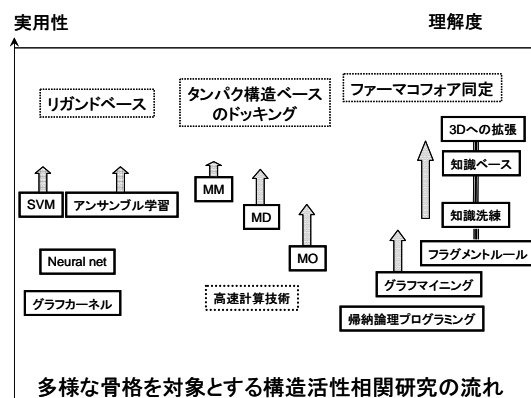


ディビジョン番号	3
ディビジョン名	理論化学・情報化学・計算化学

大項目	2. 情報化学
中項目	2-3. ケモメトリックス
小項目	2-3-2. 構造活性相関

概要（200字以内）

類似した骨格の化合物群を対象とした定量的構造活性相関や COMFA の技術は定着している。多様な骨格を対象とする場合、リガンドベースでは、機械学習法の導入が盛んである。他方、化学者による理解を重視したファーマコフォア同定が重要な課題であり、今後知識ベースが構築できればもっとも有望であろう。タンパク質構造が既知の場合は、理論化学面と高速計算技術の進歩により着実な発展が期待される。



現状と最前線

構造活性相関の研究は、Hansch, Fujita による回帰式を用いた定量的構造活性相関については、すでに確立した方法論となっており、疎水性パラメータや置換基定数を説明変数としてリード化合物最適化のため、日常的な応用が行われている。その適用分野も、医薬、農薬の活性から、一般化学物質の蓄積性や生分解性、また各種の物性との相関にまで広がっている。

さらに、リガンドの3次元構造を重ね合わせてファーマコフォアの特徴を同定する COMFA と呼ばれる試みもよく行われる。しかし、実際の分子設計では良い回帰式が得られても、実際どのような置換基を導入するかという問題に関しては、答えの得られないことも多い。

このような状況を受けて、医薬設計に関してはハイスループットスクリーニング技術を適用するというロボット技術に傾斜する状況が 90 年代に発生した。しかし、現実にはあまりにも活性を示す化合物が少ないため、活性を示す可能性の高い化合物を *in silico* であらかじめ選択する化合物ライブラリデザインの必要性が出てきた。このようなリガンドベースの構造活性相関で使われる技術の要点は、(1) 結果は必ずしも化学者によって理解できなくともよい、(2) 構造式から切り出した部分構造を bit 列としたフィンガープリントを利用する、(3) 予測精度を上げるためにニューラルネットをはじめとする機械学習分野で開発された方法を何でも使ってみる、というように変化してきている。精度はそれほど高くないが、スクリーニングという目的からすれば効果はあったと認識されている。

他方、タンパク質の3次元構造が既知の場合には、ドッキングのためのソフトウェアが多数商用化されており、in silico スクリーニングのために日常的に使われている。これらのソフトウェアの稼働原理は基本的に分子力場法である。現実にはその精度に問題があるため、結果の化学者による目視チェックが必要な状況であり、ライブラリデザインのための利用が主流となっている。

〔最前線〕 以上のような状況で、現時点での研究の最前線は、大きく分けて以下の3つの方向に分類できる。

第一の方向は、タンパク質構造を利用した高精度のドッキングシステム開発である。分子動力学法、さらに量子化学計算まで取り入れて精度を上げる試みがある。理論面の進展と共に、計算機の高速度化とそのアーキテクチャーに適合したソフトウェアの開発が中心となっている。

第2の方向は、最新の機械学習法を次々と導入し、リガンドベースでとまかく予測精度を上げることである。boosting, bagging のようなアンサンブル法やSVM法、さらにグラフカーネルの導入まで行われている。また、物理化学的あるいはグラフ理論による記述子も開発されている。しかし、この方向性に関してはすでに化学ではないという観点からの批判もある。

最後の方向は、多様な化学構造から理解できるファーマコフォアを抽出することである。データマイニング技術が主役となるが、その中でも帰納論理プログラミングやグラフに特化した各種マイニング技法に加え、多数のフラグメントを利用して命題論理の範囲内で特徴的部分構造を探索するアプローチが提案されている。現状では、後者のアプローチの方が支持化合物数の少ない活性にも対応できるため、有効と考えられている。

この研究方向におけるもう一つの問題点は、活性を識別できる部分構造を発見したとしても、それがデータベースのバイアスに依存しており、必要条件ではあっても十分条件とはなっていないことである。十分条件を表現する部分構造が得られてこそ、化学者にとり有効な知識となる。そこで、構造式空間での知識洗練により、十分条件となるファーマコフォアを抽出する試みが成功し、実際に各種の活性に対して知識ベースを構築する試みが始まっている。

T. Okada: "Mining from Chemical Graphs", in D.J. Cook & L.B. Holder Ed., "Mining Graph Data", Chap. 14, pp.347-379, Wiley - Interscience (2007)

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
構造式レベルでのファーマコフォア知識ベースの構築
良好な精度（50%以上）を与えるライブラリデザイン手法
- ・ 10年後までに解決・実現が望まれる課題
3次元構造レベルでのファーマコフォア知識ベースの構築
定量的精度でのドッキングを可能とするシステムの開発

キーワード

データマイニング, ファーマコフォア, virtual screening, ライブラリ設計, 知識ベース

(執筆: 岡田 孝)