

大規模分子の分子軌道計算 – Lysozyme とモデル DNA 分子の分子軌道 –

渡邊 寿雄^{a,b}, 稲富 雄一^c, 石元 孝佳^{a,b}, 梅田 宏明^{a,b}, 櫻井 鉄也^{a,b,d}, 長嶋 雲兵^{a,b*}

^a 科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST),

〒 332-0012 埼玉県川口市本町 4-1-8 川口センタービル

^b 産業技術総合研究所 計算科学研究部門, 〒 305-8568 つくば市梅園 1-1-1 中央第二

^c 九州大学情報基盤研究開発センター, 〒 814-0001 福岡市早良区百道浜 3-8-33-710

^d 筑波大学大学院 システム情報工学研究科, 〒 305-8571 つくば市天王台 1-1-1

*e-mail: u.nagashima@aist.go.jp

(Received: June 7, 2007; Accepted for publication: July 20, 2007; Published on Web: August 30, 2007)

フラグメント分子軌道 (FMO) 法に基づいた大規模系の分子軌道 (FMO-MO) 計算についての概要と、Lysozyme 分子およびモデル DNA 分子への適用結果を示す。FMO-MO 法は、FMO 計算によって得られた密度行列を基に Fock 行列を生成し、それを一度だけ対角化することによって近似的な分子軌道を得る方法である。本方法は巨大な分子全体に対する反復解法を行わずにすむため、従来の分子軌道法では非常に困難であった 2000 原子程度の大規模分子の分子軌道計算を、比較的短時間で行うことができる。また新たに開発した対角化法を用いると化学反応に重要な HOMO-LUMO 付近のみの軌道エネルギーと軌道を得ることが可能であり、従来法に比較して大幅な時間短縮が可能である。

Dual Opteron 246, 2GHz × 128 台を用いた並列計算では、Lysozyme 分子 (129 アミノ酸残基, 1961 原子) と溶媒 (8258 原子、STO-3G 20758 基底関数) に対する FMO-MO 計算を、5 時間程度で行うことができた。このとき、水分子の配置の仕方で HOMO-LUMO の位置が大きく変化することが確認された。また同じシステムを用いると DNA のモデル (40 核酸対、2636 原子、STO-3G 10108 基底関数) を 1 時間以内で計算可能であった。この場合、HOMO は DNA の中央に大きな振幅を持ち、LUMO は DNA の末端近くに大きな振幅を持っていた。

キーワード: Fragment Molecular Orbital method, Large scale molecule, Parallel processing, Lysozyme, Model DNA

1 はじめに

新規物質設計・製造の高度化・低コスト化技法の開発や新薬の安価・高速・安全な開発技法の開発のニーズの高まりと、無駄な廃棄物を生む化合物の合成や動物実験などを行わないなど、環境負荷の少ない開発技法への強い要請を背景に、材料シミュレーションへの期待が高まっている。

材料シミュレーションの技法として、主に孤立系の電子状態の計算に用いられる、量子力学に基づいた理

論計算の一種である分子軌道 (Molecular Orbital, MO) 法 [1] がある。MO 法では分子中の電子分布を表す多くの MO が得られる。そのうちフロンティア軌道 [2] と呼ばれる 2 つの MO と、その軌道エネルギーの解析が、さまざまな化学反応のメカニズムを理論的に解明する際に重要となるというフロンティア電子理論で 1981 年にノーベル賞が福井謙一に与えられた。近年、DNA の 1 次構造や、多くのタンパク質の立体構造が明らかになってきたが、タンパク質などの巨大生体分

子の反応機構の解明は、その分子の大きさゆえに、反応する部位を特定することすらままならないことが多い。したがって、DNA やタンパク質などの巨大分子のフロンティア軌道を求めることは、反応部位、および反応機構の理論的な解明に非常に重要であり、創薬などの分野で特に有用である。

