

GPGPU で実現するバーチャルスクリーニングのための 高速分子形状比較プログラム「FastROCS」

FastROCS: Harnessing the Power of GPU Computing to Speed Up Drug Discovery

佐藤 秀行^a, 勝山 マリコ^a

Hideyuki Sato and Mariko Katsuyama

1. はじめに

創薬研究の初期段階における活性化化合物探索方法の一つに、市販化合物や各製薬企業が所有する化合物ライブラリなどを対象として行われるバーチャルスクリーニング (VS) があります。これは、条件を満たす化合物を大量の化合物からコンピュータ上で絞り込む方法で、化合物を新たに合成することなく候補化合物の生物学的有用性の確認が可能なることから、研究の効率化の観点により近年注目されています。一方、タンパク質は結合する低分子化合物を、化学構造ではなくその形状 (Shape) と電子的特徴 (Electrostatics) に基づいて識別しています。つまり、これら特徴を利用した VS は、化学構造が異なりながら同等の活性を有する化合物を探索する有用な手法として考えられます。そこで、我々は低分子化合物の形状に着目し、活性既知の低分子化合物の構造にのみ基づいて、数百万化合物からなる大規模化合物データベースを対象として VS を実施する分子形状比較プログラム ROCS¹を開発しました。ROCS は、これまでに多くの創薬研究者によりその有用性が実証されており、タンパク質の活性部位と化合物間の相互作用に基づいて行われる VS (ドッキング) よりも優れているとの報告²もあります。本稿では、ROCS を GPGPU (General-Purpose computing on Graphics Processing Units) に対応させることで、計算速度を数百~1,000 倍と飛躍的に高速化し、1 秒間に 200 万配座の計算処理を実現した高速分子形状比較プログラム FastROCS を紹介します。

2. GPGPU による高速並列計算

GPGPU は、通常はグラフィックス表示に使用される 400 個以上のコアを配置したボードにより並列計算を実現する高速計算技術です。ワークステーション 1 台当たり 1 テラフロップス以上の計算能力を追加することができ、高速計

算技術が必要とされる分野では急速に実用化が進められています。特に、ワークステーションクラスの筐体に 4 枚の GPGPU 用のボードを設置することができることから、大規模なサーバーームを用意せずに高速並列計算を実現することが可能となり、今後益々注目される技術であると考えられます。

一方、CPU (Central Processing Unit) 用プログラムは一般的には C/C++ などの言語で作成されていますが、GPGPU で動作させるためには OpenCL などの専用言語で書きなおす必要があります。そのため、この様に高速化計算に有用な技術であっても、現時点ですべてのプログラムが GPGPU に移植されているわけではありません。本稿で紹介する FastROCS は、CPU で動作する ROCS を GPGPU 仕様とするために、OpenCL, C++, Python を組み合わせ新たに作成したプログラムです。

3. FastROCS

ROCS は化合物の形状 (Shape) と、構造に含まれる化学的特徴を球で定義したもの (Color) を組み合わせ (Figure 1)、3 次元空間上で Shape と Color に基づいて 2 分子の重なりを最適化し、Shape と Color の重なりに対して 2 分子の類似性を計算します。ROCS では、Shape と Color をそれぞれガウシアン関数で定義することで、高速かつ高精度の計算を実現しています。

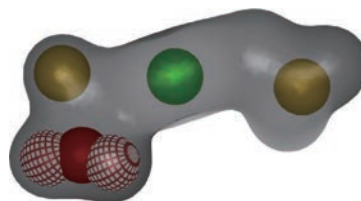


Figure 1. An example of shape (gray) and color (ball) form query.

検索対象となるデータベースに含まれる化合物の配座 (三次元構造) は、予め OMEGA³等の配座生成専用プログラムで計算しておきます。ROCS は、最初にクエリー (活

^a オープンアイ・ジャパン株式会社
連絡先 〒105-0003 東京都港区西新橋 1-6-12 アイオス虎ノ門 904
URL <http://www.eyesopen.com/>
電子メール oe_japan@eyesopen.com

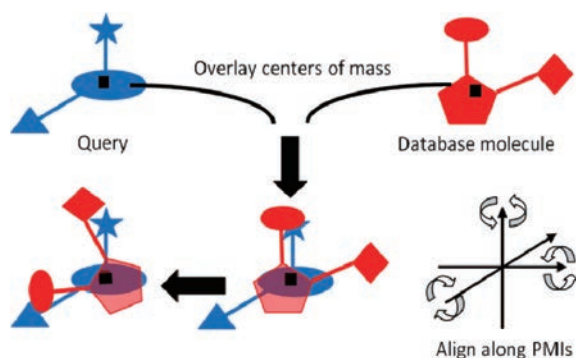


Figure 2. ROCS alignments for query and database molecule.

