

分子の構造活性相関解析のためのニューラルネット ワークシミュレータ： Neco(NEural network simulator for structure-activity COrrrelation of molecules)の開発(4) ペリラルチン類の甘味・苦味分類

田島 澄恵^a, 松本 高利^b, 長嶋 雲兵^{c*}, 細矢 治夫^a, 青山 智夫^d

^a お茶の水女子大学人間文化研究科, 〒 112-8610 文京区大塚 2-1-1

^b 物質工学工業技術研究所, 〒 305-8565 つくば市東 1-1

^c 産業技術融合領域研究所, 〒 305-8562 つくば市東 1-1-4

^d 宮崎大学工学部, 〒 889-2192 宮崎市学園木花台西 1 - 1

*e-mail: umpei@nair.go.jp

(Received: January 7, 2000; Accepted for publication: April 17, 2000; Published on Web: May 23, 2000)

分子構造によって甘味や苦みの性質を示す 22 種類のペリラルチン類の分類を 3 層のパーセプトロン型ニューラルネットをもちいて行った。入力パラメータとして、疎水性パラメータ LogP と分子構造 STERIMOL パラメータ 5 種のパラメータを用いると、非線形分類法の一つである 3 層のパーセプトロン型ニューラルネットで誤認識が全く無しに正しい分類学習が可能となる。入力層ニューロン数と出力層ニューロン数をそれぞれ、6 と 2 としたとき、再構築学習法によって最小化された中間層ニューロン数は 1 となり、最小の中間層ニューロン数を持つ。逆に中間層ニューロン数を 1 と固定し、入力パラメータの数を最小化すると、LogP と分子の幅を示す STERIMOL パラメータ W_r の合計 2 つだけで誤認識のない分類が可能であることがわかった。

STERIMOL パラメータの代わりに、分子構造の共通な骨格上の 6 原子の形式電荷とそれぞれの分子の最高占有軌道 (HOMO) と最低非占有軌道 (LUMO) エネルギーおよび両者の差 (HOMO-LUMO Gap) を選択し、それらと LogP を入力パラメータとして用いた再構築学習の結果も中間層ニューロン数が 1 となった。この場合、甘味・苦味の分類には、LogP に加え、ペリラルチン類の 6 員環の 2 重結合に関わる炭素原子の電子状態が重要である。これは W_r が重要であることに対応している。また LUMO エネルギーが重要であることがわかった。これは、甘味の発現には電子親和性の反応が関わっていることを示唆している。

分子軌道計算から得られる情報のみ、すなわち形式電荷と HOMO, LUMO エネルギーおよび HOMO-LUMO Gap を入力パラメータとした場合、中間層ニューロンの数は 3 となった。

キーワード: Perillartine derivatives, Neural Network, Quantitative Structure Activity Relationship (QSAR), Sweet/Bitter Classification, Charge Distribution, HOMO-LUMO Energy

1 はじめに

ペリラルチンは、植物のシソに含まれるペリラルアルデヒドから合成され、シソ糖と呼ばれる甘味化合物であり、その甘味は非常に強く蔗糖の数千倍とも言われている。ペリラルチンの誘導体 (Figure 1) は、甘味を示すものと苦味を示すものがあり、その分類は定量的構造活性相関の成功例としてよく知られている。[1, 2]

高橋ら [1] は、甘味や苦みの性質を示す 22 種類のペリラルチン誘導体の分類のため、説明変数として疎水性パラメータ $\log P$ (1-オクタノール/水の分配係数の対数值) と分子構造 STERIMOL パラメータ [3-6] 5 種 (Figure 2), 合計 6 種のパラメータ (Table 1) から、Fisher 比の高い 3 変数 ($\log P$, W_u , W_r) を選択した。それらを用いた KNN 法 [7] では、甘味・苦味の分類において 25 番と 51 番の分子が誤認識される。しかし線形学習機械法 [8, 9] を用いると、誤認識は起こらない。[1]

我々は、構造活性相関のためのニューラルネットワークシミュレータ Neco を開発してきた [10-13]。Neco で採用されているパーセプトロン型のニューラルネットワークは非線形分類に有効であることがよく知られており、本論文では、非線形分類法の一つである三層パーセプトロン型ニューラルネットワークと再構築学習法 [14] を用いて、ペリラルチン誘導体の構造活性相関解析を行った結果を報告する。

