



## 2008 年度下期未踏 IT 人材発掘・育成事業 採択案件評価書

### 1. 担当PM

竹内 郁雄 PM（東京大学大学院 情報理工学系研究科 創造情報学専攻 教授）

### 2. 採択者氏名

チーフクリエイター：山岸 純也(千葉大学 薬学部 薬品物理化学研究室)  
コクリエイター：なし

### 3. プロジェクト管理組織

株式会社オープンテクノロジーズ

### 4. 委託金支払額

2,994,293 円

### 5. テーマ名

GPGPU を用いた薬物親和性評価プログラムの開発

### 6. 関連Webサイト

<http://bukka.p.chiba-u.jp/>

### 7. テーマ概要

新規医薬品の開発は大きく分けて

①疾患ターゲット(原因タンパク質)の探索、②医薬品となりうる化合物の探索・最適化、③臨床試験による評価の3段階に分けられる。現在、ひとつの新規医薬品の開発には十数年の歳月と数百億円の投資が必要とされるが、今後はより複雑な疾患・

化合物を取り扱う必要に迫られ、経費・期間ともに増加の一途を辿ると言われている。

現在、製薬業界では②「医薬品となりうる化合物の探索」において、原因タンパク質と医薬品候補化合物との親和性をコンピュータ上で評価する *in silico* 創薬を利用する動きが高まってきている。それは「数百万種に及ぶ化合物や実験器具・試薬が不要であること」と「探索時間が短縮できること」を期待してのことだが、実際には「計算時間が長い」「計算結果と実験的に求められる親和性の相関が十分でない」といった問題があり、目的が十分達成されていない。これらの問題を解決するのが本開発の目標である。

GPGPU とは GPU (グラフィックボードに搭載されたプロセッサ) を一般論理演算に用いる手法である。GPU は並列計算可能なプロセッサが千単位で搭載されており、超並列計算に優れているが、並列計算の扱い難さやメモリ関連の制約により、現在のところ十分な計算速度を出せるプログラムは限られている。

本開発におけるプログラムは、原子に対する物理化学的相互作用の寄与を計算するものであり、それらはそれぞれ独立した式から構成されているため、まさに GPGPU に適したプログラムとなっている。その計算能力を活かし、あらゆる寄与を計算することで評価関数の改善を図る。

また遺伝的アルゴリズムなど適応学習アルゴリズムを取り入れることで、候補化合物がどのような内部構造で原因タンパク質に結合するか (配座探索) を効率的に探索する。

## 8. 採択理由

未踏ユースが年 2 期制になってから、捲土重来、つまり 1 回目に落ちて 2 回目の挑戦をする人が増えてきた。山岸君もその 1 人である。1 回目と 2 回目の間は半年程度の開きしかないが、さすが若い人の成長は著しい。未踏ユースに応募するだけでも人材育成の効果があると言ったら言いすぎか。不採択コメントは、PM たちがそれなりの意図をもって書いていることを知ってほしい。

閑話休題。山岸君は GPGPU を使いこなす実力を、半年前に比べて格段につけてきた。竹内の持論であるが、コンピュータでいい仕事をする人の多くはコンピュータが専門でない人たちである。山岸君は、2006 年度に未踏ユース・スーパークリエイターになった藤秀義君の後輩にあたる薬学部の学生で、薬学が専門である。最近、薬物設計 (創薬) はナマ物を扱う (*in vitro*) ほかに、コンピュータの力を借りないと (*in silico*) 効率が上がらないということがわかってきた。つまり、いかに効率よくコンピュータで候補となる化合物を絞りこむかが鍵となる。現状では問題が山積しているが、だからこそ、どんな手段であれ、目の前の壁を乗り越える努力が必要である。グラフィックスの

高速化が目的であるがゆえに、大量に出回り、かつ開発努力が継続している GPGPU を使って、分子とタンパク質の間の物理化学的親和性の計算に転用しようというアイデアは面白い。予備実験で、すでに CPU に対して 2 桁程度の高速化が達成されているという。未踏ユースには GPGPU 使いの先輩も多いので、支援も受けられるだろう。

