

考察と予測モデルの構築

(産業医学総合研究所) 猿渡雄彦, 中西良文, (富士通(株))

湯田浩太郎, (日本バイオアッセイ研究センター) 松島泰次郎

【序論】

1979 年の労働安全衛生法の改訂により始まった有害物調査制度により、現在までに 12,000 件を超える微生物を用いる変異原性試験が行われ、労働省(厚生労働省)に届出られ、試験結果の概要は公表されている。産業医学総合研究所はこれらのデータの詳細を電子ファイル上にデータベース化し管理している。これらのデータをトレーニングセットとして、変異原性を化学構造式のみから予測するシステムを開発できれば、Regulatory Science 上も Green Chemistry 上も意義が大きい。危険な物質は開発合成すらしめない、生産しない、労働者に暴露させない、環境に放出しないという技術の実現に大きく貢献する。我々はそのような予測システムの実現を目指し研究を始めた。

2001 年末までに届出られた試験物質 11322 物質のうち混合物、ポリマー等を除いた 8042 の有機化合物のデータがあり、これが現時点でのトレーニングセットになる。(今後順次増える。)これらの化学物質のうち変異原性が陽性となるものの割合は 14.7%である。

莫大な化学空間を単一の予測モデルで処理するのは困難と考えられるので、何らかのクラス分けが必要となる。まず考えられるのは置換基によるクラス分けであるが、同じ化合物の中に基準となる置換基が複数含まれる場合に矛盾が起こる。そこで我々は1つの化合物には1つのクラス・1つのモデルを対応させるを試み、Drug-likeness の指標である Lipinski パラメータの中にその基準となるパラメータを求めた。このパラメータがベストのものとなる保証はないが、最初の試みとして最も合理的なものとする。同時に

Lipinski パラメータ(分子量、LogP, 水素結合ドナー数、アクセプター数、回転可能結合数)を含めた種々の物理化学定数と変異原性の間のトレンドアナリシスも行った。さらに得られた基準で分類された化合物群について試験的な予測モデルの構築を行った。

【方法】

2958 種の電荷をもたない(イオン化していない等)化合物について 85 の物理化学パラメータを3D-TSAR により算出した。変異原性は *Salmonella typhimurium* TA100, TA1535, TA98, TA1537 および *Escherichia coli* WP2uvrA のいずれかの菌種で代謝活性化の有無のいずれかで復帰変異原コロニー数が溶媒対照値の 2 倍を超えた場合を陽性とした。結果を DIVA によりカラムアナリシスを行い集計した。サブセットの 67 化合物を用いて線型学習機械法と Bayes 判別法による試験的な予測モデルの構築を行った。

【結果】

カラムアナリシスの結果の一部を Fig. 1 から Fig. 5 までに示す。あるパラメータ値の範囲内にある陽性化合物と陰性化合物の数をカラムの高さで表す。なお Ratio を示す折れ線は

$$\text{Ratio} = N_{\text{Positive}} / (N_{\text{Positive}} + N_{\text{Negative}})$$
 で定義される。

$N_{\text{Positive}} + N_{\text{Negative}}$ が約 100 以下の場合には Ratio の値の信頼性が低いので、下地を網目にしてある。

分子量が大きくなるに従い変異原性を示す割合が小さくなる(Fig. 1)。LogP は 0.5~2.5 くらいで変異原性を起こす確率が最も高く、その前後では減少していく(Fig. 2)。水素結合ドナー数とアクセプター数のついては顕著な相関はないが、アクセプター数では陽性化合物が広く均等に分散している(Fig. 3)。

Lipinski's Rule から外れると変異原性を示す割合が顕著に減少し(Fig.4)、生体とのインタラクションという点で変異原性を Drug-likeness に準じて取り扱うことが正当であると言える。分子軌道 Total Energy が高いと変異原性の出現の割合は顕著に高くなる(Fig.5)。