2 計算結果

2.1 疎水性パラメータ $\log P$ と構造パラメータ (STERIMOL) の場合

Table 1 に示す疎水性パラメータ $\log P$ と構造パラメータ (STERIMOL) [1, 2] を用いて、入力層 6 ニューロン、出力層ニューロン 2 とし、中間層ニューロン数を再構築学習法 [14] で最小化した。このとき最小化された中間層ニューロンの数は 1 となった。

入力層ニューロン数 6、中間層 1、出力層 2 (6,1,2) としたときのバックプロパゲーションによる学習の結果を Table 2 に示す。学習条件は以下の通りである。ネットワークの初期値として全結合重みを 1.0 とした。学習誤差のしきい値は、1% 程度の誤差を容認するとして 0.0011 とした。学習の反復回数の上限を 100000 回としたが、この場合収束には 1500 回の反復が必要であった。ただし、異なる初期値を用いて生成した乱数など、他の初期値を用いてもほぼ 2000 回以内で収束した。(1,0) が甘味を示すもので (0,1) が苦味を示すものとして学習を行った。Table 2 に示されるように、誤認識はない。KNN で誤認識される [1, 2] 25 と 51 の誤差は苦味を示すもののうちで大きいとはいえ、1% 程度の誤差であり、他のものと大きな違いはない。言い換え

Table 1. Log P and STERIMOL parameters for Perillartine derivatives

Mol. ¹	sweet ² /bitter	log P	L	Wl	Wu	Wr	Wd
1	1	2.58	8.52	3.13	2.85	3.42	1.99
4	1	0.87	5.10	3.13	1.91	2.94	1.90
8	1	2.28	8.69	3.19	2.84	3.42	1.99
28	1	1.10	9.36	3.14	2.94	3.41	1.98
29	1	1.40	9.36	3.14	3.26	3.56	2.10
34	1	1.48	6.06	3.09	2.08	3.01	1.71
37	1	1.10	8.87	3.30	2.63	3.07	2.52
42	1	1.48	6.29	3.09	2.63	3.07	2.52
43	1	0.78	7.10	3.09	1.91	3.41	1.91
44	1	0.80	9.01	3.09	2.20	3.41	2.02
45	1	1.10	9.01	3.08	2.52	3.43	2.53
14	0	-0.10	0.67	3.33	4.11	3.56	2.10
15	0	-0.10	9.36	3.04	3.76	3.62	2.22
16	0	-0.92	9.37	3.14	3.56	3.56	2.10
22	0	-0.72	5.51	3.05	2.53	3.41	1.97
23	0	-0.72	6.15	3.16	2.67	3.01	1.72
25	0	0.34	6.05	3.25	2.62	3.43	2.03
33	0	0.72	10.67	3.51	4.08	3.63	2.22
48	0	1.40	7.98	3.12	3.42	5.96	2.00
49	0	0.80	7.68	3.09	2.32	5.84	1.96
50	0	1.10	7.68	3.09	2.43	5.89	2.57
51	0	1.90	5.88	2.72	2.95	3.92	3.85

1: Details of for Perillartine derivatives and STERIMOL parameters are depicted in Figures 1, 2.

2: Sweet and bitter are 1 and 0, respectively.