創薬はいまや世界中の企業や研究機関が競っている分野であるから、このプロジェクトによって山岸君の成長だけでなく、注目すべき成果が出ることも期待したい。

## 9. 開発目標

本プロジェクトでは GPGPU を用いて、薬物親和性評価の予測を高速かつ高精度に行うことを目的とする。スクリーニングソフトウェアは性質上、並列計算で高速化することが容易であるため、GPGPU を用いて高速化を図ることが出来る。また GPGPU の導入コストは並列クラスタや既存ソフトウェアのライセンス料よりはるかに安価となっている。GPGPU を用いることで個人向けの PC でもスクリーニングを行うことが可能になり、あらゆる研究グループで日常的に創薬を行うことが可能になると期待される。

## 10. 進捗概要

本プロジェクトでは科学計算の分野でも注目を浴びている GPGPU を用いてスクリーニングを高速に、かつ高精度の計算を行うことに成功し、スクリーニングソフトにも GPGPU が有効な選択肢であることを示した。既存のソフトウェアではたとえ PC クラスタを用意したとしてもライセンスの問題等から並列化への対応は難しかったが、本ソフトウェアを用いることで PC クラスタのみならず家庭用 PC でも実用レベルのパフォーマンスを引き出すことができ、創薬の可能性が広がることは間違いない。

CPU との速度比較より、現行のソフトウェアよりも莫大な計算量が必要な計算でさえも短時間で行えることを証明した。本来なら計算速度の問題等から避けたい全原子計算、水分子の問題も、アクセラレータを用いることで解決できることも示した。しかし実在する水分子の取り扱いはまだ発展途上であり、これから改善する必要がある。また今回用いた分子力学(MM)計算は分子動力学(MD)シミュレーションなどの計算にも用いられる手法であり、それらの分野への応用も容易である。

## 11. 成果

nVidia から提供されている GPGPU 用の言語 CUDA を用いて薬物親和性評価プログラムを開発した。薬物親和性評価プログラムとは化合物とタンパク質の三次元デー

タを用いて「どのようなポーズで結合するか」、「どのような強さで結合するか」を予測するソフトである。

通常、化合物は回転可能結合を持ち、同一の化合物でも様々な形を取ることができる。プログラムはまず化合物がどのような自由度をもつのかを判定し、様々なポーズで親和性を評価することで、どの形が強く結合しやすいのかを予測する。親和性はエネルギー変化で表わされる。エネルギー変化とは化合物・タンパク質がそれぞれ水中に存在するときのエネルギーと、複合体で存在するときのエネルギーの差である。通常、スクリーニングではタンパク質は同じ構造を用いるので、タンパク質が単独で水中に存在するときのエネルギーは計算する必要がない。しかし化合物が単独で水中に存在するときのエネルギーを計算しないソフトウェアも多く、精度の低さに結びついている。エネルギーの計算には分子力学 (Molecular Mechanics: MM) を用いる。MM は力場と呼ばれる原子の種類 (原子タイプ) に依るパラメーターセットを用いてエネルギーの算出を行なう。本プロジェクトでは精度の向上を目指して、全原子の MM 計算を行ない、結合親和性の予測を行なった。また試験的にタンパク質の周囲に水を発生させた状態での計算や、タンパク質の自由度を考慮した計算も取り入れる。

プログラムは大きく分けて、ポーズ探索部、エネルギー極小化部の 2 つに分けられる。ポーズ探索ではどのポーズの結合親和性が最も高くなるかを遺伝的アルゴリズム (GA) を用いて探索する。ポーズ探索では、ある一定の刻み幅での探索となり、加えて水分子を含んだ親和性評価の場合は水分子を固定してポーズ探索を行なうため、最後にエネルギーの微分方向 (力のかかる方向) に原子を動かし微調整することが必要となる (エネルギー極小化部)。本プロジェクトではそれぞれのプログラムにおいて GPGPU を用いて計算の高速化を行った。

GPGPU を用いることで家庭用の PC でも実用レベルでソフトウェアの動作が可能となる。現在は高性能 PC クラスタや高額なライセンス料を要するという費用面の問題から、大量の化合物の短時間のスクリーニングは限られた環境でのものとなっている。しかし、家庭用 PC でもアクセラレータなどを用いた高速化がすすめば、スクリーニングソフトウェアが実用レベルで使える環境になる。

最終的には外部プログラムに頼らないソフトウェアの開発を目指す。開発期間内では電荷の設定に関しては外部プログラム antechamber (AMBER) を用いた。

