ハイパフォーマンス
コンピューティング 76－4
（1999．5．14）

分散並列処理による巨大分子系の電荷計算

三浦信明† 建部修見†† 北尾修†
長嶋雲兵† 関口智嗣††

光合成細菌などの巨大な分子を非経験的に取り扱うのは非常に困難であり，分子力学等の古典的な方法論によってその電荷を解析する必要がある．電荷平衡法は分子中の原子の部分電荷を簡便に計算できる方法である．その電荷の計算には原子数と同じ次元の連立一次方程式を解く必要がある．原子数が数万になる場合，1台のワークステーションでの計算は不可能となる．そこで，本稿では分散並列処理によって電荷の計算を効率良く行う事を試みた．

Calculation of Partial Charges of Huge Molecular Systems by Parallel Computing

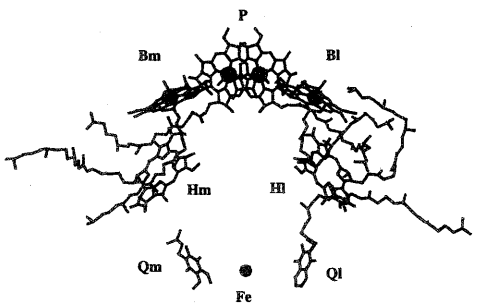
NOBUAKI MIURA,† OSAMU TATEBE,†† OSAMU KITAO,†
UMPEI NAGASHIMA† and SATOSHI SEKIGUCHI††

It is almost impossible to treat the huge molecules as photosysthetic bacteria by *ab initio* molecular orbital theory, then we need to treat the molecules by the molecular mechanic method, which is a classical methodology. The charge equilibration method is convenient method to calculate the partial charges in the molecule. In order to calculate the charges, we need to solve the linear equation whose order is the same as the number of atoms in the molecule. The case which the number of atoms becomes large (~20 000) it is difficult to solve the linear equation on the computer with single CPU. In this report, we try to efficiently calculate the partial charges of the atoms in *Rhodpseudmonas viridis* by parallel computing.

1. はじめに

光合成の初期過程は反応中心に含まれる補因子バクテリオクロロフィル2量体 (P) の光励起によって誘起される電荷分離とその後の一連の電子移動反応である．その機構には現在でも未解決な問題がある．**図1**にX線スペクトルで測定された反応中心の補因子を示す[1]．Pが光によって励起されP*となった後，電子はバクテリオフェオフィチン (H) に移動する．この際にアクセサリバクテリオクロロフィル (B) を介した2段階の反応なのか，直接電子が移動する1段階の反応なのかは解っていない．反応中心はほぼC_2対称であるにもかかわらず電子はもっぱらL側を通る．L側とM側を通る比率はおよそ100:1である[2]．我々は，光合成の初期過程でおこる光合成細菌内の電子移動反応の機構を解明するための方法論の構築を行っている．電子伝達系に対する周辺蛋白質が作る電場の影響を調べるため，電荷平衡法を用いて紅色光合成細菌の全原子の部分電荷の計算を行う事は電子移動反応の初期過程の知見を得る上で重要である．

図1 光合成反応中心の補因子



これらの反応の特徴は反応中心を取り囲む蛋白質が作る構造的な要因と電場環境に由来すると考えられるが，1984年にX線解析によってDeisenhoferらによって構造が調べられた紅色光合成細菌は約20 000原子からなり，その反応中心の電場を解析するのは非常に困難である．現在の計算資源では，このような巨大な

† 物質工学工業技術研究所
National Institute of Materials and Chemical Research
†† 電子総合技術研究所
Electrotechincal Laboratory

分子を非経験的計算で取り扱うのは不可能であり，分子力学等の古典的な方法論によらざるをえない．

通常の分子力学法では，各原子を原子種によってある決まった電荷を持った点電荷として表現して静電力を構成する．しかし，原子の部分電荷はたとえ同じ原子種であっても，その置かれる状況が異なれば異なる電荷を持つと考えるのが自然である．

