

氏名	原 秀 人
学位(専攻分野)	博士(薬学)
学位記番号	薬博第655号
学位授与の日付	平成20年3月24日
学位授与の要件	学位規則第4条第1項該当
研究科・専攻	薬学研究科医療薬科学専攻
学位論文題目	大規模情報可視化を利用した薬物動態特性の構造活性相関解析に関する研究

論文調査委員 (主査) 教授 橋田 充 教授 高倉喜信 教授 乾 賢一

### 論文内容の要旨

コンビナトリアルケミストリーやハイスループットスクリーニングをはじめとする創薬技術の飛躍的な進歩により、網羅的な医薬品探索研究が主流となり、多変量かつ大規模な化合物情報が収集されるようになってきた。このような状況において期待されるものは、得られた化合物情報を適切に処理し、探索合成研究を支援する情報科学的解析方法である。従来、重回帰分析やニューラルネットワークなどの情報科学的手法を基本として、体内動態特性と化学構造との関係を解明する定量的構造活性相関(QSAR)研究が盛んに行われてきた。しかしながら、これらは、あくまで個別の化合物特性を経験論的にモデル化するものであり、モデルによる予測結果の妥当性が問題となることに加え、しばしば互いにトレードオフの関係となる創薬化合物の様々な特性を俯瞰的に捉えて創薬研究に還元するような目的には利用できない。そこで申請者は、近年情報学分野で精力的な研究がなされている情報可視化に注目し、創薬研究によって得られる多面的な化合物特性情報を視覚的に表現することでヒトが持つ潜在的なパターン認識能力によりそのデータ構造を包括的に捉えることのできる技術開発を行った。まず大規模な化合物構造情報を縮約し、化合物特性空間の中での目的化合物の局在を効果的に視覚化する方法を考案した。また、実験的に得られる定量的データを効果的に表示する方法として、化合物を構造に基づいて階層的に分類し表示する方法を開発した。さらに、階層分類構造においてよりデータ構造を捉えやすい対称決定木を構築するための新規再帰分割法を開発した。以下、これらの研究成果について解析実施例を交えながら論述する。

#### 第I章 動態特性判別を目的とした化合物特性空間の新規可視化法の開発

化合物ライブラリー設計を行う際には、Drug likenessのように化合物特性を定性的に2値判別するモデルがしばしば利用される。一般に判別分析モデルを構築するためには事前に正例および負例のデータが必要となるが、決定的に得られる事前情報がいずれか一方しかなく、他方は仮定の下で解析することがしばしば行われている。そこで、このような問題を解決するために、主成分分析等で行われてきた化合物特性空間を用いた解析法を発展させ、情報可視化を組み合わせながらバイアスフリーで2値判別ができる方法を考案した。すなわち、化合物全体の分散に対して目的化合物の分散を最小にするように写像空間を定義することによって、化合物特性空間における目的化合物の局在位置から識別が可能になると考え、さらにこの数学的問題が化合物群の分散共分散行列を用いた一般固有値問題に帰着されることを見出した。本手法により経口医薬品の局在を解析したところ、3次元で定義される化合物の全体空間のうち経口薬が約1/6の空間に局在していることが明らかとなった。比較のために主成分分析法で化合物特性空間を定義したところ、経口薬が空間全体に分布してほぼ識別力を欠くことが示され、本方法の有効性が証明された。さらに、同様の解析は脳移行を示す薬物に対しても可能であり、本方法が様々な問題に対して適用可能であることを確認した。

#### 第II章 階層的な分類を応用した視覚的構造活性相関研究法の開発

視覚的に捉えられる2次元あるいは2次元空間にプロットする方法は分布の特徴を直感的に捉える上で有効であるが、大規模データを解析した場合には、限られた表示空間の中でデータポイントの重なりが生じたり、全体分布を捉えるために広

角表示しなければならないなどの問題が起こる。したがって、大規模なデータがもつ全体的および局所的な特徴を同時に捉えるためには新しい情報可視化方法が必要であると考え、階層的データを2次元空間に再帰的な入れ子構造のグラフとして効率よく表示する平安京ビューを導入し、化合物構造記述子に基づくクラスター解析あるいは再帰分割解析と組み合わせることで化合物の構造活性相関を視覚的に解析する方法を考案した。文献より化合物の溶解度に関するデータを収集するとともに、それらの化合物の構造記述子に基づいて化合物の階層的クラスタリングを行い、それらの結果を平安京ビューで表示したところ、類似構造による分子量やClogPの同等性が溶解度と関係し、分子量が大きくClogPが高いものほど溶解度が低くなる傾向を視覚的に捉えることに成功した。

さらに、多特性データを同時に扱う構造活性相関解析方法として、各特性データの情報利得の総和を分割基準評価式とした再帰分割解析法を考案し、得られた結果を拡張平安京ビューにより可視化する目的変数データを表示するための方法を確立した。6つのチトクロムP450 (CYP) 分子種による代謝感受性データを本方法により解析した結果、CYP2E1およびCYP2D6がそれぞれ低分子化合物、カチオン性化合物を基質とすることなど各CYP分子種の基質特異性に関して一定の法則性を見出すことに成功した。

