

○右田啓哉、荒川正幹、船津公人

東京大学大学院工学系研究科（〒113-8656 東京都文京区本郷7丁目3番1号）

[緒言]

化合物データベースを分析して物性に関するモデルを得るために様々なツールが活用されているが、化学構造の特徴を示す多様なパラメータ（化学構造記述子：molecular descriptor、以下、記述子）を計算するツールにも多くの種類がある^[1]。本発表ではQSPR解析の信頼性のために重要である「記述子の意味」に重点をおいた新しい記述子計算ツールである Descript (Descriptor - ist) とその物性推算システムへの組み込みの概観を示す。

[Descript 概要]

記述子を用いた様々な解析の試みにおいては、解析で使用する記述子の意味する内容に気を配る事で結果の化学的な信頼性が増す。また十分な知見を有す場合には計算負荷の軽減が可能となる。しかし、複雑な機構をもつ物性や複雑な構造を持つ化学構造についての解析では分析しなければならない記述子が多くなり、計算量が增大する上に解釈も困難になりがちである。そこでデータマトリックスを生成するだけでなくその物理化学的な意味を汲み取り、数値処理と物理化学的解釈の両輪による解析をサポートする新しい記述子計算ツールが必要となる。

Descript では様々な記述子に対してコンピュータ処理できる形で意味を定義し、その意味情報を利用しながら解析の援助をする。各記述子について計算方法や背景から物理化学的的属性などについて属性項目を設定し、様々な解析においてこの情報を用いることでデータの傾向や構造物性相関の解釈が進み、信頼性のある解析が可能となる。属性としては単純な部分構造数から体積や構造のフレキシビリティといった形状情報、表面電荷や分子間相互作用といった電気的性質に関する情報など様々な可能性が考えられる。解析への応用としては、例えば相関関係やPCAの主成分においてそれらの値が示す物理化学的な意味に対する解釈をサポートすることなどが挙げられる。また目的のデータ群にどのような構造特徴があるか判定する事もできるため、モデルデータセットの選定や後述するように物性推算システムへの組み込みも可能である。なお Descript は Windows アプリケーションとして作成しており、コーディングには Visual C++ を用いた。設計においては計算速度の重視とともに将来様々な拡張が為される可能性を考慮した。

[物性推算システムのモデル定義への利用]

昨年度発表した^[2]物性推算システム（図1）は、目的とする化合物の化学構造から自動的にQSPRによる物性推算を行うシステムであり、モデルを蓄積するモデルライブラリとモデルの選択と計算を行うコントローラからなる。このシステムにおいて目的の化学構造についての物性を精度良く推算するモデルを選ぶためには事前にモデルの予測対象構造を適切に定義することが重要である。この定義は厳密であれば良いというものでもなく、意味を踏まえて簡潔に記述する事が実用において重要である。

昨年度は化学構造の特徴を部分構造数ベクトルで定義することでモデルデータセットの構造特徴を記述し、一定の精度でモデル選択を自動化した。しかしながら依然としてモデルの定義など構造の特徴を踏まえ手作業で行わなければならない部分が残されている。この点については記述子の意味を踏まえた数値操作によって自動化が可能であると考えられ、Descript を用いる事で解決を図ることができる。

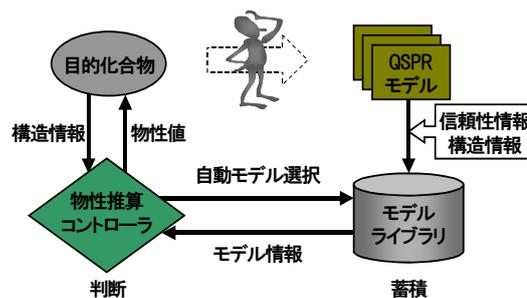


図1:物性推算システム

[参考文献]

- [1] R. Todeschini and V. Consonni, Handbook of Molecular Descriptors, WILEY-VCH (2000)
 [2] 右田啓哉、荒川正幹、船津公人、構造物性相関モデルライブラリの構築、日本コンピュータ化学会 2005 春季年会講演予稿集、2P07 (2005)