Supplementary Figures

The ANN model (Figure S1) consisted of 1 densely connected hidden layer with 100 nodes each and a 'Rectified Linear Unit' (relu) as the activation function [keras.layers.Dense(100, activation = 'relu')], 1 hidden layer with 100 nodes, and 'Dropout' as the activation function [keras.layers.Dropout], and 1 densely connected hidden layer with 100 nodes and 'Sigmoid' as the activation function [keras.layers.Dense(100, activation = 'sigmoid')] (Figure S1). The purpose of the dropout layer is to remove all the unnecessary nodes and connections that were created during the development of the model but did not play a significant role in the outcome. The removal of these nodes decreases the computational processing time and reduces the potential for over-fitting. The optimal number of hidden layers was determined by comparing four model scenarios with an increasing number of hidden layers, 3, 4, 5, and 6 (Figure S3). We examined the MAE at the end of the training cycle at 500 epochs and the overall minimum MAE. As shown in Figure S3, we did not observe large differences between the four scenarios.

The optimizer used was 'AdamOptimizer' with 0.001 as the training step [tf.train.AdamOptimizer(0.001)]. We ran the model for 500 epochs to identify the time point at which the model reaches the minimum MAE for the testing set and we set the model's training patience to that number [early_stop = keras.callbacks.EarlyStopping (monitor = 'val_loss', patience = 100)] (Figure S3), approximately 100 epochs. The model then uses its algorithm to identify the appropriate time to stop the training after 100 epochs. (Figure S2).

Figure S1. Schematic diagram of the ANN model showing the hidden layers, number of nodes, and the activation functions. The model consisted of 1 densely connected hidden layer with 100 nodes each and 'Rectified Linear Unit' (relu) as the activation function, 1 hidden layer with 100 nodes and 'Dropout' as the activation function, and 1 densely connected hidden layer with 100 nodes and 'Sigmoid' as the activation function. The purpose of the dropout layer is to remove all the unnecessary nodes and connections that were created during the development of the model but did not play a significant role in the outcome. The removal of these nodes decreases the computational processing time and reduces the potential for over-fitting.

Figure S2. MAE during the optimization of the ANN model for the training (blue) and testing sets (orange). We ran the model for 500 epochs (A) to identify the time point at which the validation MAE reached its minimum and we set the training patience of the model to that point, approximately 100 epochs. Using its supportive algorithm, the model then identified the appropriate time to stop the training after 100 epochs, which, in this case, was approximately 120 epochs (B).

Figure S3. MAE for the training and testing set of the ANN for four different numbers of hidden layers: (A) 3 hidden layers, (B) 7 hidden layers, (C) 11 hidden layers, and (D) 21 hidden layers. The dashed gray line shows the MAE of the testing set at 500 epochs. It is important to note that the model also contains a function (EarlyStopping) for identifying the time point with the smallest MAE for the testing set and it will stop the training automatically at that point. For example, in plot (C), the MAE at 500 epochs is 0.33, however, the minimum MAE is 0.31, which is located at approximately 120 epochs.

Figure S4. A typical chromatogram, and M2 spectra of several typical compounds in foods.