Rappe と Goddard III によって提唱された電荷平衡法[3]は，原子のエネルギーを電荷の汎関数として表現し，イオン化ポテンシャルや電子親和力などを用いて簡便に原子の部分電荷を計算する方法である．電荷の計算に必要な計算は，実非対称密行列を係数行列とする連立一次方程式を解くことである．紅色光合成細菌の原子上の電荷を計算するには 20 000 次元の連立一次方程式を解くことが要求されるが，LU 分解法では次元数の 3 乗に比例する時間と，2 乗に比例する記憶領域が必要である．20 000 次元の場合行列要素は約 3.5GB になる．1 台のワークステーションで計算を行う場合には，行列要素をディスクに置くなどの工夫が必要である．しかしディスクはメモリよりアクセス速度が非常に遅いため，処理速度の点で不利である．分散並列処理により 1 台あたりが保持するデータ量を減らす事で，現実的な計算環境で計算が行えると期待できる．また，係数行列は要素間に相関する部分がないためその生成処理の並列化は容易であり，処理速度の観点からも効果が期待できる．ここでは，分散並列処理を行なうことで電荷平衡法の連立一次方程式の求解を効率よく行なうことを考えた．

2. 電荷平衡法

電荷平衡法については Rappe と Goddad III の文献などに詳しいのでここではその概略を述べるにとどめる．1 原子の静電エネルギーは

$$E_i(q_i) = E_0 + \chi_i^0 q_i + \frac{1}{2} J_{ii} q_i^2$$

と表され，分子については

$$E(q_1, q_2, ..., q_N; R_1, R_2, ...,) = \sum_i^N E_i(q_i) + \sum_{i<j}^N J_{ij}(r_{ij}) q_i q_j \qquad (1)$$

ここで，q_i は原子上の部分電荷，χ_i^0 は原子の電気陰性度，J_{ii} はハードネスである．R_i は原子の核座標で r_{ij} は原子 i と原子 j の核間距離である．$J_{ij}(r_{ij})$ は，

$$J_{ij}(r_{ij}) = \int\int |\phi_{n_i\zeta_i}(r_1 - R_i)|^2 \frac{1}{r_{ij}} \times |\phi_{n_j\zeta_j}(r_2 - R_j)|^2 dr_1 dr_2, \qquad (2)$$

のようにクーロン積分の形をしており，ϕ はスレーター型関数

$$\phi_{n_i\zeta_i}(r_1 - R_i) = N_{n_i\zeta_i}(r_1 - R_i)^{n_i - 1} e^{-\zeta_i(r_1 - R_i)}$$

である．本計算では $\phi_{n_i\zeta_i}$ を Stewart の展開式を用いて 3 個の s 型ガウス関数で展開している (STO-3G)[4]．

$$\phi_{n_i\zeta_i}(r_1 - R_i) = N_{n_i\zeta_i} \sum_{k=1}^{3} c_k M_k e^{-\alpha_k(r_1 - R_i)^2}$$

分子の静電エネルギーに全体の電荷が定数であるという拘束条件をつけて停留置を求める条件から電荷を決定する方程式が導かれ，

$$\begin{pmatrix} 1 & 1 & \cdots & 1 \\ J_{21} - J_{11} & J_{22} - J_{12} & \cdots & J_{2N} - J_{1N} \\ J_{31} - J_{11} & J_{32} - J_{12} & \cdots & J_{3N} - J_{1N} \\ \cdots & & & \\ J_{N1} - J_{11} & J_{N2} - J_{12} & \cdots & J_{NN} - J_{1N} \end{pmatrix} \times \begin{pmatrix} q_1 \\ q_2 \\ q_3 \\ \\ q_N \end{pmatrix} = \begin{pmatrix} 0 \\ \chi_2^0 - \chi_1^0 \\ \chi_3^0 - \chi_1^0 \\ \\ \chi_N^0 - \chi_1^0 \end{pmatrix}. \qquad (3)$$

