS3D_10_PL	CATS3D_11_PL	CATS3D_12_PL	CATS3D_13_PL	CATS3D_14_PL
CATS3D_15_PL	CATS3D_16_PL	CATS3D_17_PL	CATS3D_18_PL	CATS3D_19_PL
CATS3D_00_NN	CATS3D_01_NN	CATS3D_02_NN	CATS3D_03_NN	CATS3D_04_NN
CATS3D_05_NN	CATS3D_06_NN	CATS3D_07_NN	CATS3D_08_NN	CATS3D_09_NN
CATS3D_10_NN	CATS3D_11_NN	CATS3D_12_NN	CATS3D_13_NN	CATS3D_14_NN
CATS3D_15_NN	CATS3D_16_NN	CATS3D_17_NN	CATS3D_18_NN	CATS3D_19_NN
CATS3D_00_NL	CATS3D_01_NL	CATS3D_02_NL	CATS3D_03_NL	CATS3D_04_NL
CATS3D_05_NL	CATS3D_06_NL	CATS3D_07_NL	CATS3D_08_NL	CATS3D_09_NL
CATS3D_10_NL	CATS3D_11_NL	CATS3D_12_NL	CATS3D_13_NL	CATS3D_14_NL
CATS3D_15_NL	CATS3D_16_NL	CATS3D_17_NL	CATS3D_18_NL	CATS3D_19_NL
CATS3D_00_LL	CATS3D_01_LL	CATS3D_02_LL	CATS3D_03_LL	CATS3D_04_LL
CATS3D_05_LL	CATS3D_06_LL	CATS3D_07_LL	CATS3D_08_LL	CATS3D_09_LL
CATS3D_10_LL	CATS3D_11_LL	CATS3D_12_LL	CATS3D_13_LL	CATS3D_14_LL
CATS3D_15_LL	CATS3D_16_LL	CATS3D_17_LL	CATS3D_18_LL	CATS3D_19_LL

(2540 categorical features in total)

Acknowledgements

This study was financially supported in part by Takeda Pharmaceutical Company Limited (COCKPI-T Funding program), and also supported by Takeda researchers K. Hidaka, M. Toyofuku, A. Okabe, K. Imamura, T. Tawaraishi, M. Mochizuki, J. Li, T. Sugimoto for screening compound selection and provision of tangible materials, and by T. Hidaka for acquisition of descriptor data. This study was also supported by IDS (Institute for Datability Science) grant of Osaka University for Y. Matsushita and K. Nakatani.