従来の MO 法による MO 計算では、全エネルギーや電子密度分布とともに、MO やその軌道エネルギーも同時に得られる。しかし、計算量が分子サイズの 4 乗に比例し、かつ行列の対角化など、並列化による高速計算が行いにくい部分が存在することなどから、1000 原子を超えるような大規模分子に対して従来の MO 法を適用すると、少し大きめの PC クラスタなどを用いた場合でも、10 日～数ヶ月といった長い計算時間が必要になる。いろいろな有機化合物から有効な薬品になる物質の候補を絞るスクリーニングなどに MO 法を利用する場合には、この長い計算時間が大きな障害となる。

北浦によって提案されたフラグメント分子軌道 (FMO) 法 [3-5] は、従来の MO 法の計算量を大幅に削減するための計算手法の一つとして、最近注目されている。FMO 法では、大規模 MO 計算をいくつかの小規模 MO 計算に分割して行い、その結果を用いて巨大分子全体のエネルギーや電子密度分布を求める。特徴的なことはその基本的な式からも判るように、タンパク質内部のフラグメント間の様々な相互作用を直接計算できることである。タンパク質のフラグメント間相互作用は、薬剤設計には欠かすことのできない情報であり、従来は研究者の知識と経験によって見積もられていたが FMO 法を用いると自然な結果としてフラグメント間相互作用が得られる。

FMO 法は並列化が非常に容易で高効率期待できるため、すでに複数のグループが並列 FMO 計算プログラムの開発・性能評価を行っており、非常に高い並列性能が出ることが示されている [6]。FMO 法による最大規模の計算は、今のところ 2005 年に池上らが行った光合成反応中心タンパク質のもので、原子数 20581、電子数 77754、基底数 164442 (6-31G*) といった計算規模である [7]。彼らは高度に並列化されたプログラムを実装し Dual Opteron (2 GHz) 300 ノード (600CPU) を用いて約 90 時間で計算を実行している。論文に書かれることはないが、この計算の入力データの作成には約 3ヶ月程度の時間を要したと聞いている。すでに計算時間より信頼の置ける入力データの作成時間の方に時間がかかるという状況が目前に迫っている。

FMO 法では、全エネルギーと電荷密度を得ることができ、そのために分子全体に広がるフロンティア軌道を求めることはできない。しかし、我々は FMO 法で得られた電子密度行列を使って全系の Fock 行列を再構築し、その対角化を行う (FMO-MO 法) ことで、フロンティア軌道をはじめとする MO や軌道エネルギーが精度よく求められることを確かめた [8]。Fock 行列の対角化も従来の Householder-Bisection 法や Householder-QR 法を用いて全固有値・固有ベクトルの計算を行うのではなく、櫻井と杉浦によって新たに開発された対角化法 [9] を用いると並列化が非常に容易であるため、計算量は多いものの、並列計算機を用いた高速化が図りやすい。我々の開発した FMO-MO 法は大規模 MO 計算が可能な FMO 法を拡張した計算法であり、FMO 法では得られなかった大規模分子の MO を求めることを可能にする。これは、従来法ではどうも解けなかった大規模系の分子軌道計算が可能となったことに他ならない。

我々は、科学技術振興機構 (JST) 戦略的創造研究推進事業 (CREST) の研究領域「シミュレーション技術の革新と実用化基盤の構築」の研究プロジェクト「グリッド技術を用いた大規模分子シミュレーションプログラムの開発」において、グリッド技術を利用したフラグメント分子軌道法に基づく分子軌道計算プログラムの開発と大規模 MO 計算の実現を目指している。特に FMO-MO 計算に適した一般化固有値解法として櫻井-杉浦法 [9] を利用することにより、グリッド環境下でも効果的な並列計算が可能となる。

本稿では紙面の都合上、FMO-MO 法を用いた Lysozyme 分子およびアデニンとチミンの対によるモデル DNA 分子の MO 計算に話を限ることとする。

2 計算方法

2.1 FMO 法

北浦によって開発された FMO 法 [3-5] は、分子を 20 ~ 30 個程度の原子で構成される小さなフラグメントに分割して、このフラグメントとフラグメントペアに対して、従来 MO 法を適用する。タンパク質においてはペプチド結合を切断することでフラグメンテーションを行うことが経験的に行われている。その結果得られる、フラグメント、ならびにフラグメントペアの密度行列 (それぞれ、 $\{\mathbf{D}_I\}$, $\{\mathbf{D}_{IJ}\}$ ($1 \leq I, J \leq N_f$)). N_f はフラグメント数) を用いて、分子全体の密度行

