

MDGRAPE-3 システムを用いた大規模タンパク質 MD シミュレーションの性能評価

菊川 豪太^{*1}, 米澤 康滋^{*2}, 中村 春木^{*2}, 神谷 成敏^{*3},
泰地 真弘人^{*4}, 姫野 龍太郎^{*5}

^{*1} 東北大学 流体科学研究所

^{*2} 大阪大学 蛋白質研究所

^{*3} 神戸大学 医学系研究科

^{*4} 理化学研究所 ゲノム科学総合研究センター

^{*5} 理化学研究所 次世代スーパーコンピュータ開発実施本部

1. 緒言

次世代スーパーコンピュータの実現によってもたらされる科学分野への貢献は、多大かつ広範にわたるものであるが、特にライフサイエンス分野においては、医学・医療応用、あるいは生命現象に対する基礎的理解といった重要な問題を内包しているため、その実現が渴望されている。例えば、創薬の際の *in silico* スクリーニング、薬物ドッキングのシミュレーションに対しては、より高精度かつ高速な探索が可能となり、必然的に大規模なシミュレーションとなる膜タンパク質等の機能構造解析に用いられれば、長大な時間・空間スケールのダイナミクスの解析を可能とすることが期待できる。ここに改めて述べるまでもないが、これらミクロスケールの生命現象の理解が、細胞・組織といったメゾスケール、あるいはマクロスケールの生体現象、生体機能の解析に有用な情報をもたらす、ライフサイエンス研究の新たな展開につながると考えられる。

本研究では、次世代スーパーコンピュータの基盤開発に合わせ、特にミクロスケールにおける、ライフサイエンス分野の応用アプリケーションに対する性能評価や将来展望を行うことを主眼として研究を進めている。ここでは、現在、理化学研究所ゲノム科学総合研究センター高速分子シミュレーション研究チームによって開発が行われている分子動力学専用計算機 MDGRAPE-3^[1,2]を用いて、大規模タンパク質系の MD 計算の性能評価を行った。この際、基盤アプリケーションとして、大阪大学蛋白質研究所（バイオ産業情報化コンソーシアム, JBIC）にて開発が進められている蛋白質構造予測シミュレーションパッケージ myPresto^[3]（特に MD 計算プログラム *cosgene*）を用いて開発を行った。

大規模な MD 計算の最大の障害となるのが、多体間に作用する相互作用の評価に膨大な時間を費やされることである。古典 MD 計算においては、ポテンシャル関数のモデル化を行って計算を行うが、この際、一般的に 2 体で表現される分子間相互作用（Lennard-Jones 相互作用, Coulomb 相互作用等）が計算時間のほとんどを占める。これらの 2 体間相互作用の計算を専用計算機に任せ、飛躍的な計算時間の向上を図るのが MDGRAPE システムの狙いである。ここでは、MD シミュレーションの大規模並列計算において特に問題となる、周期境界条件下での Coulomb 相互作用の計算手法に焦点を当て、MDGRAPE を組み合わせた性能評価を詳細に行った。

2. MDGRAPE システムとコード開発の概要

本研究に用いた計算システムは、理化学研究所ゲノム科学総合研究センターの MDGRAPE-3 クラスタシステムである^[1,2]。システムの概略図を図 1 に示す。今回の計算では、12chip を搭載した 1 MDGRAPE ボードと 1 ホストでの計算、さらに InfiniBand で連結された複数のホスト CPU を用いた並列計算を行った(詳細は後述)。MDGRAPE ボードは、単一のボードで、約 2TFlops の公称性能を発揮する (250MHz 動作)^[2]。ホストマシンについては、Intel Xeon 5150 (2.66GHz) CPU を搭載した計算機を用いた。(MDGRAPE やその他システムの詳細については、^[1,2]を参照)

MDGRAPE システムにて、MD 計算が行えるように、ゲノム科学総合研究センターより提供されているライブラリ (m3 ライブラリ系)^[1]を利用し、myPresto パッケージ^[3]の cosgene プログラムを改変した。改変によって、現在、MDGRAPE 上で以下の MD 計算手法が実行可能になっている。

