

ニューラルネットワークによる発癌物質の構造活性相関

7H-1

井須 芳美, 長嶋 雲兵, 細矢 治夫
お茶の水女子大学理学部情報科学科

1 はじめに

ニューラルネットワークは、その動作の特徴として非線形的な動作をすることが知られている。また入力と出力の間にある相関をあらかじめ解析しておくことが困難な事象に適用されている。ニューラルネットワークのもつこれらの特徴は分子の構造とその活性の間にあるような非線形的な相関を扱うのに適している。本研究では多環式芳香族炭化水素の¹³C-NMRケミカルシフトとその発癌性における特徴抽出を青山らの再構築学習法 [1] を用いて行なった。

多環式芳香族炭化水素 (PAH) の発癌性については従来から多くの研究がなされてきた [2]。従来の研究では PAH における L 領域、K 領域、bay 領域 (図 1) と呼ばれる部分の反応性が重要視されてきたため、本研究では bay 領域を含むような多環式芳香族炭化水素 11 種類について共通する構造 (bay 領域を含む) に着目し、その¹³C-NMRケミカルシフトと発癌性に対する相関とその特徴抽出を行なうことを試みた。

2 特徴抽出の方法とデータ

本研究で用いたニューラルネットワークモデルは 3 層の階層型ニューラルネットワークである。学習には青山らが提唱した再構築学習法を用いた。再構築学習法とはバックプロパゲーションを用いた学習過程の他に消去過程を取り入れたものである。消去過程では結合荷重行列を部分的に消去する。この消去過程を取り入れることにより不必要なニューロンの孤立が起こる。従って得られた結合荷重行列を見れば入力データのどのパラメータが出力に関係しているのかがわかる。

本研究で特徴抽出を行なう対象とした化合物は同じような構造を有する 11 種類の多環式芳香族炭化水素 (表 1) である。この内、発癌性の化合物が 7 種類、非発癌性の化合物が 4 種類である。入力データはこれら 11 種類の化合物の構造が共通している部分の¹³C-NMRケミカルシフト値で、出力はその発癌性である。図 2 に用いた化合物の 1 つ Benz[a]anthracene の構造を示す。図 2 において太線を引いた部分の構造が表 1 に示した化合物に共通する構造である。

表 1: 対象とした化合物名とその発癌性

図 1: K 領域, L 領域, bay 領域

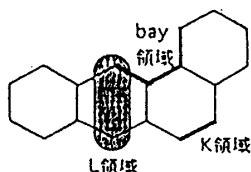
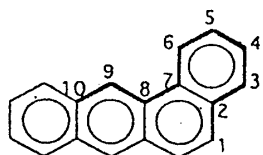


図 2: benz[a]anthracene の構造



化合物番号	化合物名	発癌性
1	Benz[a]anthracene	+
2	Dibenz[a,h]anthracene	+
3	Benzo[a]pyrene	+
4	7-Methylbenz[a]anthracene	+
5	7,12-Dimethylbenz[a]anthracene	+
6	Phenanthro[4,5-bcd]thiophene	+
7	Benzo[e]pyrene	+
8	Triphenylene	-
9	Dibenz[a,c]anthracene	-
10	1-Methylbenz[a]anthracene	-
11	Benzo[b]chrysene	-

注: +: 発癌性有, -: 発癌性無

3 結果

再構築学習法を用いた学習によって得られた結合荷重行列(表 2)によると、 $C1$ と $C10$ のケミカルシフト値が発癌性の発現に最も重要であり、次いで $C4, C8, C9$ が重要な相関を持っていることが示唆された。そこで次に $C1$ と $C10$ のケミカルシフト値を入力データとして通常のバックプロパゲーションを用いて学習を行なった後、 $C1, C10$ をそれぞれとり得る範囲で変化させ、それぞれの値に対する発癌性をニューラルネットワークに予測させることによりニューラルネットワークの分類の様子を調べた。結果を図 3に示す。非発癌性物質の $C10$ の $^{13}C-NMR$ ケミカルシフトが127.5周辺にあることがわかる。また、6番目の分子が特異的であることがわかる。そこで6番目の分子のデータを除いて学習させた結果を図 4に示すが、あまり大きな変化は見られない。