列を計算する。

$$\mathbf{D}^{\text{FMO}} = \sum_{I>J}^{N_f} \mathbf{D}_{IJ} - (N_f - 2) \sum_I^{N_f} \mathbf{D}_I$$

また、各フラグメント（フラグメントペア）に対する Fock 行列中の核-1 電子ハミルトニアン行列には、次に示すとおり、従来 MO 法で用いた行列に若干の補正が含まれる。

$$\begin{aligned} (\mathbf{H}_X^{\text{CORE}})_{ij} &= (\mathbf{H}_{X,0}^{\text{CORE}})_{ij} + (\mathbf{V}_X)_{ij} + (\mathbf{B}_X)_{ij} \\ (\mathbf{V}_X)_{ij} &= \sum_{K \neq X}^{N_f} \left\{ \sum_{A \in K} \int d\mathbf{r} \varphi_i(\mathbf{r}) \frac{-Z_A}{|\mathbf{r} - \mathbf{R}_A|} \varphi_j(\mathbf{r}) \right. \\ &\quad \left. + \sum_{k,l \in K} (\mathbf{D}_K)_{kl} \langle ij|kl \rangle \right\} \end{aligned}$$

ここで、添字の X はフラグメント、またはフラグメントペア（それぞれ I, J ）を表している。和の中に含まれる添え字 K, A および $\{k, l\}$ は、それぞれ（条件を満たす）フラグメント、原子、および基底関数の番号を表す。さらに、式中の Z_A, \mathbf{R}_A は、それぞれ原子 A の原子核電荷、核座標を示す。この補正された核-1 電子ハミルトニアン行列の右辺第一項 ($\mathbf{H}_{X,0}^{\text{CORE}}$) は、従来 MO 法とまったく同じものであり、右辺第三項 (\mathbf{B}_X) は射影演算子と呼ばれ、フラグメントに分割する際に、その切断部分を適切に処理するために必要な要素である。残りの右辺第二項 (\mathbf{V}_X) は周辺のフラグメントに属する原子核や電子から受ける静電相互作用の効果を表す部分（フラグメント間相互作用項）で、「周辺のフラグメントからの影響は、静電相互作用のみを考慮する」という FMO 法の方針を示している。この周辺のフラグメントからの静電相互作用項を表す式のうち、前半は原子核からの核-電子引力項であり、原子の座標が変わらない限り変化しない量であるが、後半の周辺フラグメントの電子との 2 電子間静電反発項は、周辺フラグメントの電子密度に依存する。

FMO 計算を実際に行う場合には、すべてのフラグメントの電子密度行列 ($\{\mathbf{D}_I\}$) の初期値を与え、その電子密度を基にフラグメント間相互作用項 ($\{\mathbf{V}_I\}$) を計算したのち、フラグメントに対する MO 計算を行う。そうすると、各フラグメントの密度行列が MO 法の反復計算の間に変化するため、フラグメントの電子密度に依存するフラグメント間相互作用項も変化することになる。したがって、フラグメントの MO 計算は、フラグメントの電子密度分布（あるいはエネルギー）が収束するまで繰り返し行い、FMO 法におけるフラグ

メントの電子密度分布を求める。こうして得られた密度分布を、Self-Consistent Charge (SCC) と呼ぶ。フラグメントペアに対しては、SCC を用いてフラグメント間相互作用項の計算を行い、1 度だけ MO 計算を行う。