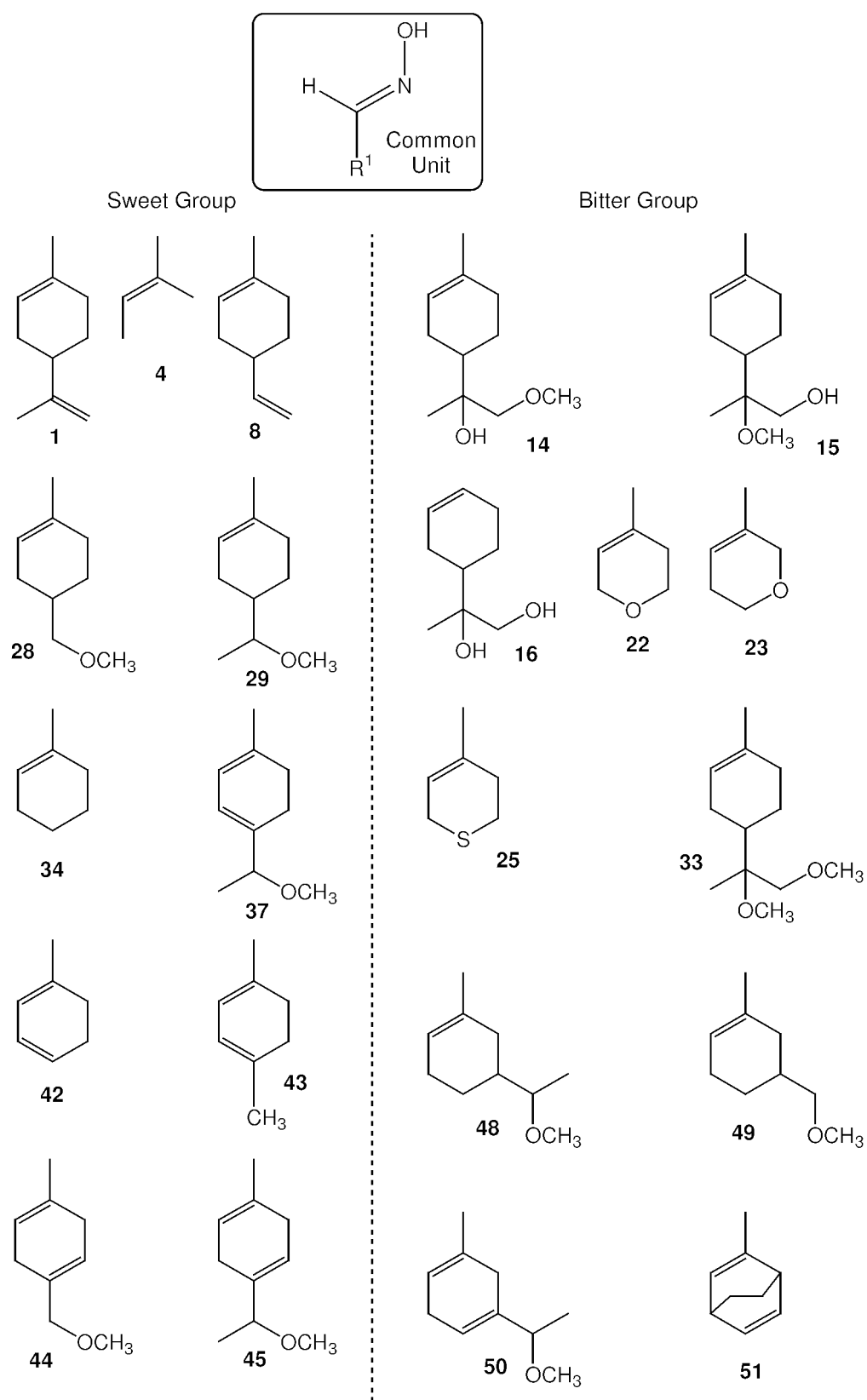


Figure 1. Sweet/bitter structures of Perillartine derivatives

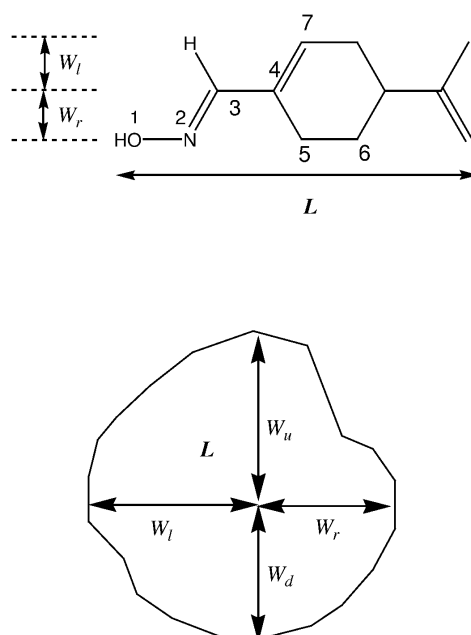


Figure 2. Details of STERIMOL parameters[1, 2] and numbering of atoms

ると 25 と 51 がとりわけ他と異なっているという結果ではない。また、中間層ニューロンの数が 1 であることは、線形学習機械法で誤認識が無いこと [1] に対応している。