- ・ Cell-index 法を用いた MDGRAPE 計算 (MDGRAPE システムによりサポートされている)
- ・ 直方体セルを含む周期境界条件の計算
- ・ ビリアル計算 (圧力テンソル計算) および NPT アンサンブル計算への対応
- ・ 1 MDGRAPE ボードーホスト CPU 並列計算 (詳細は後述)

以上のようなシステムで、水溶性タンパク質および膜タンパク質の MD 計算を行い、MDGRAPE によって、どの程度の性能向上が得られるかを評価した。

3. 計算方法

3.1 性能評価の方法

本研究では、周期境界条件下におけるタンパク質-溶媒系の MD 計算に着目し、MDGRAPE の性能評価を行ったが、以下の点を踏まえて評価した。

- 1) MDGRAPE ボードを使用する場合とそうでない場合
- 2) Coulomb 相互作用の計算法による違い (次節参照)
- 3) ホスト CPU 計算の並列化を行った場合

本稿では、計算時間に関するデータは 1 timestep あたりの計算時間の平均値として示している。また、この際のサンプルは、条件によっても異なるが、概ね 200 step~500step

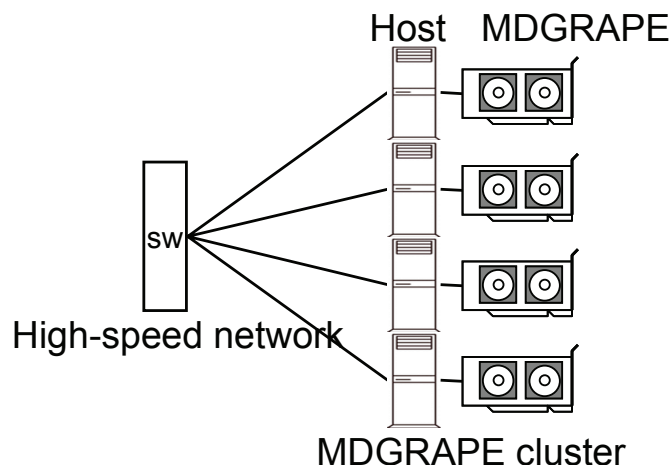


Fig.1 Schematic diagram of MDGRAPE cluster system

程度を用いて評価している。

3.2 Coulomb 相互作用に対する Wolf 法^[4,5]の導入

MD 計算において常に問題となるのが、周期境界条件下での Coulomb 相互作用の取り扱いである。ポテンシャルが相対距離の $1/r$ に比例する関数であるため、非常に収束が遅く、直接 2 体積をとって評価することは現実的ではない。一般的には、周期性を考慮して、現実的な計算時間で（数学的に）無限遠方までの相互作用を評価できる Ewald 法^[6]が用いられている。しかし、厳密には、この手法も粒子数 N に対し、計算時間が $N^2/3$ に比例するため、大規模計算では、現実的でなくなる。最近では、Ewald 法における波数空間の計算を、FFT を用いて高速化する PME 法^[7]が広く使われており、この方法では、計算時間が $N \log N$ に比例する（概ねオーダー N ）。したがって、大規模系に対しても有効な手法であり、本研究においても参照データは全て PME 法を用いた。しかしながら、PME 法も、大規模並列計算においては、内部的に使用している FFT の並列化が障害となり、高い効率が得られない。

これに対し、Wolf ら^[4]によって提案された Coulomb 相互作用の計算は、2 体間相互作用のみによって評価する方法である。通常 2 体間相互作用のみで計算する場合、現実的な距離で相互作用をカットオフする必要があるが、これを行うと著しく精度が落ちるといった問題点があった。Wolf らは、相互作用のカットオフを行いながら、ポテンシャル評価の精度を得る方法を導入し、これまでにその有効性が単純なイオン結晶や液体で示されている^[8,9]。