各フラグメントやフラグメントペアに対する MO 計算は独立に行うことができるため、FMO 法では並列処理による高速化を効率よく行うことができる [5]。また分子全体を一度に扱うことがないため、計算量はフラグメント数を m とすると $O(m(N/m)^4) \doteq O(N^2)$ 程度になり、HF 法の $O(N^4)$ に比べて大幅に減少する。この結果、FMO 法では大規模分子のエネルギー、ならびに電子密度分布を、HF 法に比べて非常に短時間に、しかも精度よく求めることが可能である [3, 13]。その一方、FMO 法では各フラグメント、フラグメントペアの軌道エネルギー（固有値）や軌道（固有ベクトル）の計算は行うが、分子全体に広がる MO を求めることはない。このため、タンパク質などの大規模分子のフロンティア軌道を求めるためには FMO 法によって得られる密度行列を用いて、全系のフォック行列を再構築して、それを対角化すれば、近似的な MO を求めることができる。

2.2 FMO-MO 法

FMO 法では、分子全体の電子密度行列を得ることができるが、MO を求めることができなかった。FMO 法で計算できる分子のエネルギーや電荷分布なども反応メカニズムを理論的に解明するための有効な情報ではあるが、活性部位などの議論のためには、やはり MO を求める必要がある。フラグメントの軌道で分子全体の軌道の代用することも提案されているが、比較的小さな分子でもフラグメントの軌道エネルギーの準位が HF 法の結果と違うことがある [10]。そこで、われわれは、FMO 法で得られる分子全体の密度行列を基に、分子全体に対する Fock 行列と重なり行列を一度だけ計算し一般化固有値問題を解くことで、分子全体における分子軌道ならびに軌道エネルギーを近似的に求める計算方法を提案した [8]。こうして得られた分子軌道を FMO 法における分子軌道 (FMO-MO) と呼び、これまでに FMO-MO が HF 法で得られる分子軌道とほぼ一致することが示されている。FMO-MO 計算では HF 法にみられる反復処理 (SCF 計算) が必要ないため、分子全体に対する Fock 行列をただ一度だけ作成するだけでよく、高速な計算が可能である。またフロンティア軌道に対応するごく一部の固有値、固有ベ

クトルのみを計算すればよいことも FMO-MO 計算の特徴の一つである。もちろん本法で得られる MO は厳密な意味での SCF-MO ではなく、SCF-MO が満たさなければならないビリアル定理等も満たさない。あくまでも近似的な MO である。本法では HF 法だけでなく、CI 法などで得られる密度行列から Fock 行列を再構築して、近似的な MO を得ることが可能である。もちろん、その MO は、CI 法などで得られた密度行列を対角化して得られる自然軌道 (Natural Orbital: NO) とは異なる。

FMO-MO 計算の主な 3 つのステップ、(1) FMO 計算による密度行列作成、(2) 分子全体の Fock 行列作成、および (3) Fock 行列の対角化、のうち、(1) は、並列処理による高速化が可能である。また、(2) の計算では、必要となる 2 電子積分計算の計算量が $O(N^4)$ となりコストが大きい、個々の 2 電子積分は独立に計算できるため、こちらも並列計算による高速化が可能である。また、(3) の行列の対角化に対しても、化学的には固有値が 0.0 (HOMO-LUMO) 付近の数個の固有値・固有ベクトルのみが必要であるので、固有値の範囲の指定ができる櫻井-杉浦法 [9] のような並列化によって高速化が可能な方法の利用ができる。

本研究で用いている櫻井-杉浦法は n 次対称行列 $A \in R^{n \times n}$ 、及び n 次正定値対称行列 $B \in R^{n \times n}$ をもつ大規模一般化固有値問題: $Ax = \lambda Bx$ (1) の解法の一つである。本法は (1) を複素平面上の指定した領域内の固有値のみをもつ小規模一般化固有値問題に帰着させ、固有値、及び固有ベクトルを求める方法である。行列 A, B 、及び非零ベクトル $w, b \in R^n$ で構成される関数 $f(z)$ を、 $f(z) \equiv w^H(zB - A)^{-1}b$ と定義し、複素平面上に中心 $\gamma \in R$ 、半径 ρ の円周 Γ を配置する。ここで、 Γ 内には m 個の相異なる固有値 $\lambda_1, \lambda_2, \dots, \lambda_m$ が含まれているとする。さらに、 Γ 上に N 個の等間隔点 $\omega_0, \omega_1, \dots, \omega_{N-1}$ をとり、これらの点における関数値 $f(\omega_0), f(\omega_1), \dots, f(\omega_{N-1})$ を計算する。 N 個の関数値は、 Γ 内の m 個の固有値のみをもつ小規模一般化固有値問題に帰着させる際に用いられる。関数値 $f(\omega_0), f(\omega_1), \dots, f(\omega_{N-1})$ の計算には、複素対称行列をもつ N 本の連立一次方程式: $(\omega_j B - A)y_j = b, j = 0, 1, \dots, N-1$ の求解が必要となる。これにより、大規模固有値問題が、固有値の範囲を限った小規模固有値問題に帰着され、高い並列性をもつようになる。

