博　士（情報科学）　アティア　モハマド　ワヒブ　モハマド

学位論文題名

De Novo Drug Design by Evolutionary Algorithms Accelerated by GPU

(GPU 高速化による進化計算を用いた新規薬剤構造の設計)

学位論文内容の要旨

De Novo drug design approaches generate novel molecules out of building blocks consisting of single atoms or fragments. Global optimization algorithms are usually employed to search the chemical space by generating new molecular structures through probing many different fragments in a combinatorial fashion. However, combinatorial drug design is naturally limited by several pivot features; a) conventional representation methods fail to model the flexibility of the receptor in the 3D space, b) failure in biasing the generated structure towards a validated structure of similar motifs, and c) generating hundreds of thousands of poses which requires a high performance parallel system to generate complex structures.

Advanced Evolutionary Algorithms (EAs) are computationally intensive and often contain search algorithms within a search algorithm. Given their performance for complex problems, they are of great interest to De Novo ligand design. Conventionally, EAs were implemented on large clusters and super computers to reduce the execution time. However, Supercomputers are not accessible to an average person or business. Processors like Graphical Processing Units (GPUs) provide an unprecedented amount of computing power in a workstation. However, EAs in their simple form are not suitable for implementation over SIMD type architectures.

This dissertation proposes EAs designed for addressing the specific needs of De Novo ligand design while performing efficiently over GPU. The first challenge was modeling the natural phenomenon of flexible receptors in the same algorithm that searches the chemical space. The main contribution for this research includes a new encoding scheme that represents candidate solutions as a series of transformations modeling structure and/or posing search. The results first proved the functionality of combining structure/posing search in one step. Second it enabled the regeneration of structures in the ZINC database reported to have used the docking in a separate step (with even manual interaction in some cases) which improved the results by up to 1.6 times (note that this is a high percentage for a process with very high precision as binding).

In the next step, a new approach for multi-objective optimization was introduced to the De Novo ligand design. The basic idea was to combine the conventional force field with the 3D overlap of a known structure. This enabled biasing the population to known effective structures and avoiding divergence into structures not confirmed by X-ray. This approach showed essential when searching for real-world ligands. I provided a detailed empirical analysis of the effect of biasing to 3D overlaps on the volume produced and the trajectory of the search procedure. Three structures of motifs similar to structures in the ZINC database were regenerated at less than have the runtime reported in ZINC and in a single algorithm (i.e. results reported in ZINC were outcome of different programs running in

serial).

For the next step, the generation of complex structures that are of real interest to the community was identified to depend on a large database of fragments (in order of thousands) and with more than 4 rotation points in the docking process. This means using a high performance parallel system is De Facto. I introduced the use of GPU in De Novo ligand design (first to the knowledge of the author). The motivation was to combine an effective search technique with the high cost efficiency to enable high level simulations that would normally need high-cost conventional clusters with a complicated workflow of programs. The designed algorithms were carefully optimized and tuned for the SIMT architecture of GPU. The optimization did not involve just mere adaptation; new techniques and data handling patterns were used to drive the efficiency higher. Comparison with other state-of-art parallel systems showed a speedup of up to 30x in some cases at a very low cost.

Finally, I targeted the LiverX receptor which is a challenging receptor requiring a highly complex structure. The results of long simulations produced new structures when compared to structures with similar motifs in ZINC, have better binging affinity. Note that this would mean that the search for the structure and the search for the posing were both competitive. This algorithm was tested against at least one commercial docking package and the results were highly encouraging. The generated structures were of better affinity while achieving a speedup for single precision GPU implementation over 42x and the speedups for double precision implementation of over 20x.

During this research I found that when accurately modeling the specifics of binding, combinatorial optimizing can be a great tool for In Silico molecular modeling. During all the above mentioned research projects a significant consideration was given to optimize the implementation in order to maximize the hardware resource utilization.