本研究では、Zhan ら^[5]によって修正された表式（以下、修正 Wolf 法）を導入して、その計算法による精度、および性能評価を行った。Wolf 法は、波数空間の演算がなく、2 体間相互作用のみで表現されるため、1)並列化性能が得やすい、2)MDGRAPE での演算が可能である、という利点を持っている。ここでは、MDGRAPE での評価を中心に示す。

修正 Wolf 法のポテンシャルの表式は、以下のようになっている。

$$V(r) = q_i q_j \left[\frac{\operatorname{erfc}(\alpha r)}{r} - \frac{\operatorname{erfc}(\alpha R_c)}{R_c} + \left(\frac{\operatorname{erfc}(\alpha R_c)}{R_c^2} + \frac{2\alpha \exp(-\alpha^2 R_c^2)}{\pi^{1/2} R_c} \right) (r - R_c) \right] \quad (1)$$

ここで、 q_i は粒子の電荷、 r は粒子間距離、 R_c はカットオフ半径、 α は収束パラメータであり、式の形からもわかるとおり、Ewald 法における α パラメータと同じ意味を持っている。すなわち、実空間での関数の収束性を決定している。要約すると、修正 Wolf 法では、通常の Ewald 法の実空間部分の計算に対し、カットオフ距離におけるソフト項（括弧内第 2 項）とダンプ項（括弧内第 3 項）が付与されている。波数空間の寄与を無視しており、基本的に α パラメータが小さく、カットオフ半径 R_c が大きいほど精度が得られるような構造になっている。

3.3 計算条件

MDGRAPE による性能評価、修正 Wolf 法の精度評価に、次の 2 つの計算系を用いた。

- 1) 水溶性タンパク質 BPTI-水溶媒系（以下 BPTI 系）
- 2) 膜タンパク質 MAO-A-POPC 脂質膜-水溶媒系（以下 MAO-A 系）

それぞれの計算系のスナップショットを図 2 に示す。BPTI 系は、BPTI、水分子 (TIP3P モデル) 6815 分子、Cl⁻ カウンターイオン 6 原子、計 21339 原子の系である。MAO-A 系は、MAO-A、FAD、POPC 脂質 163 分子、水分子 (TIP3P) 17125 分子、Cl⁻ 78 原

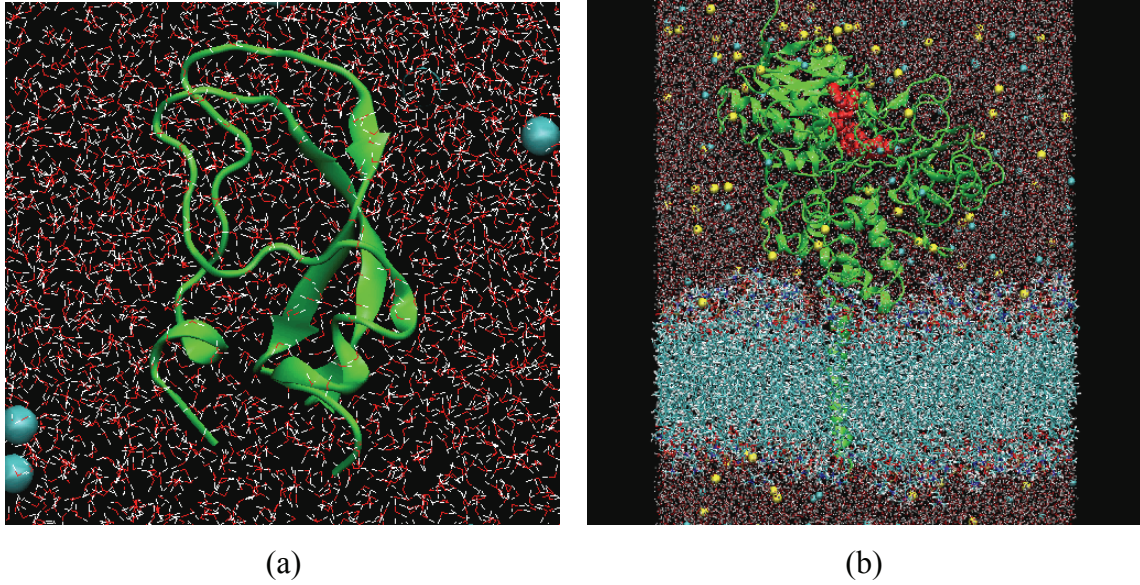


Fig.2 Snapshots of computational system; (a) BPTI system, (b) MAO-A system. Both graphics are drawn by using VMD molecular graphics^[10].