ネットワーク構造を (6,1,2) としたときのバックプロパゲーションによる学習結果のネットワークの重みを Table 3 に示した。左側が入力層と中間層の重みであり、右側が中間層と出力層の重みである。下段は高橋ら [1] による Fisher 比 ($\times 10^{-3}$) である。入力層と中間層ニューロン間の係数の絶対値は、それぞれのニューロンからの信号の重要性を反映しており、この傾向は、Fisher 比によるパラメータのランクの傾向と L と W_d の順序が逆転しているとはいえ、ほぼ一致する。つまり従来の解釈と同じように、一番重要なパラメータは疎水性パラメータ LogP であり、次に重要なのは分子の厚み W_u と幅 W_r である。より細かく見るならば、 logP が大きく、分子長 L が長かつ分子の厚み W_u や幅 W_r が短いものが甘味が強くなることを示している。

次に、中間層ニューロン数を 1 に固定して、同様に学習誤差 0.0011 という条件で、誤認識をおこさない入力パラメータ数最小の組を求めると、 LogP と W_r の 2 つのパラメータからなる組を得た。

ネットワーク構造を (2,1,2) としたときのバックプロパゲーションによる学習結果のネットワークの重みを Table 4 に示す。左側が入力層と中間層の重みであり、右側が中間層と出力層の重みである。これは、ペリラルチン誘導体の甘味・苦味の三層パーセプトロン型ニューラルネットを用いた非線形分類では LogP と W_r が本質的なパラメータであることを示している。また、 LogP より W_r の方が重要であることを示唆している。

高橋らが採用した Fisher 比の大きな (LogP , W_u , W_r) の 3 つのパラメータを用いた場合、学習誤差 0.0011 とし 100000 回以内の学習で誤認識が無くなるためには中間層ニューロン数を 4

Table 2. Results of learning by three layer perceptron (6,1,2)
with Log P and STERIMOL parameters

sweet				bitter			
No.	Item	1	2	No.	Item	1	2
1	1	0.997514	0.002486	12	14	0.003820	0.996180
2	4	0.995524	0.004475	13	15	0.003811	0.996189
3	9	0.997499	0.002501	14	16	0.003624	0.996376
4	28	0.996097	0.003903	15	22	0.003691	0.996309
5	29	0.993824	0.006176	16	23	0.004068	0.995932
6	34	0.997448	0.002552	17	25	0.009133	0.990867
7	37	0.995920	0.004080	18	33	0.006746	0.993254
8	42	0.990937	0.009062	19	48	0.003799	0.996202
9	43	0.996683	0.003317	20	49	0.005344	0.994656
10	44	0.997175	0.002825	21	50	0.004260	0.995740
11	45	0.995995	0.004004	22	51	0.007765	0.992235

Input data : Table 1

Out put data:1, 0 for sweet and 0, 1 for bitter

Table 3. Weight of network in the three layer perceptron (6,1,2)
with Log P and STERIMOL parameters

	layer1(row) and layer2(column)					layer2(column) and 3(row)		
	log P	L	Wl	Wu	Wr	Wd	1	2
1	16.37	7.63	-1.76	-10.35	-9.39	-6.38	11.62	-11.62
FR*	78.4	0.0174	0.059	60.1	49.5	9.14		

*: Fishers' ratio [1] ($\times 10^{-3}$)

Table 4. Weight of network in the three layer perceptron (2,1,2)
with Log P and STERIMOL parameters

	layer1(row) and 2(column)		layer 2(column) and 3(row)	
	log P	Wr	1	2
1	-30.23	74.82	-19.94	19.94

以上にすることが必要であった。4以下では、100000回以内の学習回数では収束せず、中間層ニューロン数1、2、3いずれの場合でも51を誤認識する。

ネットワーク構造を(3,4,2)としたときのバックプロパゲーションによる学習結果のネットワークの重みをTable 5に示す。この結果では、Fisher比の順とは異なり、Wr, LogP, Wuの順で重要となる。Wuはすべての重みがLogPとWrに比べ小さい。この結果は、先に示したようにLogPとWrがペリラルチンの甘味・苦味分類に本質的なパラメータであることを強く示唆している。

2.2 疎水性パラメータlog Pと共通骨格の形式電荷および軌道エネルギーの場合

前節で用いたSTERIMOLパラメータが量子化学的意味づけが曖昧であることのため、手軽に計算可能でかつ量子化学的意味づけが比較的明確な共通骨格部分(Figure 2の番号を打った7原子)の形式電荷とHOMO, LUMOエネルギーおよびHOMO-LUMO Gapを入力データとして学習を行った。

これは、ペリラルチン誘導体の甘味・苦味活性がそれらの共通骨格部分の電荷分布の差異に反映すると考えられること、および、HOMO, LUMOエネルギーとその差は、分子全体に広がる情報を含んでいること、活性発現の初期相互作用が、電子供与的か電子吸引的かラジカル的かの傾向を見ることができると考えられるためである。