性既知の低分子化合物)とデータベース化合物の重心を重ね合わせた後に、データベース化合物の配座を固定したまま回転と並進のみで重ね合わせを最適化します (Figure 2)。次に、クエリーとデータベース化合物に対し Shape と Color それぞれの Tanimoto 類似係数によるスコア値 (Tanimoto-Combo) を算出することによって、クエリーと類似性の高い化合物をデータベースから絞り込みます。ROCS は、1 CPU 当たり 1 秒間に 20 から 40 化合物を処理することができ、さらに計算をクラスター上に並列化することで、大規模データベースを対象としたバーチャルスクリーニングを実用的な時間内で実現します。

FastROCS は、劇的な高速化を実現するために、これまでの CPU 上での計算を単に GPGPU に移植するだけではなく、データベースを GPGPU サーバーのメモリ (RAM) 上にすべて読み込む等の基本的な仕様から見直しています。FastROCS では、クエリーとデータベース化合物の重ね合わせを行う過程のうち、重心を重ねるまでの部分を CPU で行い、後半の最適化部分のみを GPGPU で行っています。また、クエリーとデータベース化合物の類似性計算は、Shape の類似性 (ShapeTanimoto) を GPGPU で行い、Color の類似性 (ColorTanimoto) は CPU で行い、CPU で両者の和を TanimotoCombo として算出します。

さらに FastROCS は、その高速計算性能を創薬研究現場で最大限に活用出来る様に、クライアント-サーバー型での運用とし、ネットワーク上の端末から数百万化合物のデータベースに対して計算を実行し、1 分以内に結果を得ることが出来ます。そのため、例えば OpenEye 社製グラフィックス・プログラム VIDA を利用すれば、クエリー分子の入力から結果の表示までの一連の操作を容易に実施することが可能です。

4. 計算速度

PubChem⁴ から化合物データベースを入手し、薬としての有用性の低い化合物を FILTER プログラムにより削除後

約 1000 万化合物のデータベースを作成し、配座解析プログラム OMEGA により各化合物あたり 5 配座程度を発生させてデータベースを作成しました。このデータベースを NVIDIA 社製 Tesla カードを 4 枚搭載した FastROCS サーバーの RAM 上に読み込み、VIDA をクライアントとして、RSCB PDB に登録されている 3POG⁵ のリガンド構造をクエリーとして、FastROCS 計算を実行しました。その結果、ROCS では数時間程度必要であった計算が FastROCS では約 25 秒で完了し、クエリーとは化学構造が異なりながら立体的な重なりの良い化合物を見出すことができました (Figure 3)。

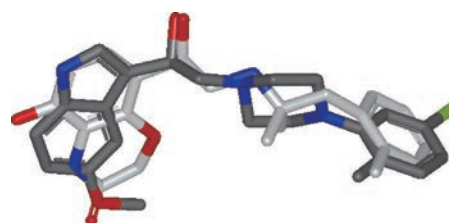


Figure 3. An illustration showing a hit database molecule (gray) overlaid on query (white).

5. まとめ

FastROCS は、活性化合物の構造にのみ基づく VS を、GPGPU により非常に高速で実現する新しいプログラムです。数千万化合物のデータベースを対象とした場合、ROCS では数時間から一晩程度の計算時間が必要であったのに対し、1 分以内での計算を可能にしました。この計算速度の劇的な向上は、単に計算時間が短縮されただけでなく、従来の 2 次元構造の類似性に基づく検索と同程度の時間で 3 次元構造に基づく類似性検索を可能とすることから、創薬研究におけるパラダイムシフトつまり薬物設計に変革をもたらすことが可能であると期待されます。

引用文献

- (1) Grant, J. A.; Gallardo, M. A.; Pickup, B. *J. Comp. Chem.* **1996**, *17*, 1653–1666.
- (2) Sheridan, R. P.; McGaughey, G.; Cornell, W. *J. Comput. Aided Mol. Des.* **2008**, *22*, 257–265.
- (3) Hawkins, P. C. D.; Skillman, A. G.; Warren, G. L.; Ellingson, B. A.; Stahl, M. T. *J. Chem. Inf. Model.* **2010**, *50*, 572–584.
- (4) <http://pubchem.ncbi.nlm.nih.gov/>
- (5) Rasmussen, S. G.; Choi, H. J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; Devree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K. *Nature* **2011**, *469*, 175–180.

(受理日 2012 年 4 月 3 日)