学位論文審査の要旨

主　査	准教授	棟朝雅晴
副　査	教　授	赤間　清
副　査	教　授	大宮　学
副　査	教　授	古川正志

学位論文題名

De Novo Drug Design by Evolutionary Algorithms Accelerated by GPU

(GPU 高速化による進化計算を用いた新規薬剤構造の設計)

新規薬剤構造の設計においては、原子もしくはタンパク質の断片を組み合わせていくことで新たに有望とされる薬剤構造を探索する必要があり、困難な組み合わせ最適化問題として、大域的最適化アルゴリズムを用いた探索が行われている。大域的最適化アルゴリズムについては、先端的な進化計算アルゴリズムに関する研究が近年活発になされており、複雑な構造を有する組み合わせ最適化問題を解くことが可能になりつつあるため、それら先端的な手法を新規薬剤構造の探索に適用する試みが進められている。

また、新規薬剤構造の設計のような大規模かつ複雑な最適化問題を解く場合、並列化が重要な課題となる。従来は共有メモリ型の並列計算機や、分散メモリ型のクラスタシステムを対象とした並列化が主流であったが、近年、GPU (Graphical Processing Unit) のような演算コアを多数有するメニーコアアーキテクチャ上での大規模並列化に関する研究も進展しており、大規模かつ複雑な組み合わせ最適化問題の解決において成果を上げている。

しかしながら、これまでに提案された手法においては、分子構造の柔軟性を考慮していない、有望な構造を重点的に探索することができていない、膨大な組み合わせとなる分子の配置に関する探索が効果的にできず現実的な計算時間で解が得られていない、などの問題がある。

本論文では、新規薬剤構造の設計にあたり、先端的な進化計算アルゴリズムを採用するとともに GPU による大規模並列最適化を行うことで、有効な問題解決を実現するためのフレームワークを提案するものであり、中でも分子構造の柔軟性を考慮したモデリングと、薬剤構造の変形と配置を考慮した解の符号化に関する新規手法を開発し、GPU 上に実装している。

第 1 章では計算機シミュレーションによる新規薬剤構造の設計について、その概略を述べるとともに、GPU のアーキテクチャについて説明している。

第 2 章では先端的な進化計算および、GPU における進化計算の大規模並列化について議論している。

第 3 章ではタンパク質の構造変換に基づく対象問題のモデル化について提案している。本手法は自然界におけるタンパク質の柔軟性を考慮したものであり、その構造の変換を記述した符号化を採用することで、新規薬剤構造の探索において解の質を改善するとともに、探索効率を大幅に向上させることができる。

第 4 章では新規薬剤構造の探索のための多目的最適化問題について定式化している。具体的には、従来手法で用いられているエネルギー関数に加えて、3 次元構造の重複を基に計算されたペナルティ項を考慮したものとなっている。これにより、物理的な形状を考慮しつつ不要な探索の回数を削減することを可能としている。提案手法を現実の問題に適用し、新規薬剤構造を探索したところ、薬剤構造の有力なデータベースである ZINC データベースに登録されている構造と同様の解をより高速に得ることができる。GPU による大規模並列化を行う際、そのアーキテクチャ上の特長である SIMT (Single Instruction Multiple Threads) を最大限活用するための最適化を行うとともに、そのメモリの階層構造を活用したデータ処理方法を提案することで、アーキテクチャの構造を活用した大規模並列化を可能としている。結果として、従来型のアーキテクチャと比較し

て 30 倍以上の高速化を実現している。

第 5 章においては、さまざまな肝臓疾患に関係する LiverX receptor を対象とし提案手法を適用することで、当該疾患に対して有望となる新規薬剤構造の探索を試みている。提案手法により得られた解は、有望な新規薬剤構造の一つとして ZINC データベースへの登録が許可されており、本論文における新規薬剤構造の探索フレームワークの有効性が、実用面でも検証される結果となった。従来手法、特に商用ソフトウェアとの比較においても、GPU を用いることでより優れた構造を有する解が得られるとともに、その計算速度についても従来比 20〜40 倍の高速化を実現できている。

第 6 章、第 7 章においては、以上の結果に関する考察と結論について述べている。計算機シミュレーションによる新規薬剤構造の探索において、提案する問題のモデル化、解の符号化、探索手法、GPU による大規模並列化が極めて有効であり、現実の問題においても有用な結果が得られていることが確認できた。

これを要するに、本論文は、先端的な進化計算アルゴリズムを用いることで、複雑な探索を効果的に行うとともに、近年主流となっているメニーコアアーキクチャによる並列計算環境を活用した大規模並列化を行うことで、大規模かつ複雑な新規薬剤構造の探索において有用となる問題解決フレームワークを実現したものであり、大規模並列進化アルゴリズムおよびバイオ情報学の分野に貢献するところ大なるものがある。よって著者は、博士 (情報科学) を授与される資格あるものと認める。