Figure S1

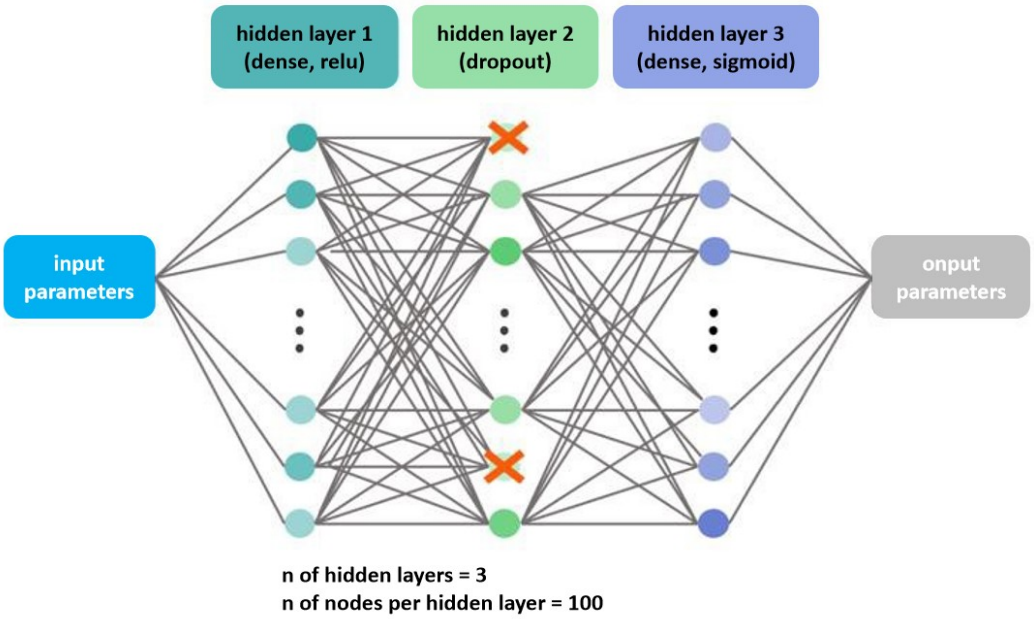

hidden layer 1
(dense, relu)
hidden layer 2
(dropout)
hidden layer 3
(dense, sigmoid)
input
parameters
onput
parameters
n of hidden layers = 3
n of nodes per hidden layer = 100

Figure S2

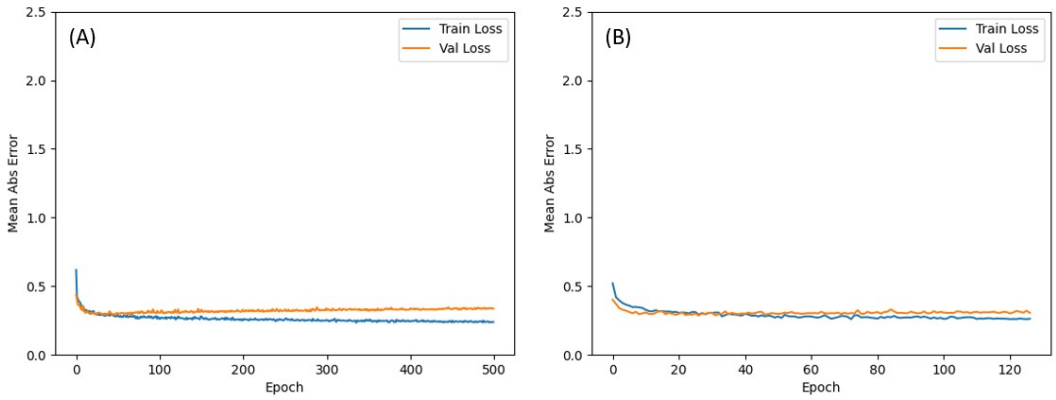

(A)
Train Loss
Val Loss
Mean Abs Error
Epoch
(B)
Train Loss
Val Loss
Mean Abs Error
Epoch