これらの値の計算はMOPAC93/AM1[15]を用いて分子構造最適化を行い、それぞれの値を求めた。入力データをTable 6に示す。表作成の都合上、有効桁数を3桁で示したが、実際の計算は有効数字6桁を入力して行った。

この場合も再構築学習法によって最小化された中間層ニューロン数は1となった。ネットワーク構造を(11,1,2)としたときのバックプロパゲーションによる学習結果をTable 7に示す。ここでも、ネットワークの初期値として全結合重みを1.0とした。学習条件は先の場合と同様、学習誤差のしきい値は0.0011である。収束には4954回の反復を要した。

先に説明したSTERIMOLパラメータを用いた場合の1500回の反復に比べ収束が遅くなっている。この収束が遅くなる傾向は、乱数など異なる初期値を用いてもほぼ同様である。(1,0)が甘味を示すもので(0,1)が苦味を示すものとして学習を行った。最小化された中間層ニューロン数は1である。ネットワーク構造(11,1,2)としバックプロパゲーションで学習させたニューラルネットワークの分類結果をTable 7に示す。ここでも誤認識はない。STERIMOLパラメータを用いた場合にKNNで誤認識される25と51の誤差は、苦味を示すもののうちで大きくはなく、むしろ小さくなっている。この場合は、28と48の誤差が大きくなっているが、その値は1%以下である。また、STERIMOLパラメータを用いた場合と同様、中間層ニューロンの数が1であることは、これらのパラメータを用いても線形学習機械法で誤認識無く分類ができる可能性を示している。

ネットワーク構造(11,1,2)としバックプロパゲーションで学習させたニューラルネットワークの重みをTable 8に示した。この場合も、入力層と中間層ニューロン間の係数の絶対値は、それぞれのニューロンからの信号の重要性を反映している。

ここでも一番重要なパラメータは、疎水性パラメータLogPであり、次に重要なのは、C3からC7の炭素の電荷である。今回用いたペリラルチン誘導体において、大きな電荷を持つ末端の酸素の部位は、長距離のクーロン相互作用を考慮すると分子認識上のフラグとして重要であ

Table 5. Weight of network in the three layer perceptron (3,4,2) with Log P and STERIMOL parameters

	layer1(row) and 2(column)			layer 2(column) and 3(row)	
	log P	Wu	Wr	1	2
1	10.81	-1.96	-24.90	7.31	-7.31
2	10.53	-0.80	-27.22	8.23	-8.23
3	22.71	-3.37	-17.90	7.54	-7.54
4	4.33	0.71	-29.97	8.84	-8.84

Table 6. Log P, Formal charge, HOMO, LUMO energy, and HOMO-LUMO Gap of Perillartine derivatives