子, Na^+ 76 原子, 計 81647 原子によって構成されている. それぞれ, ポテンシャルパラメータには, AMBER および CHARMM 力場を用いている. セルサイズは, それぞれ, 約 $60 \times 60 \times 60 \text{ \AA}^3$, $70 \times 90 \times 120 \text{ \AA}^3$ となっている. また, 温度条件は, それぞれ, 300K, 310K を用いた.

4. 結果と考察

4.1 修正 Wolf 法の精度評価

まず, 修正 Wolf 法の精度を確認するため, 式(1)における α パラメータとカットオフ半径 R_c を変化させて, 同一のコンフィグレーションでの PME 法の結果との誤差を調べた. 図 2 には, BPTI 系でのポテンシャルの相対誤差, 図 3 には, カベクトルの差ノルムの相対誤差を示す. 相対誤差は, PME 法で計算される系の RMSF (root mean square force) を基準にして算出している. 結果から, 前述した α パラメータを小さくし, カットオフ半径 R_c を大きくすると精度がよいという傾向が再現されており, 今回行った計算のパラメータ範囲では, 特に, $\alpha = 0.1 \text{ \AA}^{-1}$, $R_c = 20.0 \text{ \AA}$ としたときに, 高い精度が得られた. このとき, ポテンシャルの相対誤差は, 約 0.05%, カベクトルの相対誤差は, 約 1.1% となっている. なお, MAO-A 系に関してはここには示していないが, 同程度の誤差に収まっている.

次に, 実際にタンパク質のダイナミクスに対してどの程度の影響があるか評価するために, RMSF (root mean square fluctuation) および, 動的相関マップ (以下, DCCM) を調べた. これらは, 残基ごとの平均構造からの揺らぎ, 残基間の揺らぎの相関を意味し, 以下の式で表される.

$$\text{RMSF}(i) = \sqrt{\langle (\mathbf{R}_i - \langle \mathbf{R}_i \rangle)^2 \rangle} \quad (2)$$

$$\text{DCCM}(i, j) = \frac{\langle (\mathbf{R}_i - \langle \mathbf{R}_i \rangle) \cdot (\mathbf{R}_j - \langle \mathbf{R}_j \rangle) \rangle}{\sqrt{\langle (\mathbf{R}_i - \langle \mathbf{R}_i \rangle)^2 \rangle \langle (\mathbf{R}_j - \langle \mathbf{R}_j \rangle)^2 \rangle}} \quad (3)$$

ここで、 \mathbf{R}_i は、残基の位置座標(ここでは、特に C_α 原子のみで評価した)を示しており、 $\langle \dots \rangle$ は、アンサンブル平均を示す。式(3)における DCCM は、-1 から 1 の値を持っており、1 で完全な正の相関、-1 で負の相関、0 で相関がないということを意味する。BPTI 系での RMSF と DCCM の結果を図 4,5 にそれぞれ示す。なお、この際修正 Wolf 法のパラメータは、 $\alpha = 0.1 \text{ \AA}^{-1}$ 、 $R_c = 20.0 \text{ \AA}$ および $R_c = 14.0 \text{ \AA}$ を用いた。結果から、RMSF, DCCM とともに、2 つの Coulomb 計算法による結果が非常によく一致しており、2 つの手法によるデータの相関係数は、それぞれ、0.961 と 0.977 となっている。以上のことから、修正 Wolf 法は、タンパク質の計算系においても非常に有効な方法であることがわかった。

