

1. はじめに

環境ホルモンと考えられている化学物質の構造は様々で、あらゆる化学物質にその疑いが指摘されている。現在までに合成されたすべての化学物質の影響を、実験により測定していくことはきわめて困難である。そこで、私たちはニューラルネットワークという手法を用い、化学物質の種々のプロパティーとその毒性との構造活性相関の解析を行い、毒性値を容易に予測することを試みてきた。これまでに化学物質のプロパティーとして①ligand のフラグメント情報 に加え新たに、②ligand の IP, EA をはじめとする物理量や原子間距離などの構造データ ③ligand と receptor との相互作用情報 ④ligand の活性型構造に関する情報を取り込んだ新規 QSAR 解析を行った。[1]

更なる高精度な予測のためには、化学物質のプロパティーとして環境ホルモンの毒性と親密に影響していると考えられている logP, 溶解度, pKa といった各種物性値を新たに取り入れることが有効であると考える。なぜなら、化学物質がその毒性発現の作用点となる細胞内のレセプターに到達する為には、細胞外の親水性相から親油性相である細胞膜を透過しなければならない。すなわち化学物質は親水性、親油性の両方の性質を適度に持っていることが必要である。すなわちオクタノール/水の分配係数である logP は化学物質の親油性を表す重要な役割を持つ因子となる。また溶解度や pKa もまた血液への溶解性という点から毒性評価において重要な因子のひとつである。

しかしながら、我々の対象とした化学物質に対

して、実験によりこれらの物性値すべてが測定されているわけではない。また、数多くの予測手法が提案されているものの、その多くが任意性のある分子軌道計算を行う必要があり、手間もかかるものである。

そこで、この環境ホルモンの毒性と相関が深いと考えられる、logP, 溶解度, pKa という各物性値を簡単にかつ高精度で予測することを試みた。本発表ではこれらの各種物性値 (logP, 溶解度, pKa) を化学物質の構造式のみから簡単にかつ高精度で予測する手法について発表する。

これらは医薬品などの開発においても重要な物性値であり、新規化学物質の合成前にある程度これらの値を予測できることのメリットはきわめて大きいことから、logP, 溶解度, pKa の各種物性値を予測する手法の開発は、それ自身を予測することにも大きな意味があると考える。

2. 方法

【階層型ニューラルネットワーク】

データ解析手法には、非線形解析手法であるニューラルネットワーク法を用いた。実際のプログラムには階層型ニューラルネットワークシミュレーター PSDD(Perceptron-type Neural Network Simulator for Drug Design)を用いた。[2]

【インプットディスクリプター】

データ解析のためのディスクリプター (入力ノード) には分子の 2 次元構造式から簡単に定義できる分子のフラグメント情報を logP では 127 個、溶解度では 57 個、pKa では 55 個用いた。また、対象とした化学物質は各物性値それぞれ約 600 物質である。

3. 結果

【fitting 結果】

logP の fitting 結果について、推算値と測定値の相関関係を Fig. 1 に、推算値の誤差の分布を Fig. 2 に示した。大きな誤差を与えるような化合物はなく、全体の 95%以上の化合物が 3σ の範囲に入った。Table 1 には logP, 溶解度, pKa の fitting 結果の統計値を示した。いずれの物性値も相関係数(r^2)は 0.99 程度であり、誤差の標準偏差(s)も 0.15 以下であった。我々の定義したフラグメントとニューラルネットワーク法を組み合わせることで、きわめて良い相関関係を構築することができた。

【他の予測手法との比較】

logP の fitting 結果について、中馬らの予測手法 [3]との比較を行った。彼らは 166 個の化合物を 3 つのグループに分け、それぞれのグループごとに線形解析を行っている。そこで我々のデータセットと共に 161 個の化学物質を選び出し、比較した結果を Table 2 に示した。Group1, Group2 ではほぼ同等であるが、Group3 では明らかに我々の手法が良い結果を得た。また彼らは分子のディスクリプターとして SCRF の MO 計算結果を用いているため、推算を行う為に大変手間がかかる。一方我々の手法は分子の 2 次元構造式のみから容易に推算が可能である。

