The University of Osaka
Institutional Knowledge Archive

Title	創薬データベースの統合による活性化合物探索手法に関する研究
Author(s)	河村, 元
Citation	大阪大学, 2008, 博士論文
Version Type	
URL	https://hdl.handle.net/11094/49269
rights	
Note	著者からインターネット公開の許諾が得られていないため、論文の要旨のみを公開しています。全文のご利用をご希望の場合は、大阪大学の博士論文についてをご参照ください。

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

【27】

氏　　名	河　村　　元（かわ むら げん）
博士の専攻分野の名称	博　士（情報科学）
学　位　記　番　号	第　２２１６６　号
学 位 授 与 年 月 日	平成 20 年３月 25 日
学 位 授 与 の 要 件	学位規則第４条第１項該当 情報科学研究科バイオ情報工学専攻
学　位　論　文　名	**創薬関連データベースの統合による活性化合物探索手法に関する研究**
論 文 審 査 委 員	（主査） 教　授　松田　秀雄 （副査） 教　授　四方　哲也　　教　授　清水　　浩　　教　授　前田　太郎

論 文 内 容 の 要 旨

近年、バイオテクノロジーなどの急速な発展により、創薬に関連するゲノム情報、化合物情報、疾患や薬理など、多様な種類の情報も急速に増大し創薬関連データベースへ蓄積されるようになってきた。このように蓄積した情報を活用しどう役立てていくかが今後のバイオインフォマティクスにおける大きな課題として認識されている。

しかし、実際に創薬関連データベースを利用して、ターゲットタンパク質に作用する化合物の探索を行おうとすると、非常に複雑な手順が必要になる。そこで、本論文ではインターネット上に蓄積した創薬関連データベースをメタデータとグリッド技術を用いて統合し取扱う方法を提案する。

また、創薬のプロセスには化合物構造間の類似性評価を行う必要がある。化合物構造の数値化の一つとして部分構造の有無をビットで表したフィンガープリントとよばれる表記法が提案されている。このビット列を距離関数で評価することで、化合物間の類似距離を求めることができる。しかし、従来の方法では全てのビットが常に等価に扱われていることが多い。本研究では、化合物の構造類似度を評価する方法として教師なしの Tanimoto 係数による距離尺度と、教師ありの距離尺度の Random Forest の proximity measure を線形判別法で組合せた化合物探索法の提案を行う。この組合せ尺度を用いることで化合物の活性クラスに関与した部分構造の特徴も考慮しながら精度良く化合物探索が行えることが期待できる。

さらに、ターゲットタンパク質に対して相同なタンパク質と作用が既知のリガンドを用いて化合物データベースで類似度探索を行うとターゲットタンパク質に対して作用するリガンド候補が得られるという考え方がある、しかし、この考え方によって新規ターゲットタンパク質に作用する活性候補化合物の候補が探索できるが、膨大な数の偽陽性な結果が拾いあげられてしまう可能性がある。よって各段階において、タンパク質間の相同性を考慮しながら化合物の検索精度を向上させ、活性化合物を選択するかは重要な課題である。そこでタンパク質の相同性検索を提案した化合物探索手法と組合せて未知リガンドを取得する方法を試みる。本手法に適用することでターゲットタンパク質に作用するリガンドのセットが有効に取得できることをしめす。

本研究により、現在の分散した創薬関連データベースの情報統合とリガンド探索などの創薬関連情報のデータマイニングが容易に実行できるようになり創薬プロセス全体の効率化に繋がることが期待できる。

論文審査の結果の要旨

本論文では、インターネット上に分散した創薬関連データベースを統合し取扱う手法を提案し評価を行っている。また、化合物の構造類似性の指標を組み合わせた活性化合物の探索手法を提案し、さらにその手法をタンパク質の相同性検索と組み合わせることで新規タンパク質へ作用する活性化合物の探索手法についての提案と評価を行っている。

ゲノム創薬の各プロセスでは多種の創薬関連データベースが必要になる。本論文では、近年それらの創薬関連データベースが分散しかつ大規模になっており、創薬プロセスにおいてそれら複数のデータベースを連携した探索が困難になっていることを述べている。さらに創薬関連データベースの直接的な統合が困難であることを指摘している。この解決のために複数の概念的なメタデータと OGSA-DAI と呼ばれるデータリソース統合のためのグリッドミドルウェアを用いた統合手法を提案している。この統合手法の実装を行い、複数の創薬関連データベースの接続性の評価によって従来手法よりも短い手順で創薬関連情報を検索できることを示している。次に、化合物構造間の類似度の計算には従来は化合物の部分構造表記である MACCS key と Tanimoto 係数と呼ばれる類似尺度が利用されているが、Tanimoto 係数では全てのビットが常に等価に扱われていることなどから、化合物探索の精度が十分でない可能性があることを指摘している。そこで、化合物構造間の類似度を評価する方法として Tanimoto 係数と、機械学習法の Random Forest を利用した類似尺度である proximity measure を線形判別法で組合せて化合物探索を行う方法を提案している。この組合せ尺度による評価実験では従来手法に比べて提案手法が有効であることが示されている。

また、活性化合物が未知のタンパク質と相同なタンパク質に対して既知の活性化合物との関係を利用して化合物データベースから類似度探索を行うことでタンパク質に作用する新たな活性化合物候補が得られるという考え方がある。本論文ではこの考え方で活性化合物探索を行う場合に、大量の偽陽性な化合物が拾いあげられる可能性があることを指摘している。この問題の解決のために上述した活性化合物探索手法を適用することを提案し、評価実験を行った結果、本手法が有効であることが示されている。

以上により、本論文の成果は、タンパク質に作用する活性化合物の探索手法に関する研究の発展に貢献するものと考えられる。よって、博士（情報科学）の学位論文として価値あるものと認める。