開発した主要プログラムと、それらの流れの模式図を図 2.1.1 に示す。図の中の赤い部分が GPGPU を使った部分である。

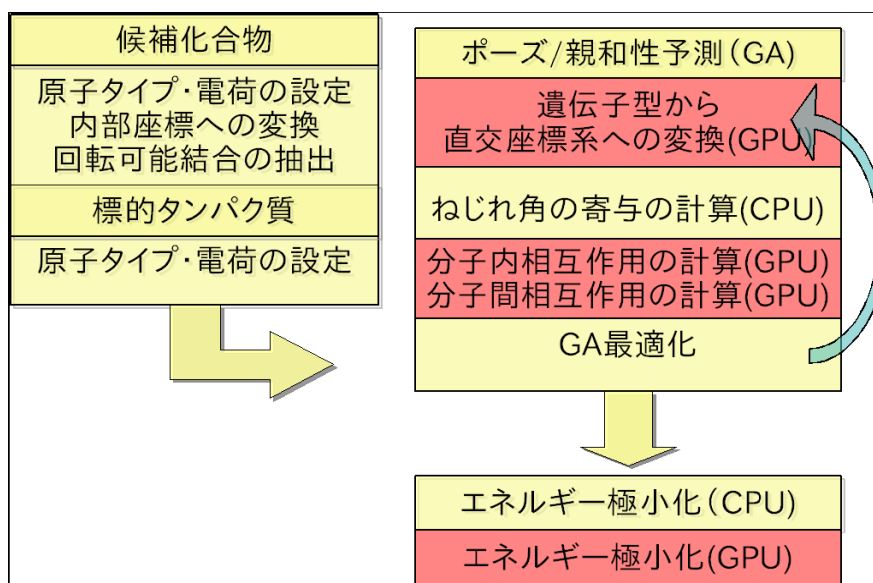


図 2.1.1 主要プログラムの模式的な流れ

個々のプログラムの詳細については省略するが、従来の親和性評価プログラムでは近似的に省略していたような少し遠い原子間の相互作用や、タンパク質と薬物の間に介在する水分子の影響を計算に取り込むなど、分子力学の本義に立ち戻ったプログラムを開発した。その結果、むしろ精度が落ちるという現象も生じたが、今後、改良できると考える。

このような本格的な分子力学計算を行なえるようになったのも、GPGPU の能力を最大限に活かしたからである。実際、GPGPU と CPU (single thread) との比較をすると、結合ポーズ探索では表 2.1.1、エネルギー極小化では表 2.1.2、プログラム全体としては表 2.1.3 となっている。ここで CPU×4 は、OpenMP を用いて CPU を 4 並列で実行させた結果（参考）であり、GPGPU\* は CPU と GPGPU を非同期に同時実行させて計算時間を短縮した結果である。なお、実験環境は

- ・タンパク質: HIV-1 Reverse Transcriptase
- ・タンパク質の原子数: 3784
- ・候補化合物の原子数: 44
- ・水分子の数: 4000 前後
- ・GA 実行回数: 8
- ・GA の個体数: 128
- ・GA の世代数: 200
- ・CPU: Intel Core2 Quad 3.0GHz
- ・GPGPU: nVidia Geforce GTX 295

である。

表 2.1.1 結合ポーズ探索プログラムの実行速度の比較

	CPU	CPUx4	GPGPU	GPGPU*
遺伝子型から直交座標への変換	14.10 ms	3.73 ms (3.78 倍)	1.88 ms (7.5 倍)	-
分子内非結合性相互作用の計算	5.07 ms	1.41 ms (3.60 倍)	0.726 ms (6.98 倍)	-
分子間非結合作用の計算(化合物-タンパク質)	953.60 ms	258.84 ms (3.68 倍)	5.70 ms (167.3 倍)	-
分子間非結合相互作用の計算(化合物-水)	109.20 ms	30.57 ms (3.57 倍)	2.50 ms (43.68 倍)	-
二面角ポテンシャルの計算	3.16 ms	1.27 m (2.49 倍)	(3.16 ms(CPU))	0 ms
次世代の選択	0.106 ms	-	-	-
合計(非水和環境)	965.96 ms	268.01 ms (3.60 倍)	10.39 ms (92.97 倍)	8.175 ms (118.1 倍)
合計(水和環境)	1086.14 ms	296.89 ms (3.66 倍)	13.70 ms (79.28 倍)	10.33 ms (105.1 倍)
200 世代実行時間(非水和環境)	192820 ms	No data	No data	1632 ms (118 倍)
200 世代実行時間(水和環境)	214049 ms	No data	No data	2082 ms (102 倍)