Figure S3

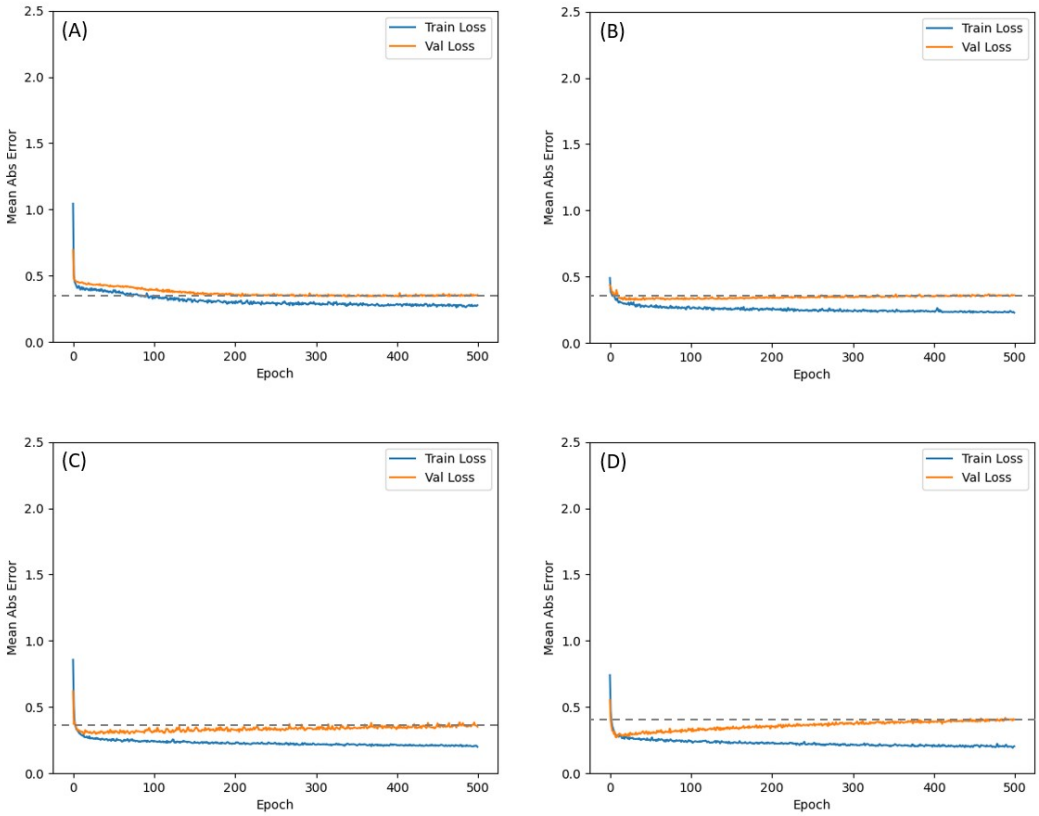
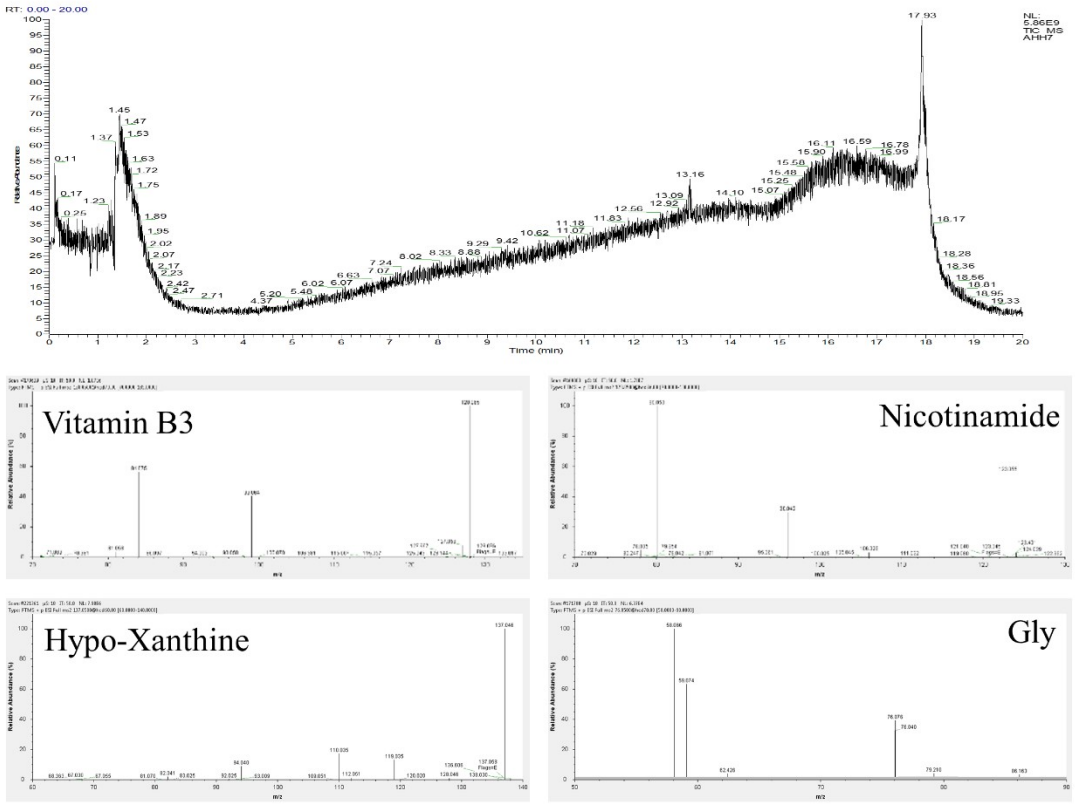

(A)
(B)
(C)
(D)
Train Loss
Val Loss
Mean Abs Error
Epoch
0.0
0.5
1.0
1.5
2.0
2.5
0
100
200
300
400
500

Figure S4


RT: 0.00 - 20.00
NL:
5.66E9
TIC_MS
AHH7
Time (min)
Vitamin B3
Nicotinamide
Hypo-Xanthine
Gly

Supplementary Tables

Table S1: Detailed information of material.

