

論文 / 著書情報  
Article / Book Information

題目(和文)	化合物群の部分構造の共通性に着目した高速バーチャルスクリーニング手法の開発
Title(English)	Fast structure-based virtual screening with commonality of compound substructure
著者(和文)	柳澤 溪甫
Author(English)	Keisuke Yanagisawa
出典(和文)	学位:博士(工学), 学位授与機関:東京工業大学, 報告番号:甲第11198号, 授与年月日:2019年3月26日, 学位の種別:課程博士, 審査員:秋山 泰,篠田 浩一,村田 剛志,関嶋 政和,石田 貴士
Citation(English)	Degree:Doctor (Engineering), Conferring organization: Tokyo Institute of Technology, Report number:甲第11198号, Conferred date:2019/3/26, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Category(English)	Doctoral Thesis
種別(和文)	論文要旨
Type(English)	Summary

## 論文要旨

THESIS SUMMARY

系・コース： Department of, Graduate major in	情報工学 知能情報	系 コース	申請学位 (専攻分野)： 博士 (工学) Academic Degree Requested Doctor of
学生氏名： Student's Name	柳澤 溪甫		指導教員 (主)： Academic Supervisor(main) 秋山 泰 教授
			指導教員 (副)： Academic Supervisor(sub)

### 要旨 (和文 2000 字程度)

Thesis Summary (approx.2000 Japanese Characters)

近年、薬剤開発の高度化および低コスト化を目的として、多数の化合物と薬剤標的タンパク質との結合力を計算機で評価し、薬剤候補化合物を選別するバーチャルスクリーニング (Virtual Screening, VS) が広く実施されている。バーチャルスクリーニングは標的タンパク質の立体構造情報を活用する構造ベース法、既知薬剤の情報を活用する化合物ベース法に大分することができる。そのうち、構造ベース法は物理化学的な相互作用を考慮でき、得られる薬剤候補化合物の新規性の高さから注目されている一方、そこで用いられるタンパク質-化合物ドッキング計算に要する時間が膨大であることが問題となっていた。また、計算時間削減のためにあらかじめ化合物を高速に評価するプレスクリーニングが提案されてきているものの、構造ベースのプレスクリーニング手法の速度は十分ではない。本論文では、構造ベース法による数千万化合物のバーチャルスクリーニングの実現を目標として、化合物の部分構造の共通性に着目した高速なプレスクリーニング手法およびドッキング計算手法を提案した。

まず、化合物の部分構造について、最小単位「フラグメント」を単結合の回転による構造変化が発生しない部分構造であると定義し、化合物分割手法を提案した。同分割手法を 3 種類の化合物データベースに適用すると、化合物数の 12 分の 1 から 110 分の 1 程度の種類のフラグメントで全化合物が表現できることを確認した。加えて、化合物データベースが大きくなるほどフラグメントの共通性が高まることを確認した。近年研究されている仮想的な大規模化合物ライブラリに対しては有用性がさらに高まることが示唆される。

次に、この化合物分割によって得られたフラグメントをドッキングの前処理として候補化合物の絞り込みを行うプレスクリーニングのための手法「Spresso」を提案した。Spresso の処理は (1) 化合物のフラグメント分割、(2) フラグメントのドッキング計算、(3) フラグメントスコアを用いた化合物評価、の 3 段階に分けることができる。分割されたフラグメントの評価を独立に行い、その評価スコアの最大値や非線形和などを取ることで、近似的ではあるが高速な化合物の評価を可能にしておき、フラグメントの評価に商用ドッキングツールである Glide を利用した例では、既存のプレスクリーニング手法である Glide HTVS モードと比して約 300 倍の高速化を達成した。さらに、高速なプレスクリーニング手法がどのように創薬プロジェクトに貢献できるのか、スクリーニングに関する複数のシナリオの比較検討を行い、Spresso の有用性を確認した。

また、化合物分割手法をドッキング計算に適用する方法も本論文で述べている。フラグメント伸長に基づくドッキングアルゴリズムは FlexX (Rarey, et al., 1996) など既存ドッキングツールでも採用されているが、このアルゴリズムを用いることで「フラグメント計算結果の再利用」に加えて、「フラグメント伸長の途中結果の再利用」も可能であることを示した上で、フラグメント伸長の途中結果の再利用効率を高める方策として、化合物評価順序の並び替えを提案した。一方、フラグメント計算結果の再利用においては、保持されるべき計算結果はデータサイズが大きく、一度に保持するデータ数に上限があることから、計算結果の保持の取捨選択が必要である。これは主メモリ空間をキャッシュ空間に見立てたキャッシュ問題として考えられる。