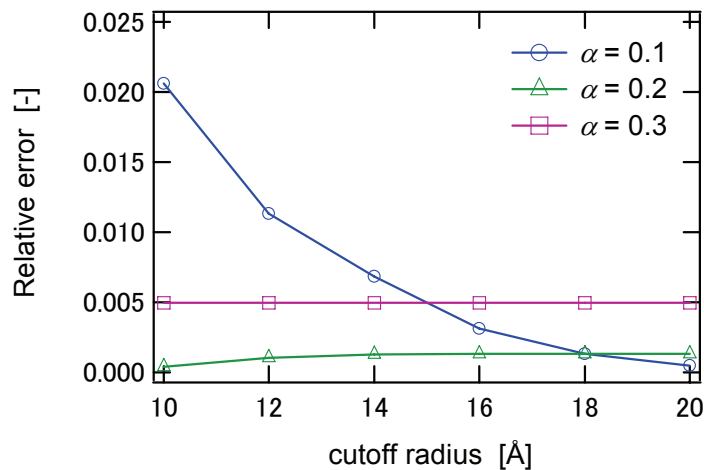


Fig.2 Relative error of potential value calculated by modified Wolf method and PME method.

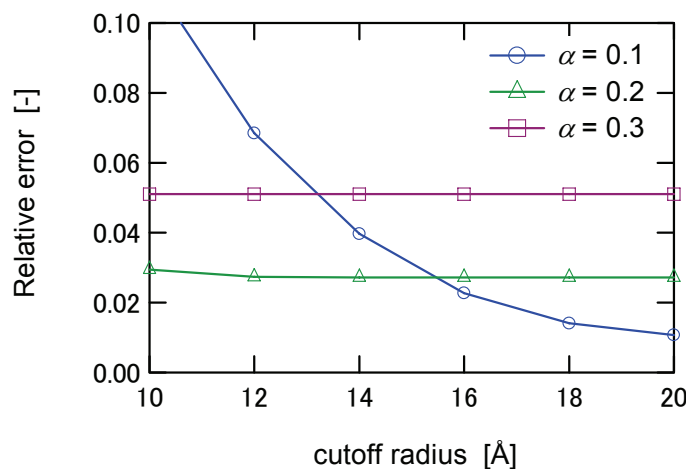


Fig.3 Relative error of the norm of force vector calculated by modified Wolf method and PME method.

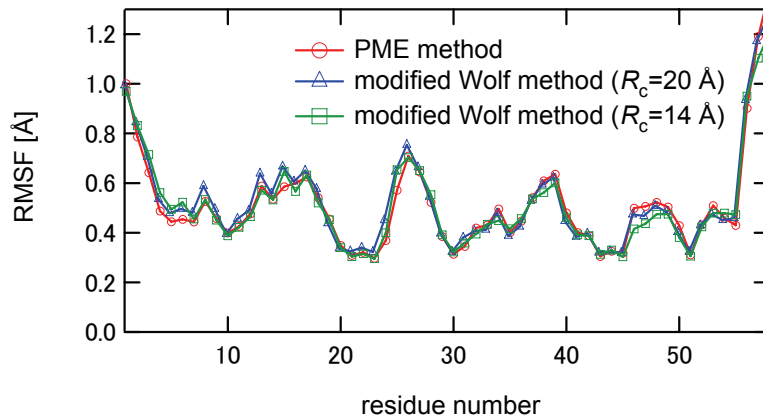


Fig.4 RMSF of the BPTI system by the modified Wolf method and PME method.