There is detailed information on the 70 compounds covered in the article (e.g. classification, name, manufacturer, origin, purity, formula, m/z, CAS, and SMILES).

Table S2: Detailed information on liquid chromatography and mass spectrometry conditions.

There is detailed information on the liquid chromatography and mass spectrometry conditions used in the article (e.g. scan mode, source parameters, liquid chromatography column, mobile phase (A/B), elution mode, etc.).

Table S1

Classification	Name	Manufacturer	Origin	Purity	Formula	m/z	CAS	SMILES
Amino acid	Glycine	Sigma Aldrich	St. Louis, MO, USA	99.00%	C2H5NO2	76.04	56-40-6	OC(CN)=O
	Alanine	Sigma Aldrich	St. Louis, MO, USA	99.00%	C3H7NO2	90.056	56-41-7	OC([C@H](C)N)=O
	Leucine	Sigma Aldrich	St. Louis, MO, USA	98.00%	C6H13NO2	132.102	61-90-5	OC([C@H](CC(C)C)N)=O
	Isoleucine	Sigma Aldrich	St. Louis, MO, USA	≥98.00%	C6H13NO2	132.102	73-32-5	OC([C@@H](N)[C@@H](C)CC)=O
	Valine	Sigma Aldrich	St. Louis, MO, USA	≥98.00%	C5H11NO2	118.087	72-18-4	OC([C@@H](N)C(C)C)=O
	Proline	Sigma Aldrich	St. Louis, MO, USA	≥99.00%	C5H9NO2	116.071	147-85-3	OC([C@H]1NCCC1)=O
	Phenylalanine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C9H11NO2	166.086	63-91-2	OC([C@@H](N)CC1=CC=CC=C1)=O
	Methionine	Sigma Aldrich	St. Louis, MO, USA	99.00%	C5H11NO2S	150.059	63-68-3	OC([C@H](CCSC)N)=O
	Tryptophan	Sigma Aldrich	St. Louis, MO, USA	98.00%	C11H12N2O2	205.097	73-22-3	OC([C@@H](N)CC1=CNC2=CC=CC=C21)=O
	Serine	Sigma Aldrich	St. Louis, MO, USA	99.00%	C3H7NO3	106.05	56-45-1	OC([C@H](CO)N)=O
	Glutamine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C5H10N2O3	147.077	56-85-9	OC([C@H](CCC(N)=O)N)=O
	Threonine	Sigma Aldrich	St. Louis, MO, USA	≥98.00%	C4H9NO3	120.066	72-19-5	O=C(O)[C@@H](N)[C@H](O)C
	Cysteine	Sigma Aldrich	St. Louis, MO, USA	97.00%	C3H7NO2S	122.027	52-90-4	OC([C@H](CS)N)=O
	Asparagine	Sigma Aldrich	St. Louis, MO, USA	98.00%	C4H8N2O3	133.061	70-47-3	O=C(O)[C@H](N)CC(N)=O
	Tyrosine	Aladdin Reagent Co., Ltd	Shanghai, China	≥99.00%	C9H11NO3	182.081	60-18-4	OC([C@@H](N)CC1=CC=C(O)C=C1)=O
	Aspartic acid	Sigma Aldrich	St. Louis, MO, USA	99.00%	C4H7NO4	134.045	56-84-8	O=C(O)[C@H](N)CC(O)=O
	Glutamic acid	Sigma Aldrich	St. Louis, MO, USA	98.00%	C5H9NO4	148.061	56-86-0	OC([C@H](CCC(O)=O)N)=O
	Lysine	Sigma Aldrich	St. Louis, MO, USA	98.00%	C6H14N2O2	147.113	56-87-1	OC([C@H](CCCCN)N)=O
	Arginine	Sigma Aldrich	St. Louis, MO, USA	98.00%	C6H14N4O2	175.119	74-79-3	OC([C@H](CCC/N=C(N)/N)N)=O
	Histidine	Sinopharm Chemical Reagent Co., Ltd	Shanghai, China	99.00%	C6H9N3O2	156.077	71-00-1	O=C(O)[C@@H](N)CC1=CN=CN1
	Citrulline	Aladdin Reagent Co., Ltd	Shanghai, China	98.00%	C6H13N3O3	176.103	372-75-8	OC([C@H](CCCNC(N)=O)N)=O
	Taurine	Aladdin Reagent Co., Ltd	Shanghai, China	98.00%	C2H7NO3S	126.022	107-35-7	NCCS(O)(=O)=O
Metabolites	Hypotaurine	Sigma Aldrich	St. Louis, MO, USA	≥98.00% (TLC)	C2H7NO2S	110.027	300-84-5	NCCS(O)=O