水素原子を含む場合にはスレーター型関数の軌道指数 ζ_i

$$\zeta_i = \zeta_i + q_i$$

と表現し，電荷の値が収束するまで繰り返し計算を行う必要がある．

3. 計算の詳細

本計算では電荷平衡法のプログラムを DEC Alpha Station のクラスタ etlwiz を用いた．連立一次方程式の求解には建部のガウスの消去法のライブラリ[5]を用いた．**図 2** に etlwiz の構成を示す．etlwiz は 32 台構成で，それぞれのプロセッサにおけるメモリは 512MB×8 台，256MB×8 台，128MB×16 台のような 3 段構成である．これは並列処理を実行する上で並列度に依存しない一定のメモリ量を確保するためである．ネットワークは 100Base/TX を ether switch で接続した構造になっており，ether switch のバックプレーンは 1.2Gbps である．

作成したプログラムコードを用いて，問題の大きさを 1000~10000 の範囲で変えながら処理時間を測定し処理時間を問題サイズの関数で表せるか検討した．約 20 000 次元の問題を分散並列処理で解こうとした際に各ノードに均等にデータを配分すると 1 台あたりに約 95.3MB のデータを割り当てることになる．etlwiz ではメモリが 128MB のプロセッサがあるためページングが起こり効率が著しく落ちる可能性がある．まず 128MB のメモリを実装 8 台のプロセッサ，および 512MB のメモリを実装する 8 台のプロセッサを用いて変えながら処理時間を計測し，ページングについて検討した．この場合 10 000 次元が上記の 1 台あたり 95.3MB のデータ配分に当たる．

図 3 128MB, 512MB のメモリを実装するプロセッサを用いた場合の処理時間の比較. a) 係数行列の生成. b) 連立一次方程式の求解.