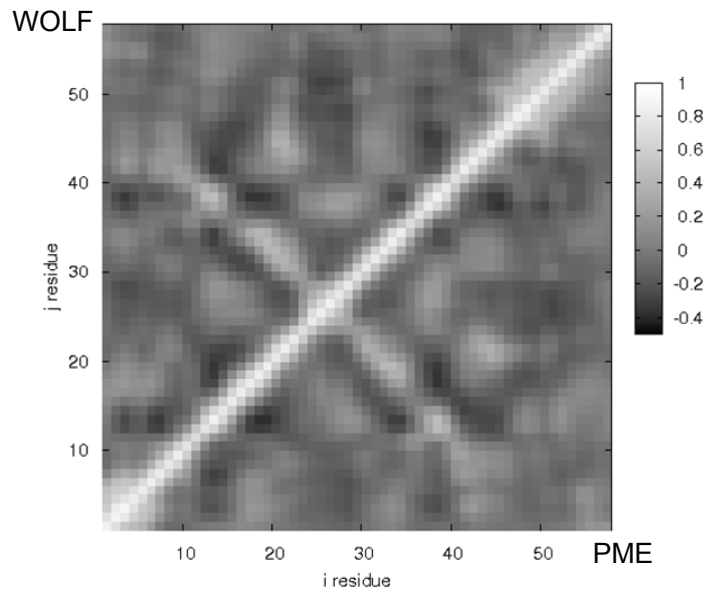


Fig.5 DCCM of the BPTI system by the modified Wolf method and PME method. Upper left triangular matrix corresponds to the result by the modified Wolf method, and lower right one the result by the PME method.

4.2 MDGRAPE-3 速度向上評価

次に MDGRAPE システム (本節では, 1 MDGRAPE ボード-1 ホストプロセスの結果) でのパフォーマンス評価の結果を示す. 表 1 には, BPTI 系での PME 法, 修正 Wolf 法の結果を, MDGRAPE 使用時とそうでない場合について示す. 修正 Wolf 法のパラメータは, $\alpha = 0.1 \text{ \AA}^{-1}$, $R_c = 20.0 \text{ \AA}$ を用い, PME 法は, 一般的に用いられている $\alpha = 0.4 \text{ \AA}^{-1}$, $R_c = 12.0 \text{ \AA}$ を用いて評価している. 結果から, PME 法では約 15 倍, 修正 Wolf 法では約 33 倍 (MDGRAPE なしの PME 法と比較して) の速度向上が得られた. PME 法で速度上昇率が大きくない原因は, 波数空間の演算をボードで計算することができず, ボトルネックとなっていることにある. これについては, 次節で詳細なデータを示す. さらに MAO-A 系の結果についても表 2 に示す. 結果から, BPTI 系と同程度の速度上昇率を得ており, やはり修正 Wolf 法のパフォーマンスがよいこ

Table 1 Performance of MDGRAPE calculations in the BPTI system by using the PME method and modified Wolf method. In the table, “Total” indicates the total time [sec] for 1 MD step, and “Ratio” the speed up ratio to the PME method without MDGRAPE board.

	PME		WOLF	
	w/o MDGRAPE	w/ MDGRAPE	w/o MDGRAPE	w/ MDGRAPE
Total	2.454	0.198	10.647	0.089
Ratio	1.0	14.91	0.277	33.17

Table 2 Performance of MDGRAPE calculations in the MAO-A system by using the PME method and modified Wolf method. In the table, “Total” indicates the total time [sec] for 1 MD step, and “Ratio” the speed up ratio to the PME method without MDGRAPE board.

	PME		WOLF	
	w/o MDGRAPE	w/ MDGRAPE	w/o MDGRAPE	w/ MDGRAPE
Total	14.115	0.969	57.136	0.404
Ratio	1.0	14.57	0.25	34.94

とがわかる。

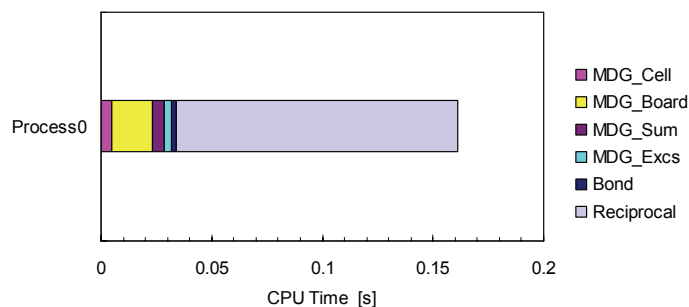
4.3 ホスト並列計算による MDGRAPE-3 性能評価

前節では、PME 法の波数空間が原因となり、MDGRAPE のパフォーマンスを活かせないということを述べた。そこで、この波数空間の演算部分やその他のホスト側での計算をホスト側で並列化することで、性能の向上を図った。すなわち 1 MDGRAPE ボード-N ホストプロセス並列での計算である。(説明の都合上、MDGRAPE に演算を渡すプロセスをプロセス 0 と呼ぶ) 並列化の実装は MPI を用いて行い、結果をわかりやすくするため、プロセス 0 では、MDGRAPE の演算のみを行い、プロセス 0 を除く全てのプロセスで、その他の演算の並列計算を行った。