	L-Hydroxyproline	Solarbio Reagent Co., Ltd	Shanghai, China	≥99.0%	C5H9NO3	132.066	51-35-4	O[C@H]1CN[C@H](C(O)=O)C1
	Spermidine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C7H19N3	146.165	124-20-9	NCCCNCCCCN
	N-omiga-acetylhistamine	Sigma Aldrich	St. Louis, MO, USA	98.00%	C7H11N3O	154.097	673-49-4	CC(NCCC1=CN=CN1)=O
	L-pyroglutamic acid	Sigma Aldrich	St. Louis, MO, USA	98.00%	C5H7NO3	130.05	98-79-3	OC([C@H]1NC(CC1)=O)=O
	Tyramine	Sigma Aldrich	St. Louis, MO, USA	≥98.50% (GC)	C8H11NO	138.091	51-67-2	NCCC1=CC=C(O)C=C1
	Histamine	Sigma Aldrich	St. Louis, MO, USA	≥99%	C5H9N3	112.087	51-45-6	NCCC1=CN=CN1
	Creatinine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C4H7N3O	114.067	60-27-5	O=C1CN(C)C(N)=N1
	L-Dihydroorotic acid	Sigma Aldrich	St. Louis, MO, USA	≥99.00%	C5H6N2O4	159.04	5988-19-2	O=C(C[C@@H](C(O)=O)N1)NC1=O
	L-Carnosine	Aladdin Reagent Co., Ltd	Shanghai, China	98.00%	C9H14N4O3	227.114	305-84-0	OC([C@@H](NC(CCN)=O)CC1=CN=CN1)=O
	Uric acid	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C5H4N4O3	169.036	69-93-2	O=C(N1)NC2=C1NC(NC2=O)=O
	Betaine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C5H11NO2	118.086	107-43-7	[O-]C(C[N+](C)(C)C)=O
	Sarcosine	Macklin Biochemical Co., Ltd	Shanghai, China	99.00%	C3H7NO2	90.056	107-97-1	OC(CNC)=O
Nucleobases	Cytosine	Sigma Aldrich	St. Louis, MO, USA	≥99.0%	C4H5N3O	112.051	71-30-7	NC(N1)=CC=NC1=O
	Hypoxanthine	Macklin Biochemical Co., Ltd	Shanghai, China	99.00%	C5H4N4O	137.046	68-94-0	O=C1C2=C(N=CN2)N=CN1
	Xanthine	Aladdin Reagent Co., Ltd	Shanghai, China	≥99.5% (HPLC)	C5H4N4O2	153.041	69-89-6	O=C(N1)NC(C2=C1N=CN2)=O
	Guanine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C5H5N5O	152.057	73-40-5	O=C1C2=C(N=CN2)N=C(N)N1
	Uracil	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C4H4N2O2	113.035	66-22-8	O=C(N1)C=CNC1=O
	Adenine	Aladdin Reagent Co., Ltd	Shanghai, China	≥99.5% (HPLC)	C5H5N5	136.062	73-24-5	NC1=C2C(N=CN2)=NC=N1
Nucleosides	Beta-Thymidine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%	C10H14N2O5	243.097	50-89-5	O=C(C(C)=C1)NC(N1[C@H]2C[C@H](O)[C@@H](CO)O2)=O
	Cytidine	Aladdin Reagent Co., Ltd	Shanghai, China	≥99.0% (HPLC)	C9H13N3O5	244.093	65-46-3	O=C(N=C(N)C=C1)N1[C@H]2[C@H](O)[C@H](O)[C@@H](CO)O2
	Inosine	Sigma Aldrich	St. Louis, MO, USA	≥99.0% (HPLC)	C10H12N4O5	269.088	58-63-9	O[C@H]1[C@@H](O)[C@H](N2C(N=CNC3=O)=C3N=C2)O[C@@H]1CO
	Xanthosine	Aladdin Reagent Co., Ltd	Shanghai, China	≥98.00% (HPLC)	C10H12N4O6	285.083	146-80-5	O[C@H]1[C@@H](O)[C@H](N2C(NC(NC3=O)=O)=C3N=C2)O[C@@H]1CO
	Guanosine	Aladdin Reagent Co., Ltd	Shanghai, China	98.00%, for cell culture	C10H13N5O5	284.099	118-00-3	O[C@@H]1[C@@H](CO)O[C@@H](N2C(N=C(N)NC3=O)=C3N=C2)[C@@H]1O
	Uridine	Aladdin Reagent Co., Ltd	Shanghai, China	99.00%, for cell culture	C9H12N2O6	245.077	58-96-8	O[C@H]1[C@@H](O)[C@H](N2C(NC(C=C2)=O)=O)O[C@@H]1CO
	Adenosine	Aladdin Reagent Co., Ltd	Shanghai, China	＞99.50% (HPLC)	C10H13N5O4	268.104	58-61-7	O[C@@H]1[C@@H](CO)O[C@@H](N2C3=NC=NC(N)=C3N=C2)[C@@H]1O
Vitamins	Pyridoxine	Sigma Aldrich	St. Louis, MO, USA	≥99.00%	C8H11NO3	170.081	65-23-6	OCC1=C(CO)C=NC(C)=C1O