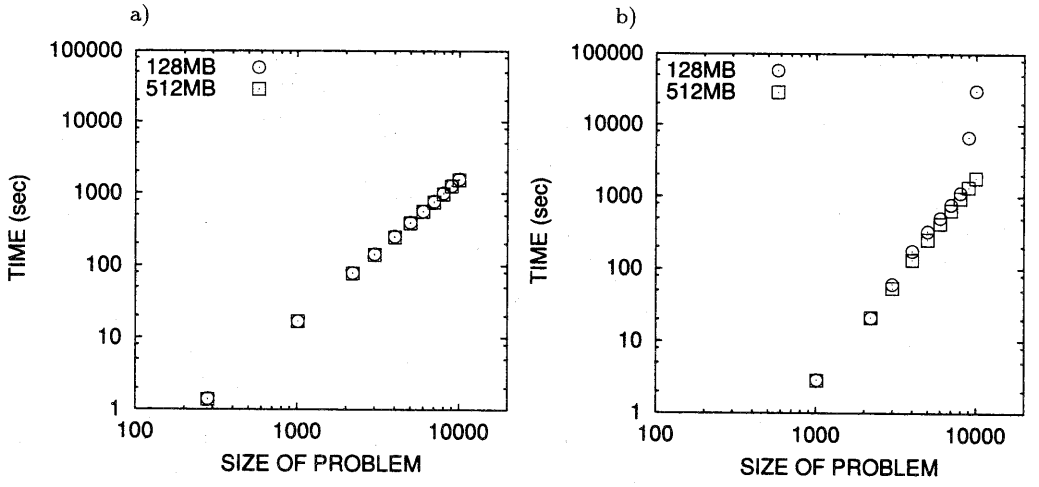


図 2 etlwiz の構成図

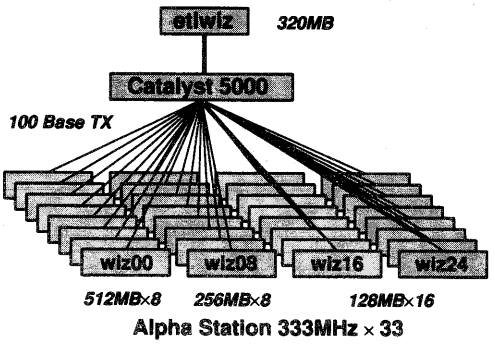
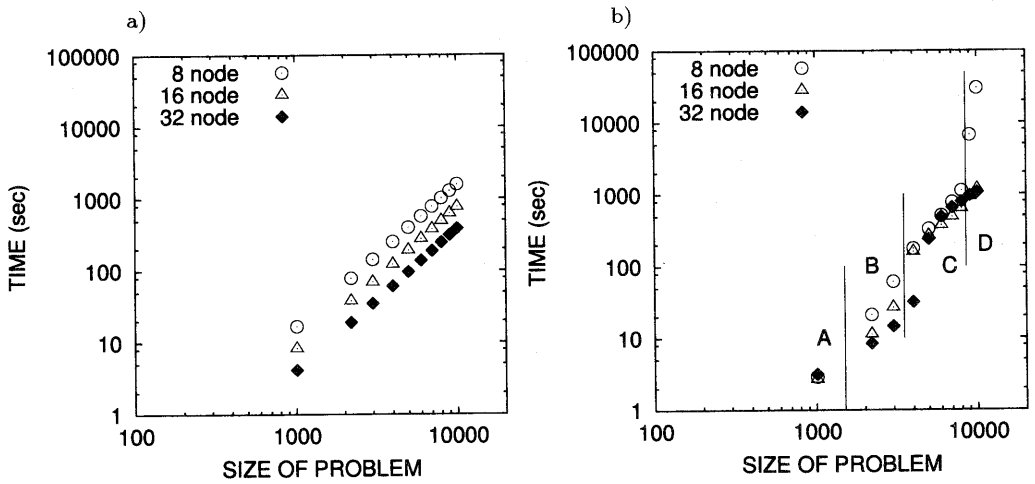


図 3 に 128MB のメモリを実装した 8 台のプロセッサ (構成 1) 及び 512MB のメモリを実装したプロセッサ 8 台 (構成 2) を用いて処理時間を計測した結果を示す. 計算は, 行列の生成と一次方程式の求解の 2 つの部分からなっている. それぞれ問題の大きさの N に対して N^2, N^3 に比例する. 構成 1 に対して, 最小自乗法によって求めた比例係数はそれぞれ 1.54×10^{-5}, 1.79×10^{-9} である. 係数行列の生成の処理時間は実装するメモリに依存しない結果となった. これは、行列要素間に依存する関係がないために生成された要素が再び参照されることがないためページングの影響を受けていないと考えられる. 一方, 一次方程式の求解では 9000 次元で構成 1 では著しい速度低下が見られ, ページングが起こっていると考えられる. 上記の比例係数はこれら 2 点を除いてフィッティングを行ったものである. 多くのメモリを実装するプロセッサを用いて計算を行えば, ページングを回避する事が可能であることがわかる.