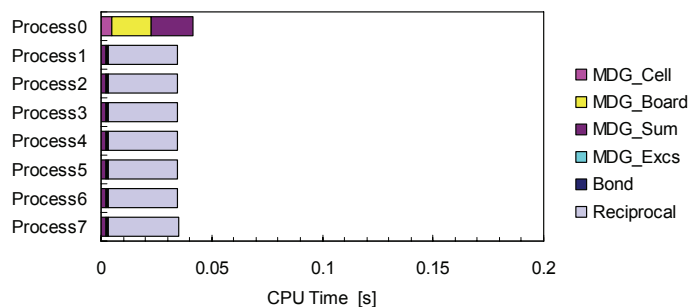
まず、BPTI 系の結果を表 3 に示す。結果は、最も性能が得られた並列数の結果のみ示す。PME 法では、8 プロセス並列計算のとき、修正 Wolf 法では、1 プロセス時に最も性能が得られたため、2 プロセス並列計算の結果を用いた。また、各プロセスの計算時間の内訳を PME 法と修正 Wolf 法について、それぞれ図 6,7 に示す。図中各項目については、MDG_Cell (MDGRAPE 計算用の cell-index 登録)、MDG_Board (ボード演算、厳密にはオーバーラップ計算^[2]の登録)、MDG_Sum (MDGRAPE 演算待ち時間、後処理)、MDG_Excs (ホスト側での分子内寄与の補正)、Bond (共有結合相互作用の計算)、Reciprocal (PME 法における波数空間の計算) である。このうち後 3 項目

Table 3 Performance of MDGRAPE calculations with host-based parallelized algorithm in the BPTI system by using the PME method and modified Wolf method. In the table, “Total” indicates the total time [sec] for 1 MD step, and “Ratio” the speed up ratio to the PME method without MDGRAPE board.

	PME			WOLF		
	w/o MDG	1 process w/ MDG	8 process w/ MDG	w/o MDG	1 process w/ MDG	2 process w/ MDG
Total	2.952	0.198	0.087	10.647	0.089	0.093
Ratio	1.0	14.57	33.93	0.25	33.17	31.74

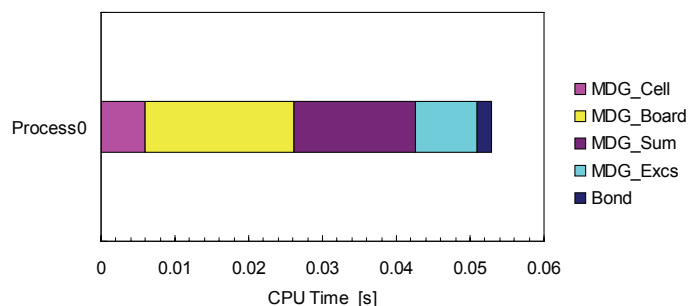


(a)

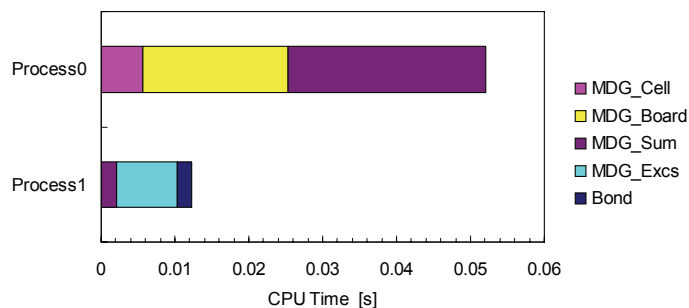


(b)

Fig.6 CPU time of the each process for 1 MD step with multi-process — 1-MDGRAPE board by the PME method in the BPTI system; (a) 1 process — 1 board, (b) 8 process — 1 board.