	Pyridoxal	Bide Pharmatech Ltd.	Shanghai, China	≥99.00%	C8H9NO3	168.066	66-72-8	OCC1=C(C=O)C(O)=C(C)N=C1
	Pyridoxamine	Medchemexpress	New Jersey, USA	≥99.00%	C8H12N2O2	169.097	85-87-0	NCC1=C(CO)C=NC(C)=C1O
	Thiamine	Macklin Biochemical Co., Ltd	Shanghai, China	98.00%	C12H17ClN4OS	265.112	59-43-8	NC1=NC(C)=NC=C1C[N+]2=CSC(CCO)=C2C
	Biotin	Sigma Aldrich	St. Louis, MO, USA	≥98.00%	C10H16N2O3S	245.096	58-85-5	O=C1N[C@]2([H])[C@]([C@H](CCCCC(O)=O)SC2)([H])N1
	D-Pantothenic acid	Macklin Biochemical Co., Ltd	Shanghai, China	≥98.00%	C9H17NO5	220.118	79-83-4	CC(C)(CO)C(C(=O)NCCC(=O)O)O
	Nicotinic acid	Macklin Biochemical Co., Ltd	Shanghai, China	≥99.50% (HPLC)	C6H5NO2	124.04	59-67-6	O=C(O)C1=CN=CC=C1
	Niacinamide	Aladdin Reagent Co., Ltd	Shanghai, China	≥99.80%	C6H6N2O	123.055	98-92-0	NC(C1=CN=CC=C1)=O
Heterocyclic amines	Harman	TRC	Toronto, Canada	95.00%	C12H10N2	183.092	486-84-0	CC1=NC=CC2=C1NC3=CC=CC=C23
	DMIP	TRC	Toronto, Canada	98.00%	C8H10N4	163.098	132898-04-5	CC1=CC2=C(N=C1)N=C(N2C)N
	PhIP	TRC	Toronto, Canada	98.00%	C13H12N4	225.113	105650-23-5	CN1C2=C(N=CC(=C2)C3=CC=CC=C3)N=C1N
	IQ[4,5-b]	TRC	Toronto, Canada	98.00%	C11H10N4	199.098	156215-58-6	CN1C2=CC3=CC=CC=C3N=C2N=C1N
	MeIQ	TRC	Toronto, Canada	98.00%	C12H12N4	213.114	77094-11-2	CC1=CC2=C(C=CC=N2)C3=C1N(C(=N3)N)C
	7,8-DiMeIQx	TRC	Toronto, Canada	98.00%	C12H13N5	228.124	92180-79-5	CC1=C(N=C2C(=N1)C=CC3=C2N=C(N3C)N)C
	8-MeIQx	TRC	Toronto, Canada	98.00%	C11H11N5	214.109	77500-04-0	CC1=CN=C2C=CC3=C(C2=N1)N=C(N3C)N
	MeAaC	TRC	Toronto, Canada	98.00%	C12H11N3	198.103	68006-83-7	CC1=CC2=C(NC3=CC=CC=C32)N=C1N
	IQ	TRC	Toronto, Canada	98.00%	C11H10N4	199.098	76180-96-6	CN1C2=C(C3=C(C=C2)N=CC=C3)N=C1N
	IQx	TRC	Toronto, Canada	98.00%	C10H9N5	200.093	108354-47-8	CN1C2=C(C3=NC=CN=C3C=C2)N=C1N
	Phe-p-1	TRC	Toronto, Canada	97.00%	C11H10N2	171.092	33421-40-8	C1=CC=C(C=C1)C2=CN=C(C=C2)N
	AaC	TRC	Toronto, Canada	98.00%	C11H9N3	184.087	26148-68-5	C1=CC=C2C(=C1)C3=C(N2)N=C(C=C3)N
	Norharman	TRC	Toronto, Canada	≥98.00%	C11H8N2	169.076	244-63-3	C1=CC=C2C(=C1)C3=C(N2)C=NC=C3
Internal standard	Hypoxanthine-d3	CIL	USA	2,8,9-D3, OD,98.00%	C5HD3N4O	139.058	NA	C1=NC2=C(N1)C(=O)NC=N2