図 4 は 8 台及び 16 台, 32 台のプロセッサを用いて処理した際の問題サイズに対する処理時間を表している. 8 台のプロセッサを用いた処理は構成 1 の結果を示してある. 係数行列の生成は, 処理に用いるプロセッサの台数を増やすとそれに連れて処理時間が短くなっており, 効率良く並列化がなされている. 連立一次方程式の求解について検討する. 前述の通り 8 台のプロセッサ (構成 1) を用いた計算では, 9 000 次元になったときにページングを起こして処理時間が 1 桁長くなっていることがわかる. 並列処理の効率を考える場合, 速度の点ではデータは均一に分散させたほうが有利だが, 計算機の仕様によってデータの分散を均一にすることが著しい効率の低下を招く事があるといえる. また, 処理時間の N 依存性は 4 つの傾向を示している. 8 台のプロセッサを用いた計算では, ページングによる効率の低下 (領域 D) を除けば, ほぼ完全に N^3 に比例している. これに対して 16 ノードと 32 ノードの場合は 2000~3000 次元 (領域 B) では台数に比例した効果が得られているが, 2000 次元未満の領域 (領域 A) と 4000 次元以降 (領域 C) では台数増加による効果が得られていない. それぞれの領域における処理の台数効果を表したグラフを**図 5** に示す. 領域 A では処理する問題が小さすぎるために処理時間が通信に依存していると考えられる. 注目する点は領域 B でスーパーリニアスピードアップが見られることである. 問題サイズが 2000 次元から 3000 次元であるこの領域が今回用いた etlwiz では最適な問題サイズと考えられる. 領域 C ではプロセッサ数を増やすことによる効率の向上は見込めない事が分かる. 恐らくこの領域はプロ

図 4　8 台, 16 台, 32 台のプロセッサを用いた際の処理時間の比較. a) 係数行列の生成. b) 連立一次方程式の求解.



セッサの台数を増やす事によって通信量が増えるために台数分の効率が引き出せないのではないかと考えられるが, 現時点では理由ははっきりとは解らない. また, 領域 D では見かけ上スーパーリニアスピードアップが見られる. これは, 8 台のプロセッサを用いた計算の処理時間が非常に長いために見かけ上見られるものであり, このような相対的な評価のみでは効率を議論出来ない事を示唆している.

ページングを回避する方法してはデータを不均一に分散し, メモリを多く実装しているプロセッサに対して多くのデータを割り当てることが考えられる. 具体的には 512MB のメモリを実装する 8 台と 256MB のメモリを実装する 8 台に対して 128MB のメモリを実装する 16 のプロセッサの倍の量のデータを分配した.

4. 結　　果

反応中心にあるバクテリオクロロフィル 2 量体 (スペシャルペア) の L 側と M 側それぞれのユニットの中心の Mg と窒素原子, いくつかの炭素原子のの電荷を**表 1** に示す. 各原子の配置は**図 6** の通りである. わずかな違いではあるが, L 側のユニットが M 側に比べて負電荷を帯びているように見て取れる. ユニットの電荷の和についても M 側が 0.226 であるのに対して L 側は 0.254 であり, 電荷分布はわずかに L 側に偏っていることがわかった.

表 2 は, 紅色光合成細菌の原子上の部分電荷を求める際に要した時間を示す. 予測時間は問題サイズを変えながら処理時間を計測した際に行列生成と一次方程式の求解にかかる時間を問題サイズの 2 次関数, 3 次関数へのフィッティングによって求めた. 32 台のプロ

図 6　バクテリオクロロフィル二量体のポルフィリン骨格と原子の番号付け. この図には L 側のユニットについて示した. M 側はこの図と左右反転になる.

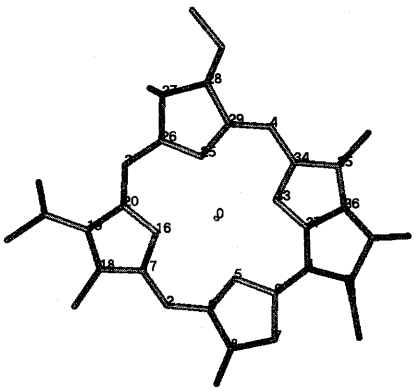


表 1　いくつかの原子上の電荷

原子	番号	電荷	
		M	L
Mg	0	0.559	0.551
C	1	−0.075	−0.074
C	2	−0.110	−0.129
C	3	−0.124	−0.133
C	4	−0.099	−0.090
N	5	−0.307	−0.303
N	16	−0.301	−0.305
N	25	−0.253	−0.299
N	33	−0.306	−0.319