分子量が低く分子軌道 Total Energy の高い化合物が変異原性を示す確率が高いと言える。

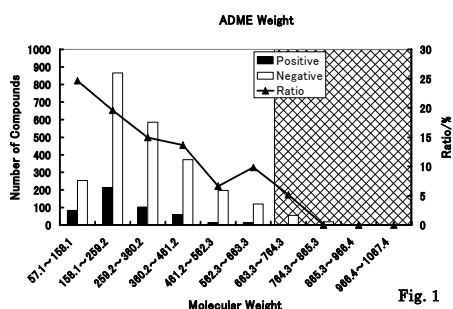


Fig. 1

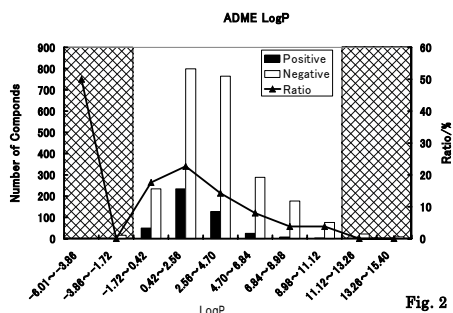


Fig. 2

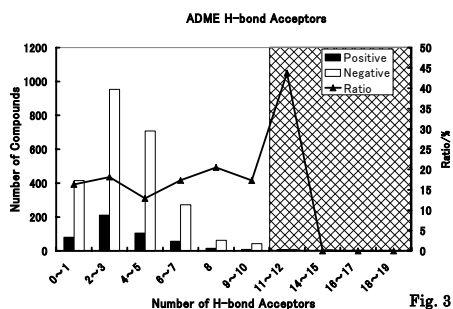


Fig. 3

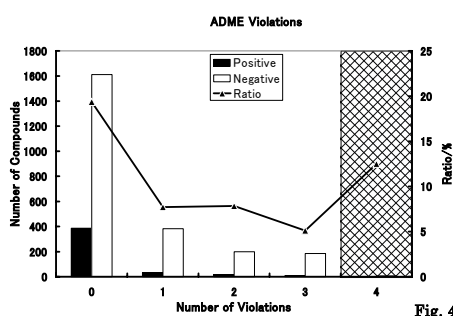


Fig. 4

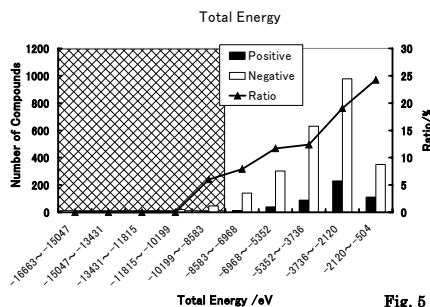


Fig. 5

我々は現在開発を試みている予測システムにおいて、化合物を1化合物1モデル対応のクラス分けする基準として水素結合アクセプター数を取ることにした。理由は①Lipinski Parameters の一つであること、②クラス数が小さくないこと、③変異原性陽性物質が偏りなく分散していること、などである。

水素結合アクセプター数 0~1 のクラスの化合物 495(陽性:81, 陰性 414)化合物から無作為に 67(陽性:17, 陰性 50)化合物を取り出し予測モデル

線型学習機械法/全体:100%				
		Calc		Classifi- cation/%
		Pos	Neg	
Obs	Pos	17	0	100
	Neg	0	50	100

Table 1

Bayes判別分析 /全体82.1%				
		Calc		Classifi- cation/%
		Pos	Neg	
Obs	Pos	13	4	76.5
	Neg	8	42	84.0

Table 2

予測率(LLM) 全体:82.1%				
		Calc		Classifi- cation/%
		Pos	Neg	
Obs	Pos	11	6	64.7
	Neg	6	44	88.0

Table 3

を作成した。分類率を Table 1, 2 に示す。

線型学習機械法 (LLM) によるモデルの Leave one out による予測率の評価を Table 3 に示す。

分類率と予測率に差があるのはト

レーニングセットのサンプル数が少ないためと考えられる。

以上のように線型学習機械法によるモデル構築の有用性が示された。