No.	Item	LogP	Formal Charge							orbital energy		
			O1	N2	C3	C4	C5	C6	C7	HOMO	LUMO	Gap
		1	2	3	4	5	6	7	8	9	10	11
1	1	2.58	-0.26	-0.06	-0.10	-0.08	-0.14	-0.13	-0.12	-9.17	0.21	9.38
2	4	0.87	-0.26	-0.06	-0.10	-0.08	-0.14	-0.19	-0.18	-9.19	0.19	9.38
3	8	2.28	-0.26	-0.06	-0.10	-0.07	-0.14	-0.13	-0.12	-9.14	0.24	9.38
4	28	1.10	-0.26	-0.06	-0.10	-0.08	-0.13	-0.13	-0.12	-9.16	0.22	9.38
5	29	1.40	-0.26	-0.06	-0.10	-0.08	-0.13	-0.13	-0.12	-9.12	0.26	9.38
6	34	1.48	-0.26	-0.06	-0.10	-0.08	-0.14	-0.13	-0.12	-9.16	0.23	9.39
7	37	1.10	-0.25	-0.06	-0.10	-0.06	-0.11	-0.12	-0.12	-8.71	-0.40	8.31
8	42	1.48	-0.26	-0.06	-0.10	-0.07	-0.11	-0.15	-0.12	-8.66	-0.24	8.42
9	43	0.78	-0.26	-0.06	-0.10	-0.08	-0.10	-0.15	-0.12	-8.52	-0.24	8.28
10	44	0.80	-0.26	-0.06	-0.10	-0.07	-0.13	-0.09	-0.09	-9.05	0.12	9.17
11	45	1.10	-0.26	-0.06	-0.10	-0.07	-0.14	-0.12	-0.12	-9.16	0.21	9.38
12	14	-0.10	-0.26	-0.06	-0.10	-0.09	-0.13	-0.14	-0.12	-9.26	0.13	9.39
13	15	-0.10	-0.26	-0.06	-0.10	-0.07	-0.14	-0.13	-0.12	-9.28	0.12	9.39
14	16	-0.92	-0.26	-0.06	-0.09	-0.07	-0.12	-0.12	-0.12	-8.95	0.41	9.36
15	22	-0.72	-0.25	-0.05	-0.10	-0.06	-0.18	0.00	-0.16	-9.38	-0.01	9.37
16	23	-0.72	-0.25	-0.06	-0.10	-0.11	-0.13	-0.17	0.01	-9.38	-0.00	9.38
17	25	0.34	-0.25	-0.05	-0.11	-0.07	-0.14	-0.25	-0.12	-8.55	-0.01	8.55
18	33	0.72	-0.26	-0.06	-0.09	-0.07	-0.12	-0.12	-0.15	-8.89	0.47	9.36
19	48	1.40	-0.26	-0.06	-0.10	-0.07	-0.14	-0.13	-0.12	-9.18	0.23	9.41
20	49	0.80	-0.26	-0.06	-0.10	-0.07	-0.15	-0.13	-0.12	-9.16	0.25	9.41
21	50	1.10	-0.26	-0.06	-0.10	-0.08	-0.13	-0.11	-0.08	-8.97	0.17	9.14
22	51	1.90	-0.26	-0.06	-0.09	-0.08	-0.13	-0.06	-0.05	-9.19	0.15	9.34

Data of sweet/bitter are not shown in this table because these are the same as in Table 1.

Table 7. Results of learning by three layer perceptron (6,1,2)
with Log P, formal charge, and etc.

		sweet		bitter			
No.	Item	1	2	No.	Item	1	2
1	1	0.998462	0.001538	12	14	0.006965	0.993035
2	4	0.998271	0.001729	13	15	0.002789	0.997211
3	8	0.998461	0.001539	14	16	0.002788	0.997212
4	28	0.990624	0.009376	15	22	0.002788	0.997212
5	29	0.997931	0.002069	16	23	0.002788	0.997212
6	34	0.998155	0.001845	17	25	0.002789	0.997211
7	37	0.998462	0.001538	18	33	0.002788	0.997212
8	42	0.998462	0.001538	19	48	0.013562	0.986438
9	43	0.998462	0.001538	20	49	0.002794	0.997206
10	44	0.992449	0.007551	21	50	0.004430	0.995570
11	45	0.991941	0.008059	22	51	0.002789	0.997211

Input data : Table 6

Output data: 1, 0 for sweet and 0, 1 for bitter

ると考えられる。しかしこの結果をみると酸素の電荷の寄与は、甘味・苦味の分類においては小さい。これは、酸素と窒素の部位がペリラルチン誘導体が活性サイトにアプローチする際には重要であるが、生体内での甘味・苦味の活性には、それらの部位よりむしろ分子骨格の炭素原子が形成する電子状態が重要であることを示唆している。

Table 8. Weight of network in the three layer perceptron (11,1,2)
with Log P and Formal charge etc.

Weight matrix between layer1(row) and layer2(column)											
LogP	Formal Charge							orbital energy			
	O1	N2	C3	C4	C5	C6	C7	HOMO	LUMO	Gap	
1	61.56	-8.90	15.21	-49.97	-39.83	49.39	44.83	-38.86	-9.30	-34.69	-14.40
Weight matrix between layer 2(column) and 3(row)											
	1	2									
1	12.35	-12.35									

また、HOMO, LUMO 軌道エネルギーおよびその差は分子全体を見るためのパラメータとして導入したパラメータであるが、重みを見ると LUMO のエネルギーの寄与が大きい。これは、基質からの電子受容性が小さいほど甘味を感じるということを示している。またこれらの LUMO は、分子骨格の炭素上の軌道から構成されていることから、電荷の解析からも示唆されるように、分子骨格上の炭素原子が形成する電子状態が、ペリラルチン誘導体の甘味・苦味の活性に重要であるといえる。