表 2.1.2 エネルギー極小化プログラムの実行速度の比較

	CPU	GPGPU
結合長(化合物)	0.004 ms	-
結合角(化合物)	0.020 ms	-
二面角(化合物)	0.070 ms	-
分子内非結合相互作用(化合物)	0.087 ms	-
分子間非結合相互作用(化合物-タンパク質)	12.97 ms	0.459 ms (28.3 倍)
結合長・結合角(水)	0.44 ms	-
分子間非結合相互作用(化合物-水)	12.89 ms	0.687 ms (18.8 倍)
分子間非結合相互作用(タンパク質-水)	868.95 ms	4.72 ms (184.1 倍)
分子間非結合相互作用(水-水)	632.53 ms	8.63 ms (73.3 倍)
その他の処理	0.002 ms	-
合計(非水和環境)	13.04 ms	0.641 ms (20.3 倍)
合計(水和環境)	1536.70 ms	15.00 ms (102.4 倍)
1000step 実行時間(非水和環境)	13080 ms	630 ms (20.7 倍)
1000step 実行時間(水和環境)	1519753 ms	15906 ms (95.5 倍)

表 2.1.3 プログラム全体の実行速度比較

	CPU	GPGPU	GPGPU*2
非水和環境	27.6 分	20.7 秒 (80.0 倍)	10.8 秒 (153.3 倍)
水和環境	3.89 時間	2.46 分 (94.9 倍)	1.26 分 (185.2 倍)

表を見てわかるように全体として 100~200 倍の速度が出ていることがわかる。これは計算量の多い上位 3 項目を GPGPU に移植したことや、また GPGPU と CPU の計算を非同期並列に実行させることで、二面角ポテンシャルの計算時間を完全に隠蔽することに成功したことが効いている。

ほぼすべての計算部分を GPGPU に移植したので、1 世代あたりの計算時間のうち 99% 以上を GPGPU 上での計算に割り振れた。これは GPGPU をアイドルにすることなく効率的に使えていることを示している。これは、ほぼすべての計算を GPGPU が担っているので 1 個の GPGPU あたり 1 個の CPU で十分であり、それを超える並列化が必要ないことを意味する。CPU の計算部分が少ないので、CPU のスペックにそれほどこだわらなくてもよい。本プロジェクトの目的である家庭用 PC での実用レベルの動作を十分に達成できたわけである。

CPU と GPGPU の対を増やしたときに、数効果が得られることも確認できた。これは計算の独立性が非常に高いゆえである。

計算精度については、RMSD という指標で評価する。RMSD はそれぞれの原子の距離の二乗の平均値の平方根で、2 以下だとかなり構造を再現できていると言え、4 以上だとあまり再現できていない。実験した 6 個の複合体中、4 個において RMSD が非水和環境でも 2 以下となり、実験のポーズを再現することに成功した。しかし残り 2 個は非水和環境では RMSD が 4 以上とあまり再現性はよくない。

水分子を発生させることでポーズ探索において改善が見られると考えていたが、全体として改善は見られなかった。ほとんどの複合体で化合物が水を求めて結合部位から外れてしまうという結果が起こった。このような結果から、水と化合物の相互作用だけではなくタンパク質と水との相互作用も考慮しなくてはいけないことが示唆される。化合物が結合部位から外れるということは、逆に結合部位には水が配置するということになるため、タンパク質が不安定になることは薬学的に推測できる。今後はこのような寄与の計算も実装し、精度の向上に努める必要がある。

さらに、計算で得られた親和性と、実験で得られた親和性との相関も調べたが、それほど良い結果が得られなかった。原因としてタンパク質と水の相互作用を考慮していないことが考えられる。また本ソフトウェアでは結合部位の中心からある距離までの水分子を発生しているが、実在の系では水は無限遠まで広がっている。クーロンポテンシャルの式からかなり遠くの水分子まで考慮しない限り水分子の影響は収束しない（無視できない）ことがわかり、その影響も誤差に含まれる。今後は水分子のミクロな系は水分子を発生させて考慮し、マクロな水（遠方の水）は連続体として近似する必要も出てくるかもしれない。精度向上については今後さらなる改良が必要である。