### 第三章 対称的構造を持った決定木構築のための再帰分割解析法の開発

第II章でも利用した再帰分割解析法は、各ノードが単純な評価式で定義される決定木を与え、階層的データ管理を行う上でも有効な方法である。しかしながら、貪欲アルゴリズムと呼ばれるトップダウン的モデリング法で木構造が形成されるため、最終的に得られた決定木が非対称となり、場合によっては全体の構造が捉えにくくなることがある。そこで、各階層ノードでの評価式が同一となるようにし対称的な決定木を構築するための再帰分割解析法を考案した。これは、各階層における評価式を網羅的に探索して分割後の誤分類率を計算し、赤池情報量基準AICを指標として最適な評価式を選択するというものである。医薬品の経口バイオアベイラビリティに関する232個のデータを解析したところ、種々の分子記述子の中で化合物の電荷がバイオアベイラビリティを規定する第一の要因であり、カチオン性薬物ではバイオアベイラビリティが低くなる傾向が認められた。この対称的決定木モデルは、自由度が低いため従来の再帰分割法で得られるモデルよりもやや予測精度が低下するものの、対称な構造のためにデータの特徴が理解しやすく、化合物特性上の問題の抽出や改善を行う上で有効なツールとなることが示された。

以上、大規模情報を適切に処理し可視化することで、薬物の化学構造と動態特性との間の関係を直感的に理解しうる方法を開発した。本研究成果は、創薬研究で求められる意思決定を効率的に行うための方法論および情報を提供するものと考えられる。

## 論文審査の結果の要旨

近年、網羅的な医薬品探索研究が主流となり多変量かつ大規模な化合物情報が収集されるようになったことから、得られた化合物情報を適切に処理し探索合成研究を支援するための情報科学的解析方法が注目を集めている。従来行われてきた重回帰分析やニューラルネットワークなどの情報科学的手法を基本に体内動態特性と化学構造との関係を解明する定量的構造活性相関 (QSAR) 解析は、個別の化合物特性を経験論的にモデル化するもので創薬化合物の特性を俯瞰的に捉えて創薬研究に還元するには適さないことから、申請者は情報可視化に注目し、創薬研究によって得られる多元的な化合物特性情報を視覚的に表現することでヒトが持つ潜在的なパターン認識能力によりそのデータ構造を包括的に捉えることを目指して3種に大別される技術開発を行った。

一般に、化合物ライブラリーの設計にはDrug likenessのように化合物特性を定性的に2値判別するモデルが利用されるが、決定的に得られる事前情報が判別分析モデルの構築に必要な正例および負例のデータの一方であることも多い。そこで、主成分分析等で行われてきた化合物特性空間を用いた解析法を発展させ、情報可視化を組み合わせながらバイアスフリーで2値判別ができる方法として、化合物全体の分散に対して目的化合物の分散を最小にするように写像空間を定義し化合物特性空間における目的化合物の局在位置から識別することを試み、この数学的問題が化合物群の分散共分散行列を用いた一般固有値問題に帰着されることを見出した。本手法により経口投与薬が3次元で定義される化合物の全体空間の約1/6の空間に局在し、一方、主成分分析法ではほぼ識別力を欠くことが示されその有効性が証明された。次に、情報の2次元あるいは3

次元空間プロットは分布の特徴を直感的に捉えるのに有効であるが、大規模データを解析し全体的および局所的な特徴を同時に捉えるためには新しい情報可視化手法が必要と考えられることから、階層的データを2次元空間に再帰的な入れ子構造のグラフとして効率よく表示する平安京ビューを導入して化合物構造記述子に基づくクラスター解析あるいは再帰分割解析と組み合わせることにより、化合物の構造活性相関を視覚的に解析する方法を考案した。本方法で化合物の溶解度と化合物の構造記述子に基づいて化合物の階層的クラスタリングを行い、類似構造による分子量やClogPの同等性と溶解度の関係を視覚的に捉えることに成功し、さらに、多特性データを同時に扱う構造活性相関解析方法として、各特性データの情報利得の総和を分割基準評価式とした再帰分割解析法を考案し、得られた結果を拡張平安京ビューにより可視化する目的変数データを表示するための方法を確立して6種のチトクロムP450分子種の代謝における基質特異性に関して法則性を見出した。さらに、再帰分割解析法は各ノードが単純な評価式で定義される決定木を与え階層的データ管理を行う上でも有効であるが、最終的に得られた決定木が非対称となり全体の構造が捉えにくいことから、各階層ノードでの評価式が同一となるようにして対称的な決定木を構築するための再帰分割解析法を考案し、医薬品の経口バイオアベイラビリティに関する232個のデータを解析して、バイオアベイラビリティを規定する要因を明らかにし、本法が化合物特性上の問題の抽出や改善を行う上で有効なツールとなることを証明した。

以上、大規模情報を適切に処理し可視化することで、薬物の化学構造と動態特性との間の関係を直感的に理解しうる方法を開発した。本研究成果は、創薬研究で求められる意思決定を効率的に行うための方法論および情報を提供するものと考えられる。

よって本論文は博士（薬学）の学位論文として価値あるものと認める。

さらに、平成20年2月22日論文内容とそれに関連した口頭試問を行った結果、合格と認めた。