STERIMOL パラメータは、分子の最安定構造の電子構造のみならず、様々に複雑な効果が繰り込まれたパラメータであると考えられるが、本計算結果からも、LUMO のような求電子的効

果も含まれていると考えることができる。

次に、単純な分子軌道計算から得られるデータのみで分類が可能かどうかをみるために、疎水性パラメータ LogP を除いて電荷と軌道エネルギーならびにその差のみを入力として、再構築学習法によって中間層ニューロン数を最小化するとそれは 3 となった。ネットワーク構造 (10,3,2) としバックプロパゲーションで学習させたニューラルネットワークの重みを Table 9 にしめす。学習条件はこれまでのものと同様である。この場合、出力層に大きな重みを持つ中間層ニューロン 1 と入力パラメータの重みをみると、酸素の寄与が大きいが窒素は大きな寄与をもたないことがわかる。

先にも述べたように酸素の大きな負電荷は、長距離のクーロン相互作用を考慮すると分子認識上のフラグとして重要であると考えられる。そのため LogP は、遠距離の相互作用と分子全体の電子状態の効果を含んでいると考えられる。また C7 炭素の重みが小さいことは、LogP と STERIMOL パラメータを用いた場合の STERIMOL パラメータ W_r が重要であるということによく対応している。

Table 9. Weight of network in the three layer perceptron (10,3,2) with Formal charge and etc.

		Weight matrix between layer1(row) and layer2(column)										
		Formal Charge					orbital energy					
		O1	N2	C3	C4	C5	C6	C7	HOMO	LUMO	Gap	
1		-21.73	0.16	-46.64	-15.99	26.41	25.12	1.24	9.96	-14.38	-11.87	
2		-29.84	-18.19	-53.48	-15.89	30.99	17.00	-4.80	38.33	9.92	-9.75	
3		-56.14	8.87	-127.08	-28.30	11.43	19.72	32.38	2.79	15.07	26.03	
		Weight matrix between layer 2(row) and 3(column)										
		1	2									
1		16.94	-16.94									
2		11.43	-11.43									
3		11.56	-11.56									

3 まとめ

疎水性パラメータと STERIMOL パラメータを入力データとして用いて、3 層のパーセプトロン型ニューラルネットの学習を行った結果、誤認識が全く無しに正しい分類が可能であることがわかった。また入力層ニューロン数と出力層ニューロン数をそれぞれ 6 と 2 としたとき、再構築学習法によって最適化された中間層ニューロン数は 1 となり、最小の中間層ニューロン数を持つことがわかった。これは線形学習機械法で誤認識が無いこと [1] に対応している。このときのニューラルネットの結合重みの絶対値の大きな 3 つの入力パラメータは、Fisher 係数のそれと一致した。

また中間層ニューロン数を 1 に固定し、同様の学習条件で入力パラメータ数を最小化すると、LogP と W_r の 2 つのパラメータでも誤認識無しに正しい分類が可能であることがわかった。

STERIMOL パラメータの代わりに、LogP と分子構造の共通な骨格上の 6 原子の形式電荷とそれぞれの分子の HOMO と LUMO エネルギーおよび HOMO-LUMO Gap を入力データとして用いた学習の結果も最適な中間層ニューロン数が 1 となった。この場合、甘味・苦味の分類には、ペリラルチン誘導体の 6 員環の 2 重結合に関わる炭素原子の電荷が重要で、酸素や窒素原子の電荷は大きな影響を与えないことがわかった。これは、STERIMOL パラメータの W_r が重

要であるということに対応している。また LUMO エネルギーが重要であることが示唆された。これは、ペリラルチン誘導体の電子受容性が甘味・苦味活性に重要であることを示唆している。

形式電荷と HOMO, LUMO エネルギーおよび HOMO-LUMO Gap のみで学習を行った場合、最小の中間層ニューロン数は 3 となった。この場合、酸素の電荷が重要であることが示され、LogP が遠距離の相互作用と分子全体の電子状態の効果を含んでいることが示唆された。