## 12. プロジェクト評価

本来グラフィックスのために開発された GPGPU を、高並列マシンとして、それ以外の分野の計算に応用しようという動きが盛り上がっている。nVidia が CUDA という GPGPU 用の言語を開発・公開したことでそれが一気に加速した。山岸君のプロジェクト



トは、これを in silico 創薬（コンピュータを利用して疾患ターゲットのタンパク質と候補化合物の親和性を評価して薬物を発見すること）に応用したものである。いまや、in silico 創薬の方法論に頼らない in vitro 創薬（実際に薬物を使って試験する方法）では、時間もコストもかかりすぎる。しかし、その in silico 創薬といえども、計算時間は膨大である。

まだ発展途上の CUDA を使いこなすのも容易ではないのに、分子間の親和性評価の基本的な方法を熟知して、真正面から正攻法でプログラムの並列化と高速化に取り組むためには薬学の専門家である以上の才能が必要である（山岸君は未踏期間中にちゃんと薬剤師国家試験をパスした）。親和性評価のプログラム（これによって実際に in vitro で試す薬物の候補を劇的に減らすという意味でスクリーニングプログラムとも呼ぶ）はすでにいくつかあるが、高価、遅い、精度が低いなどの問題がある。山岸君の挑戦は家庭用の PC に最新の GPGPU を組み込めば、フリーで使えるスクリーニングソフトを開発することである。これができれば大学等の研究機関には大きな朗報である。

真正面から正攻法というのが、実はこのプロジェクトの肝である。従来のスクリーニングソフトはあまりにも多い計算量を減らすために、かなり乱暴な近似を行なっている。山岸君の発想は、親和性評価のようにそれぞれの分子間の相互作用の計算の独立性の高いものは、GPGPU を使えば並列度が最大限に上がるはずだから、変な近似を行なわなくても、正攻法できちんと計算ができるはずだというものである。だから、分子力学の真っ当な式をそのままプログラムに組み込んだ。

GPGPU の使いこなしもすばらしかった。当初はまだ手に余しているところもあったが、ぐんぐん習熟し、最終的には CPU と GPGPU をうまく並列連携させて、そうでなかったときよりも 2 倍の速度向上を果たし、トータルで CPU 単独の 150~180 倍の計算速度を出した。一般的な応用プログラムでここまでの倍率を出したことは高く評価できる。特に CPU と GPGPU の並列連携は並みのプログラマではできないレベルのものである。CUDA のバージョンアップに対応して、今度は Mapped Memory も利用してさらなる高速化を図るとのこと。

極めて安価に、かつ高速な親和性評価の計算が行なえるところまでは行ったが、残念ながら精度については既存のソフトウェアをまだ超えていない。しかし、これはあくまでも未踏期間中ということであり、真正面から正攻法でアタックしているかぎり、精度が向上するのは時間の問題であろう。特に、すでに取り掛かっている水分子の影響や、標的タンパク質の変形まで考慮するという、まさに前人未踏の挑戦を、山岸君は行なっている。この方針が間違っているわけがない。もっとも、究極的に精度を上げなくても、数百万種の化合物をたとえば実質 1/10000 以下までにスクリーニングできれば、in silico 創薬としては十分とも言えるので、精度向上のための改良は近い将来に一件落着するのではないかと期待している。

思えば、平成 20 年度上期に応募してきたときは、まだ GPGPU をあまり理解していなかったと思われる山岸君が、ここまで伸びてきたのは驚異的である。このプロジェクトに関しては、GPGPU のプロである電通大の大島聡史君（未踏ユース OB、現在同大産学連携研究員）にレビューのとき同行・助言してもらった。こういう形の縦の連携は今回初めてであったが、山岸君には大きな刺激になったと思う。

これまで開発に余念がなかったが、なるべく早く一旦きちんとまとめて、国際会議等で発表する戦略をとるとよい。精度はもうちょっと上がったほうがいいかもしれないが、高い評価を受けるはずである。ソフト自体の公開が企画に乗っているとのことだが、このような発信は素晴らしいことである。

### 13. 今後の課題

研究室の Web サーバーにソフトウェアを無償公開(URL: <http://bukka.p.chiba-u.jp/>)し、このソフトウェアを用いた創薬研究を行い、論文の執筆や学会発表を繰り返していくことで、ソフトウェアの有用性を検証し、普及させることに努めたい。