

課題名 (タイトル) :

## 生体高分子生化学機能解析のための分子計算技術の開発

利用者氏名 : ○中村 春木<sup>1,2)</sup>, 山中 秀介<sup>2)</sup>所属 : 1) 和光研究所 次世代計算科学研究開発プログラム  
2) 大阪大学蛋白質研究所附属プロテオミクス総合研究センター

## 1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

現在ライフサイエンス分野において、生体高分子の量子力学(QM)シミュレーションを起点とした分子動力学(MD)さらには粗視化モデル(CGM)を用いた研究が盛んに行われている。生命を記述するのにいわば最も高い解像度である多電子系としての分子の記述から、ボトムアップ式に巨大な生体内イベントを記述する為には、大規模な超並列計算が可能なプログラムの開発が必要である。この目的の為、我々は、「次世代生命体統合シミュレーションソフトウェアの研究開発」(生体高分子生化学的機能解析のための分子計算技術の開発)の一環として、QM/MD/CGM 統合プログラム Platypus を開発している。本課題は、この Platypus の超並列計算を実施し、その並列性能を解析・吟味し、さらなる高速化・高度化に資する事を目的としている。

## 2. 具体的な利用内容、計算方法

QM/MD/CGM 統合プログラム Platypus のうち、最も計算時間を要するのは量子力学(QM)部分である。このため我々は Platypus の QM 部分、Platypus-QM の超並列化を進めているが、今回、この超並列性能を解析したので、以下に報告する。

我々が目的としている超並列計算の確認・吟味には、RICC での並列計算が必須である。簡易利用では 2000 コアまでの並列計算を実施し、さらに大規模並列ジョブの利用により、8000 コア×30 分の CPU 時間を確保していただき、最も汎用的に使用される事が予想されるスピン制限密度汎関数理論(RDFT)の計算を実施した。この Platypus-QM は他の種類の QM 計算も可能であるが、RDFT の数値計算パーツは分子積分や DFT

数値積分といった、他の計算法との共通部分も多いため、下記結果は Platypus-QM の並列性能の主たる部分の解析にほぼ対応している。

## 3. 結果

2010年1月から2月に行った並列性能解析を、その間の改善点も含め、時系列に沿って記述する。

## 【1月】

計算対象系は光合成活性中心バクテリオクロロフィル a とそれに配位するヒスチジン側鎖の 2 量体/基底関数 6-31G(原子軌道=1624 次元)である。Platypus-QM の主たる部分の並列計算時間および並列化率を表 1 に上げる。

表 1: Platypus-QM(1月)の各部分の計算時間 (秒) および並列化率(8000 コア/256 コアで算出)

プロセス	コア数					並列化率
	256	512	1024	2000	8000	
<i>Vxc</i>	54.5	27.5	14.4	8.1	3.1	99.99
<i>dxs</i>	25.8	15.0	10.7	8.3	6.3	99.89
<i>ERI</i>	455.0	450.5	448.5	492.9	493.4	-
<i>Swz</i>	186.4	186.4	187.7	206.5	208.0	-
<i>dERI</i>	532.6	495.1	483.2	474.7	469.9	-
全体	1155.0	1066.9	1034.1	1063.9	1280.2	-

*Vxc*: DFT の交換相関項、*dxs*: DFT の交換相関項の勾配、*ERI*: 二電子積分、*Swz*: シュワルツ公式による積分スクリーニング、*dERI*: 二電子積分の勾配

表 1 の通り、この段階での Platypus-QM コードでは、*ERI*, *Swz* 等が特に並列性能を低下させ、1024 コアから 2000 コアにコア数を増加させると逆に計算時間が増加し、並列化が頭打ちとなった。しかしながら DFT 部分は高い性能効率を達成している事が判明した。

この 1 月の結果をもとにコードを解析し、2 電

子積分部分のループ構造と分岐文の関係などを改善し、2月の末に再度8000コアを用いた超並列計算を行った。

【2月】

2電子積分部分の並列性能が劣化した主因を取り除き、さらに計算対象の原子数はそのまま基底関数(電子構造記述の解像度)を大きくし、光合成活性中心バクテリオクロフィル a とそれに配位するヒスチジン側鎖の2量体/基底関数6-31G\*(原子軌道=2488次元)に対する並列化性能を測定した。また、1月の計測では、SCFを6回としたが、系が大きいため、2月の計測では1回のみSCFとした。主たる部分の並列計算時間および並列化率を表2に示す。

表2:Platypus-QM(2月)の各部分の計算時間(秒)および並列化率

プロセス	コア数			並列化率	
	1024	2000	8000	(1)	(2)
Vxc	4.64	2.66	2.43	99.99	99.92
dExc	11.5	8.85	13.8	99.89	-
ERI	15.3	12.7	12.9	99.99	99.92
Swz	0.08	0.09	2.62	-	-
dERI	94.2	56.9	38.7	99.98	99.95
全体	202.7	146.5	140.4	99.93	99.82

並列化率は (1)2000 コア/1024 コアおよび (2)8000 コア/1024 コアで算出

表2では、基底関数の数が表1の場合の1.5倍になっているにもかかわらず、実行時間が大幅に減少した。特に8000コアの超並列計算においては、2電子積分の勾配が大幅に改善される一方、DFT部分が交換相関項およびその勾配が、ともに並列性能において低下した。

これらのDFTの並列性能劣化に関しては原因を解析中であるが、MPI\_REDUCEを実施している部分と、共通で基底関数の2乗×電子数に依存した非並列計算部分が残っていることが原因と考えられるため、対処を始めている。

4. まとめ

現段階で、RICCの簡易利用で実施可能な最大並列計算規模、8000コアに対し、Platypus-QMは並列化性能を劣化させる事なく実施できる事を確認した。また、各部分の計算実測値から、高速化の余地のある部分を洗い出した。

5. 今後の計画・展望

2010年2月の時点で、実質8000コアまでの並列化が達成された。その並列性能向上は並列化率にして99.821%であり、さらなるチューニングの余地もある事が分かった。

シュワルツ部分やDFTの勾配部分など、既に性能劣化の原因の解析を開始しており、この解析を反映した新規コードを用いてさらに並列性能を向上させ、性能計測を実施する予定である。

6. RICCの継続利用を希望の場合。これまで利用した状況(どの程度研究が進んだか、研究においてどこまで計算出来て、何が出来ていないか)や、継続して利用する際に行う具体的な内容

上記結果の通り、当初2010年1月の時点では、1000並列で並列性能向上が頭打ちしていたPlatypus-QMを改良し、8000コア並列計算を達成した。来年度も今年度の解析・検証に基づきさらにPlatypus-QMの性能向上を行いたく、2010年度も継続利用を希望する。当面の目的としては並列化率99.99%を目標としており、QM部分の並列化向上と並行して、Platypus MDとの連成計算(QM/MD)を今後計画している。

7. 利用研究成果が無かった場合の理由

現時点での目的が生命科学の特定の現象の解析を目的としたシミュレーションではなく、それに先立つプログラムの超並列へ向けたチューニングであったため、科学的な成果は現段階ではない。しかしながら、8000コアの並列化を計算速度が飽和せずに実施できるQMコードは世界的に見ても数少なく、科学研究基盤としてのQMコードの開発としては、準備段階的な成果は達成されていると言える。