Table 1 log P, 溶解度、pKa の fitting 結果

	n	r^2	s	近似曲線
logP	627	0.988	0.139	$y = 0.989x + 0.016$
溶解度	331	0.993	0.114	$y = 0.998x - 0.002$
pKa	310	0.998	0.150	$y = 0.997x + 0.029$

Table 2 logP 推算結果の比較

n	Our method		Chuman's method		
	r^2	s	r^2	s	
Group 1	72	0.984	0.152	0.982	0.160
Group 2	38	0.954	0.228	0.962	0.208
Group 3	51	0.967	0.198	0.908	0.330
all	161	0.982	0.187	0.971	0.235

Table 3 log P, 溶解度、pKa の prediction 結果

	n	r^2	s	近似曲線
logP	202	0.911	0.305	$y = 0.959x + 0.133$
溶解度	58	0.903	0.375	$y = 0.950x - 0.180$
pKa	67	0.959	0.704	$y = 1.049x - 0.002$

【prediction 結果】

次に我々の手法の予測能力を確認するために、fitting に用いていない化学物質についてその物性値の予測を行った。logP では 202 分子、solubility では 58 分子、pKa では 67 分子の予測結果を Fig. 3 に示した。Table 2 には、各物性値の予測結果の統計値を示した。予測結果についても相関係数(r^2)は 0.9 以上であり、高精度に予測することが可能であることが分かった。

以上の結果より、ニューラルネットワーク法を用いることで、我々の定義した分子の 2 次元構造情報のみから高精度で広範囲な化学物質に適用可能な logP, solubility, pKa の新規予測式の構築ができた。

今後、環境ホルモンの毒性との構造活性相関の解析において、これらの物性値を取り込んだ新たな QSAR 解析を行うことで、更なる高精度での毒性予測が行えることと期待できる。

4. References

- [1] K.Yamagishi, T. Harada, H.Tokiwa, U.Nagashima, 日本薬学会第 124 年会(2004)
- [2] T.Aoyama, Y.Suzuki, H.Ichikawa, *J. Med. Chem.*, **33** (1990)
- [3] H.Chuman, A. Mori, H. Tanaka, *Anal. Sci.*, **18** (2002)

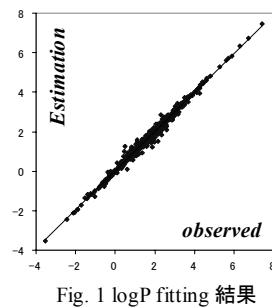


Fig. 1 logP fitting 結果

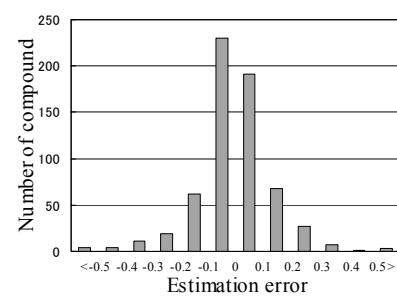


Fig. 2 fitting 誤差の分布

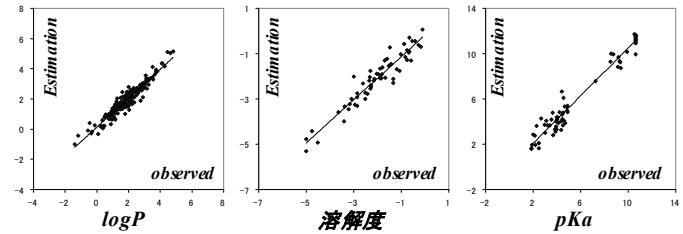


Fig. 3 logP, 溶解度, pKa の予測結果