

ディビジョン番号	3
ディビジョン名	理論化学・情報化学・計算化学

大項目	2. 情報化学
中項目	2-2. ケモインフォーマティクス
小項目	2-2-1. 構造活性相関

概要（200字以内）

分子の構造記述子の開発についてはCoMFA以降注目すべき展開は余り見られない。構造記述子とその分子の生理活性との関係を示すモデルの開発については、新たに開発されたKohonenネットワークやSVMなどのパターン認識手法の導入が積極的に行われ、その評価が盛んに行われている。今後は、CoMFAに代わる3D-QSARのための構造記述子の開発と、パターン認識手法の特徴の整理が望まれる。

```

graph TD
    A[分子] --> B[記述子の開発と利用]
    B --> C[構造記述子]
    C --> D[関係モデルの開発と利用]
    D --> E[生理活性]
  
```

現状と最前線

構造活性相関（SAR）は分子の構造（Structure）と生理活性（Activity）との関係（Relationships）を明らかにする研究領域である。生理活性に対して物性（Property）との関係を研究する構造物性相関（SPR）も方法論を共有する関連領域として、近年定着し始めている。さらに、それらの定量的（Quantitative）関係を研究する領域を定量的構造活性相関（QSAR）、定量的構造物性相関（QSPR）と呼ぶ。

方法論的には、1）分子の構造情報を構造記述子によって表記する、2）生理活性や物性と構造記述子を関係づけるモデルを構築する、3）そのモデルを検証する、という流れを繰り返す事により、生理活性や物性を最もうまく予測できる構造記述子のサブセットを求めるといえる。従ってこの領域の研究は、どのような構造記述子を用いるかと、どのような方法でモデルを作るのかという視点で見ることができる。

構造記述子としては、分子の部分構造群をビット表現したフィンガープリント、分子をグラフと見なしてそのグラフ不変量を用いるトポロジカル記述子、分子の三次元構造、さらに分子を配置した三次元格子の格子点における相互作用エネルギーによる CoMFA（Comparative Molecular Field Analysis）などがある。SAR 用の記述子は多様な分子を表現する必要があるが、QSAR の場合は類似した分子群を扱えば良い。そのため、CoMFA は分子群の重ね合わせ処理を行わなければならないが、QSAR において広く用いられている。

また、構造記述子の代わりに活性以外の物理的性質（例えばpKaなど）を用いる研究も、モデル化の方法論を共有できるため構造活性相関研究として取り扱われている。

モデルを作る方法は、SAR と QSAR で共通に使われるものと独自のものがある。SAR はリード化合物の探索（バーチャルスクリーニング）等のように、分子が活性を持つか否かを識別する事が目的である。そのため、単純なK-最近隣法や判別分析に加えて、バックプロパゲーション学習アルゴリズムを用いたニューラルネットワーク、Kohonen の自己組織化マップやSVM (Support Vector Machine)

	SAR	QSAR
構造記述子	フィンガープリント トポロジカル記述子	CoMFA
モデル構築法	Kohonen マップ ニューラルネットワーク SVM Fusion	回帰分析 PLS

などの最近開発されたパターン認識手法、さらに類似度係数のランキングを複数の係数で融合 (Fusion) させる方法等が検討されている。

QSAR は複数の要素からなる構造記述子に対して、それぞれの要素の加重係数を求めることが目的である。そのため回帰分析を始めとする様々な統計解析手法が使われている。さらに近年は、ニューラルネットワークやSVMなどの手法の導入についても検討されている。なお、CoMFAなどの大量の要素からなる構造記述子を用いた場合はPLS (Partial Least Square) 法が一般に用いられている。

活性の発現メカニズムが生体高分子のレベルで分かって来ると、構造の類似性からではなく、ドッキング問題として構造活性相関を扱えるようになる。最近はそのようなアプローチの研究も進んでいる。

J. Gasteiger, T Engel (Eds), Chemoinformatics - a textbook. Wiley-VCH, Weinheim. 680p (2003)

P. Willett, Drug Discovery Today, 11 (23/24), 1046-1053 (2006).

将来予測と方向性

- ・ 5年後までに解決・実現が望まれる課題
 - CoMFA に代わる 3D-QSAR のための構造記述子の開発
 - 多様なモデル化手法に対する整理と特徴のまとめ
- ・ 10年後までに解決・実現が望まれる課題
 - 高精度なドッキング問題の解法の開発

キーワード

構造記述子、バーチャルスクリーニング、パターン認識、CoMFA、PLS

(執筆者：中山 伸一)