貴重なご助言をいただいた物質研の田辺和俊首席研究官に深く感謝する。また本論文に関して建設的なご意見をいただいた、審査員の方々に深く感謝する。

参考文献

- [1] 宮下芳勝, 佐々木慎一, ケモメトリックス 化学パターン認識と多変量解析, 共立出版 (1995), p.143.
Takahashi, Y., Miyashita, Y., Tanaka, Y., Abe, H., Sasaki, S., *J. Med. Chem.*, **25**, 1245 (1982).
- [2] 青山義弘, 吉村忠与志, 山田明文, 化学とソフトウェア, **20**, 59 (1998).
- [3] Verloop, A., Hoogenstraaten, W., Tipker, J., *Drug Design*, **7**, 165 (1976).
- [4] Iwamura, H., *J. Med. Chem.*, **23**, 308 (1980).
- [5] Iwamura, H., *J. Med. Chem.*, **24**, 572 (1981).
- [6] Iwamura, H., Fujita, T., Koyama, S., Koshimizu, K., Kumazawa, Z., *Phytochemistry*, **19**, 1309 (1980).
- [7] Kowalski, B.R., Bender, C.F., *Anal. Chem.*, **44**, 1405 (1972).
- [8] Jurs, P.C., Kowalski, B.R., Isenhour, T.L., *Anal. Chem.*, **41**, 21 (1959).
- [9] Stuper, A.J., Jurs, P.C., *J. Am. Chem. Soc.*, **97**, 182 (1975).
- [10] 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **2**, 76 (1994).
- [11] 井須芳美, 長嶋雲兵, 細矢治夫, 大島茂, 坂本曜子, 青山智夫, *J. Chem. Software*, **3**, 1 (1996).
- [12] 藤谷康子, 小野寺光永, 井須芳美, 長嶋雲兵, 細矢治夫, 青山智夫, *J. Chem. Software*, **4**, 19 (1998).
- [13] Isu, Y., Nagashima, U., Aoyama, T., Hosoya, H., *J. Chem. Info. Comp. Sci.*, **36**, 286 (1996).
- [14] Aoyama, T., Ichikawa, H., *Chem. Pharm. Bull.*, **39**, 1222 (1991).
- [15] Deware, M. J. S., Zoebitsch, E. G., Healy, E. F., Stewart, J. J. P., *J. Am. Chem. Soc.*, **107**, 3902 (1985).
MOPAC93.00, Stewart, J. J. P., Fujitsu Ltd., Tokyo, Japan, 1993. Available from Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN, USA.

Development of a Neural Network Simulator for Structure-Activity Correlation of Molecules: Neco (4) – Sweet / Bitter Classification in Perillartine Derivatives –

Sumie TAJIMA^a, Takatoshi MATSUMOTO^b, Umpei NAGASHIMA^{c*}, Haruo HOSOYA^a and Tomoo AOYAMA^d

^aDepartment of Human Culture and Sciences, Graduate School of Ochanomizu University
2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan

^bNational Institute of Materials and Chemical Research
1-1 Higashi, Tsukuba, Ibaraki 305-8565, Japan

^cNational Institute for Advanced Interdisciplinary Research
1-1-4 Higashi, Tsukuba, Ibaraki 305-8562, Japan

^dFaculty of Technology, Miyazaki University
Gakuenkhanadai Nishi, Miyazaki 889-2192, Japan

**e-mail: umpei@nair.go.jp*

The relationships between molecular structure and taste quality: sweet or bitter, or several perillartine derivatives were examined using a perceptron type neural network simulator for structure-activity correlation of molecules: Neco with reconstruction of weight matrix method. The reconstruction of weight matrix method was used to optimize the number of neurons in hidden layer.

In the case of using six parameters: hydrophobic(log P) and the STERIMOL(L, Wl, Wu, Wr, and Wd) parameters as inputs, the number of neurons in hidden layer is minimized to one by the reconstruction learning method. Even in this case, there is no misclassified compound. The prediction rate by leave-one-out procedure was also 100%. The most important three parameters were the same as predicted by Fisher ratio.

The number of input parameters was minimized by holding the number of neurons in hidden layer to one. Two parameters, namely Log P and Wr were found to describe the sweet/bitter activity of perillartine derivatives.

Instead of STERIMOL parameters, atomic charges of common molecular skeleton, HOMO and LUMO energies, and HOMO-LUMO energy difference were selected as input data. MOPAC93/AM1 was used to evaluate these parameters. The optimum number of neurons in the hidden layer was also one. Atomic charges and LUMO energy are also important for sweet/bitter classification. This suggests that electronic structure around common molecular skeleton and electron elimination reaction are essential to sweet/ bitter activity of perillartine derivatives.

Keywords: Perillartine derivatives, Neural Network, Quantitative Structure Activity Relationship (QSAR), Sweet/Bitter Classification, Charge Distribution, HOMO-LUMO Energy