このドッキング計算における計算結果の再利用を実現するためのキャッシュ問題は、化合物評価順序があらかじめ決定できるというオフライン性の仮定の下で、重み付きオフラインキャッシュ問題と考えることができる。この問題は最小費用流問題に帰着できることが知られており、多項式時間で最適なメモリ利用戦略が導出可能である。さらに、帰着された最小費用流問題のグラフはトポロジカルソートが可能であり、かつ疎グラフであり、各辺の流量が限定的であることから、これらの特徴を利用した、動的計画法による効率的なキャッシュアルゴリズムを提案した。実際の化合物から作られた例題と、ランダムアクセスを仮定した例題のそれぞれについて従来手法との比較を行い、最大約 9 倍高速に最適なメモリ利用戦略を得られることを確認した。また、提案手法が必要とする計算量と、フラグメント計算結果の再利用によって期待される高速化の程度との比較を行い、提案したメモリ利用最適化がドッキング計算の高速化のために有用であることを確認した。

Spresso やドッキング計算における計算結果再利用の枠組みは、計算量の削減を通して従来手法に対する高速性を担保している。また、化合物数が増えるほど部分構造の共通性が高まることも確認されている。より多くの化合物で構成されるデータセットを用いた、構造に基づくバーチャルスクリーニングの実現のために有用な手法であり、今後薬剤開発における構造ベースバーチャルスクリーニング手法として広く利用されることが期待される。

備考：論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

(博士課程)  
Doctoral Program

## 論文要旨

THESIS SUMMARY

系・コース : Department of, Graduate major in	情報工学 知能情報	系 コース	申請学位 (専攻分野) : Academic Degree Requested	博士 Doctor of	( 工学 )
学生氏名 : Student's Name	柳澤 溪甫		指導教員 (主) : Academic Supervisor(main)	秋山 泰 教授	
			指導教員 (副) : Academic Supervisor(sub)		

要旨 (英文 300 語程度)

Thesis Summary (approx.300 English Words )

Virtual screening (VS) is a computational method which is widely used to evaluate the binding affinity between a vast number of compounds and a protein before conducting *in vitro* assays. In particular, structure-based VS (SBVS), which uses protein tertiary structures, is attracting attention because it can consider the binding affinity in a physico-chemical manner. However, SBVS is too computationally heavy to evaluate 10 million-order of compounds. Pre-screening methods filtering out unfeasible compounds have been proposed to avoid the problem of computational cost. Most of them are ligand-based, prediction from known inhibitors, and structure-based pre-screening methods are still needed to be accelerated.

In this dissertation, we designed a pipeline for faster virtual screening with a fundamental idea: substructure commonality. Commonality can be used to reuse of intermediate results. We firstly defined “fragment” to maximize meaningful common substructures and we revealed our definition express compounds with much smaller number of fragments than the RECAP, an existing decomposition rule. The fact leads to faster calculation of pre-screening and docking calculation.

As a first step of the structure-based virtual screening pipeline, we proposed the pre-screening method, Spresso. It evaluates compounds based on the binding scores of their fragments, and thus the calculation results of fragments are reused to multiple compounds. It is capable of exhaustive primary screening and completed pre-screening approximately 200 times faster than the conventional method (Glide HTVS) while precision is slightly compromised. When Spresso was applied to a more than 26 million compounds, the calculation time was within a day.

The second step of the pipeline is docking calculation. We designed a docking calculation procedure with reuse of intermediate results. The procedure is based on fragment-based anchor and grow algorithm. We found that one of the intermediate information, called the fragment grid, is an essential factor while it spends huge memory space; therefore, we formulated and reduced the problem of optimization of the fragment grids reuse as the minimum cost flow problem, which is known as a polynomial-time problem. Because the reduced graph of the problem is topological sortable and sparse, we finally proposed approximately 9 times faster algorithm.

備考 : 論文要旨は、和文 2000 字と英文 300 語を 1 部ずつ提出するか、もしくは英文 800 語を 1 部提出してください。

Note : Thesis Summary should be submitted in either a copy of 2000 Japanese Characters and 300 Words (English) or 1copy of 800 Words (English).

注意 : 論文要旨は、東工大リサーチリポジトリ(T2R2)にてインターネット公表されますので、公表可能な範囲の内容で作成してください。

Attention: Thesis Summary will be published on Tokyo Tech Research Repository Website (T2R2).