Table S2

Mass spectrometry conditions		
Mass spectrometer	Scan mode	Source parameters
Q Exactive HF-X	ESI(+) Full MS	Sheath gas flow rate: 50 psi Auxiliary gas flow rate: 1 arb Auxiliary gas heater temperature: 370 ℃ Funnel RF level: 50.0 Spray voltage: +3.6 kV Capillary temperature: 360 ℃
Liquid phase conditions		
Liquid chromatography column	Mobile phase (A/B)	Elution mode
Waters HSS T3 column (2.1 mm × 100 mm, 1.8 μm)	A: Pure water; B: Pure methanol	5% B isocratic elution 20 min 20% B isocratic elution 20 min 35% B isocratic elution 20 min 50% B isocratic elution 20 min 65% B isocratic elution 20 min 80% B isocratic elution 20 min

		95% B isocratic elution 20 min
	A: 0.1% formic acid water; B: 0.1% formic acid methanol	1% B: 0-1.5 min; 1-99% B: 1.5-16.5 min; 99% B: 16.5-20min gradient elution 5% B isocratic elution 20 min 20% B isocratic elution 20 min 35% B isocratic elution 20 min 50% B isocratic elution 20 min 65% B isocratic elution 20 min 80% B isocratic elution 20 min 95% B isocratic elution 20 min 1% B: 0-1.5 min; 1-99% B: 1.5-16.5 min; 99% B: 16.5-20min gradient elution

Table S3: Detailed information on the training set for each model. (Excel)

There is detailed information on the training set for each model (e.g. predicted_values, true_values, all selected descriptors, and original data).

Table S4: Detailed information on the testing set for each model. (Excel)

There is detailed information on the testing set for each model (e.g. predicted_values, true_values, all selected descriptors, and original data).

Table S5: Detailed information on the outliers outside the application domain of each model. (Excel)

There is detailed information on the outliers outside the application domain of each model (e.g. standardized residuals, leverage Value, and measured information).

Table S6: correlation between IE of compounds and the content of organic modifiers. (Excel)

There is a correlation between the IE of compounds and the content of organic modifiers. Correlation parameters were obtained from the correlation between the content (from 5% to 95%) of the isocratic elution mobile phase B (y-axis) and the mass spectral IE (x-axis) of the compound under this condition. Asterisks indicate that the point used for regression under this condition is less than or equal to 2.