図 5　A～D の領域における処理の台数効果．点線は理想値である．

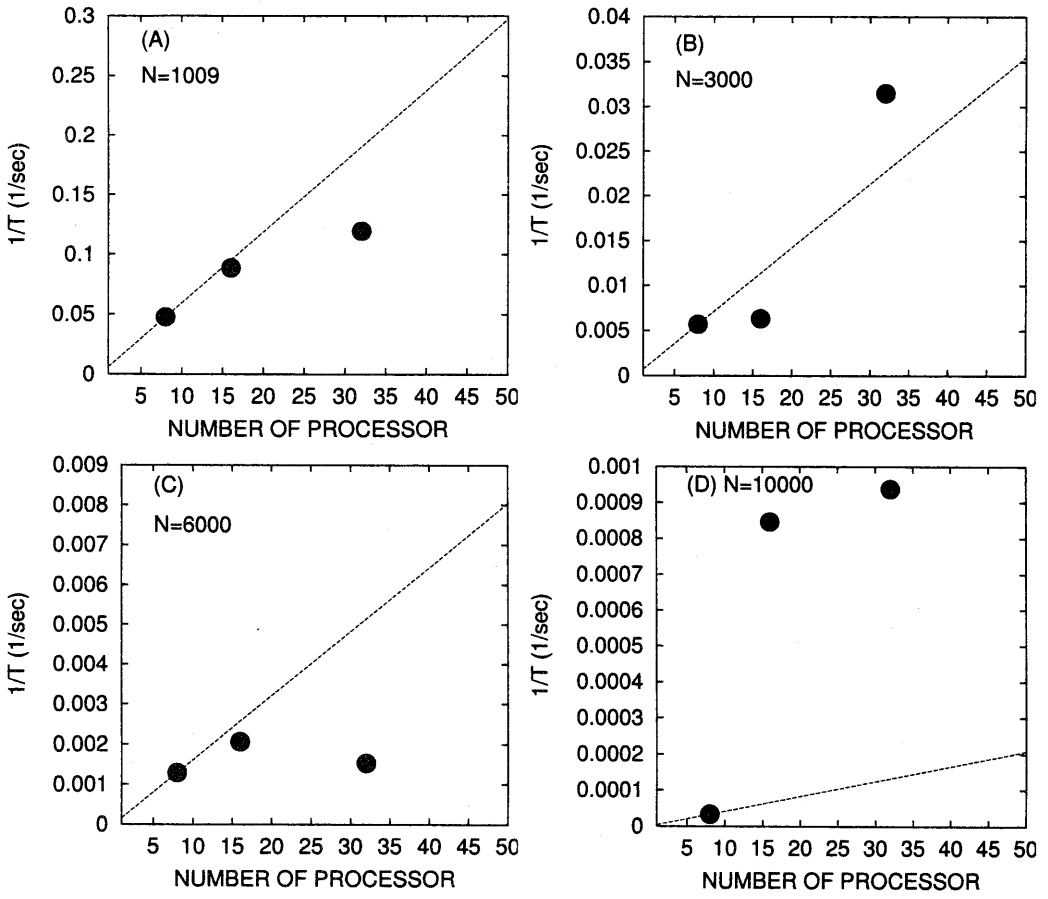


表 2　紅色光合成細菌の電荷計算に要した時間
均一分散は 32 台のプロセッサを用いた場合と 512MB のメモリを実装した 8 台のプロセッサを用いて行った結果を示す．不均一分散については 32 台のプロセッサを用いた場合の結果を示す．

	#of node	予測値		実測値	
		matrix	solv.	matrix	solv.
均一分散 (Wiz)	32	1665.74	10949.55	1767.54	70697.96
均一分散 (Wiz)	8	6612.96	15861.87	6558.53	15367.62
不均一分散 (Wiz)	32	2346.43	13666.81	2309.85	14580.44