表 2: 再構築学習法により得られた結合荷重行列

Weight mtx. between layer1(row) and layer2(column)		1	2	3	4	5	6	7	8	9	10
1	0.00	0.01	0.01	-0.00	0.01	0.01	0.01	-0.00	4.28	0.01	
2	0.00	0.01	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.01	
3	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00	
4	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	
5	<u>-7.21</u>	-0.01	2.75	-0.01	-1.86	-3.52	-0.01	4.62	-0.01	0.01	
6	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.01	0.01	0.01	
7	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	
8	0.01	0.00	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.00	
9	3.42	0.00	-0.01	5.27	-0.01	0.00	0.01	-0.00	-1.79	<u>-10.18</u>	
10	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	

図 3: $C10, C1$ を入力データとした場合の分類結果 (1)

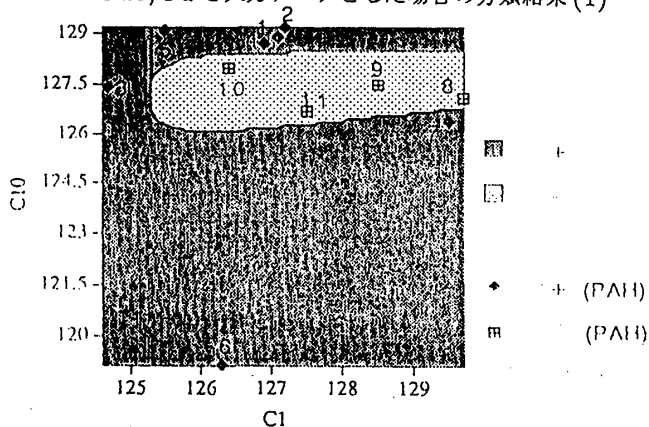
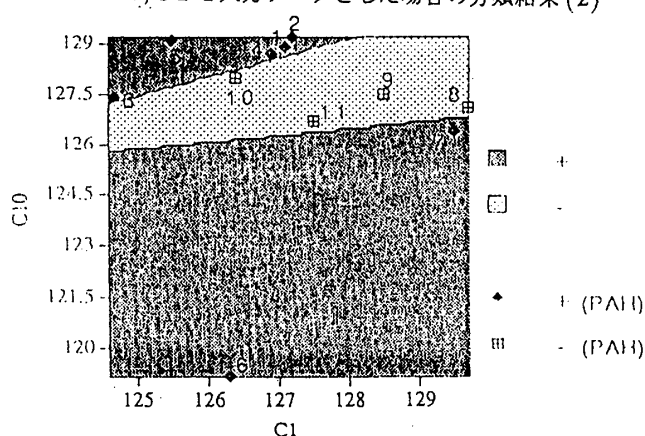


図 4: $C10, C1$ を入力データとした場合の分類結果 (2)



4 考察

多環式芳香族炭化水素の共通する構造に着目した場合、 $C10, C1$ のケミカルシフト値が発癌性の発現に最も重要であることが再構築学習法により示唆された。また、 $C10, C1$ のみならず $C4, C8, C9$ も発癌性に対して相関をもつことが示唆された。 $C8, C9$ は従来の研究で発癌性に対して重要視されているbay領域に対応する。また、 $C10, C9$ の重要性はL領域説に対応するものである。 $C4$ の重要性については従来の研究では言われていないものであるが、発癌性に何らかの相関があるものと思われる。ただし現在手に入るデータが少ないためこの結果の信頼性については少々難ありと思われる。また、なぜこれらのケミカルシフトが重要であるかは量子化学的には明らかではない。 $^{13}C-NMR$ と種々の特質に対する発癌性のシステマティックな実験および解析が必要である。

参考文献

- [1] T.Aoyama and H.Ichikawa, Chem. Pharm. Bull., 39,1222(1991)
- [2] 米澤 貞次郎, 永田 親義, 加藤 博史, 今村 隆, 諸熊 奎治, 量子化学入門 (下), 第13章, 化学同人, (1964)