(a)



(b)

Fig.7 CPU time of the each process for 1 MD step with multi-process — 1-MDGRAPE board by the modified Wolf method in the BPTI system; (a) 1 process — 1 board, (b) 2 process — 1 board.

Table 4 Performance of MDGRAPE calculations with host-based parallelized algorithm in the MAO-A system by using the PME method and modified Wolf method. In the table, “Total” indicates the total time [sec] for 1 MD step, and “Ratio” the speed up ratio to the PME method without MDGRAPE board.

	PME			WOLF		
	w/o MDG	1 process w/ MDG	16 process w/ MDG	w/o MDG	1 process w/ MDG	2 process w/ MDG
Total	14.115	0.969	0.341	57.136	0.404	0.369
Ratio	1.0	14.57	41.39	0.25	34.94	38.25

が、ホスト側での並列化が可能なものであり、並列計算を行っている。また、図には数値積分や通信時間など、(現在のコードでは)並列不可能な寄与は示していない。結果から PME 法では、波数空間の演算時間が各ホストに分散し、期待通りパフォーマンスの向上が得られ、修正 Wolf 法と同等の性能を得るに至った。逆に修正 Wolf 法は、1 ホストで MDGRAPE の性能を使い切っており、ホスト側で並列化しても性能向上は得られないことがわかった。

同様に、MAO-A 系についての結果を表 4 に示す。MAO-A 系についても、並列化によって、PME 法の性能が大きく向上し、最終的に、MDGRAPE を使用しない場合の 40 倍強の性能が得られた。

5. 結論

MDGRAPE を用いて、タンパク質-溶媒系の MD シミュレーションを行い、MDGRAPE の性能評価を行った。特に、Coulomb 相互作用の評価に関して、PME 法における波数空間計算を行わない、修正 Wolf 法を導入し、その有効性(精度および計算速度)を示した。単純な 1 MDGRAPE ボード-1 ホストプロセス計算では、MDGRAPE を使用しない計算に対し、PME 法で約 15 倍、修正 Wolf 法では 30 倍強の性能を示した。さらに、ホスト側の演算を並列化することによって、PME 法でも 40 倍程度の性能を得ることに成功した。

現状では、MD コード側の制約で、複数の MDGRAPE ボードを並列利用できないが、今後、並列計算アルゴリズムを改良することでこれを可能にし、さらなる性能向上を目指している。

謝辞

MDGRAPE システムの利用やライブラリ利用法をはじめ、技術的な面で、理化学研究所ゲノム科学総合研究センター 成見哲氏にご助力いただいた。ここに記して謝意を表す。

参考文献

- [1] <http://mdgrape.gsc.riken.jp/>
- [2] Narumi T., Ohno Y., Okimoto N., Koishi T., Suenaga A., Futatsugi N., Yanai R., Himeno R., Fujikawa S., Ikei M. and Taiji M., *Proc. Supercomputing 2006*, in CD-ROM, (2006).
- [3] http://www.jbic.or.jp/activity/st_pr_pj/mypresto/index_mypr.html
- [4] Wolf, D., Keblinski, P., Phillpot, S.R. and Eggebrecht, J., *J. Chem. Phys.*, 110, pp.8254-8282, (1999).

- [5] Zahn, D., Schilling, B. and Kast, S.M., *J. Phys. Chem. B*, 106, pp. 10725-10732, (2002).
- [6] Allen, M.P. and Tildesley, D.J., *Computer Simulation of Liquids*, Oxford Univ. Press, (1987).
- [7] Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G., *J. Chem. Phys.*, 103, pp.8577-8593, (1995).
- [8] Demontis, P., Spanu, S., and Suffritti, G.B., *J. Chem. Phys.*, 114, pp.7980-7988, (2001).
- [9] Fennel, C.J., Gezelter, J.D., *J. Chem. Phys.*, 124, 234104, (2006).
- [10] Humphrey, W., Dalke, A. and Schulten, K., *J. Molec. Graphics*, 14, pp.33-38, (1996).