セッサを用いた計算では，係数行列の生成については全データ，連立一次方程式の求解については領域 C と領域 D のデータをフィッティングに用いた．

均一分散の場合は予測時間 10950 秒に対してページングの影響によって実際には 70698 秒と予測の 7 倍の時間がかかっている．これに対して不均一にデータを分散させた場合 (不均一分散) は行列の生成・求解ともに予測通りの時間で処理が行われている．512MB のメモリを実装する 8 台のプロセッサを用いて電荷の計算を行った際の処理時間も tabreftab:time に載せてある．不均一分散の場合と比べると係数行列の生成で約 3 倍の処理時間であるが，連立一次方程式の求解については 5%処理時間が長いだけである．今回のシステムでは 512MB のメモリを実装した 8 台のプロセッサを用いて行った計算がもっとも効率良く処理できていると考えられる．言い換えれば，etlwiz で約 20 000 次

元の問題を解くに当たってはプロセッサの数を増やしても処理の効率はあがらないことを示唆している.

5. ま と め

分散並列処理を行うことで紅色光合成細菌の原子上の部分電荷の計算が効率よく行えた. 反応中心のバクテリオクロロフィル2量体のL側, M側のユニットの電荷を比較したが, わずかにL側に電荷が偏っていることが解った. 結果は電子移動反応経路がL側に著しく偏る原因が電場環境にあるとはっきりいえるほどには電荷分布は偏ってはいなかった. 電子移動後の系に対する電荷の計算との比較が必要である.

解く問題のサイズが大きくなると分散並列処理を行ったとしても1台あたりのノードが保持するデータ量は大きくなる. 不均一にメモリを実装したクラスタで計算を行う場合, その分散のさせ方如何ではページングなどにより計算効率が著しく低下する. このような場合はデータを不均一に分散させることが有効である事が解った. しかしながら, 今回の問題ではプロセッサの数を増やすよりも大容量メモリを実装したクラスタで計算を行うことが効率の面で最適であることが解った.

数値計算的な視点からみると, 問題毎に有効な並列化の環境があり, 闇雲にプロセッサの数を増やす事が必ずしも高速化にはつながらないことが解った. これは, 各々の問題に対して最適な並列環境を構築する必要性を示唆しており, 安価に構築できるPCクラスタは構成の自由度が大きいので大規模科学技術計算に有効であると考えられる.

最後に今後の課題を挙げる. 電子移動後の系の電荷計算を行う事は反応機構を議論する上で必要である. 数値計算的な観点では, 本計算によって示唆された問題ごとにあるであろう最適な並列環境を決定してる因子を調べる必要があると考えられる. 今回示した(C)の領域の支配的要因を調べる事は必要で, プロセッサ数を増やすことによるメリットとデメリットを定量的に見積もる必要がある.

謝辞 この研究はCOEプロジェクト「光反応制御・光機能材料」, ならびに工業技術院国際特定共同研究「ハイパフォーマンスコンピューティングシステム性能評価技術の研究」に基づいて行われた. また、計算機環境に関して御協力頂いた東北大学超臨海溶媒工学センター猪俣宏教授に感謝する.

参 考 文 献

1) J Deisenhofer, O Epp, K Miki, R Huber and H Michel, *J. Mol. Biol.*, **184**, 385 (1984).
2) J Berton, J L Martin, A Migus, A Antonetti and A Orszag, *Proc. Natl. Acdd. Sci. USA*, **83**, 5121 (1986)
3) A K Rappe and W A Goddard III, *J. Phys. Chem.*,**93**, 7320 (1991)
4) Stewart, *J. Chem. Phys.*, **52**, 431 (1970)
5) 建部 修見, 「分散メモリ型並列計算機によるLU分解」, 情報処理学会研究報告, 95-HPC-57, SWoPP'95 別府, pp.55-60, 1995年8月