以下では大規模なクラスター型並列計算機である AIST Super Cluster (ASC) [11] を用いた Lysozyme 分子とアデニンとチミンの核酸対で作られたモデル DNA 分

子に対する FMO-MO 計算を報告する。

3 FMO-MO 計算

3.1 プログラムと計算機

FMO 法の計算には、国立衛生研の中野らが開発した並列 FMO 計算プログラム、ABINIT-MP[12] を用いた。ABINIT-MP は Fortran90 でコーディングされている。FMO-MO 計算における全系の Fock 行列作成には、著者らが新たに成した 2 電子積分プログラム [13] を用いた。本プログラムは C 言語で記述されている。これら 2 つのプログラムは、MPI[14] によって並列化されている。

計算は、すべて ASC を用いて行った。ASC は 3 つのクラスター計算機システムで構成されており (Table 1)、FMO 計算による密度行列生成は、通信量が計算量に比べて比較的小さいため、Gigabit Ethernet で結合された F32 クラスターで計算を行った。また Fock 行列生成は、並列化効率の高い 2 電子積分計算で多くのプロセッサを利用したいこと、ならびに行列のリダクションの際に大きなネットワーク負荷がかかるため、2048 プロセッサが高速なネットワーク Myrinet で結合された P32 クラスターを用いた。一般化固有値問題の求値は P32 クラスター中の 1 プロセッサを用いて行った。F32 クラスターでは、コンパイラとして Intel Fortran Compiler (version 8.0)[15] を、また MPI ライブラリは MPICH(version 1.2.5.2)[16] を用いた。一方、P32 クラスターでは、コンパイラとして PGI コンパイラ (バージョン 5.1.3) [17] を、また、MPI ライブラリは、高速なクラスターとしての利用を容易にするために P32 クラスターに導入されている Score クラスターシステムソフトウェア [18] が提供するものを用いた。

Table 1. Detail of three clusters on AIST Super Cluster

Cluster	CPU	#proc(node)	Memory (GB/node)
P32	Opteron (model 246, 2.0GHz)	2048(1024)	6
M64	Itanium2 (1.3GHz)	512(128)	16
F32	Xeon (3.06GHz)	536(256)	4

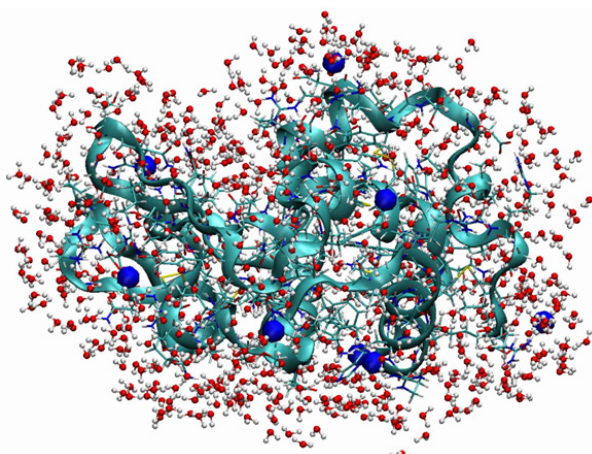


Figure 1. Structure of Lysozyme with 713 H₂O (<3.5 Å)

3.2 モデル分子

モデル分子は、卵白 Lysozyme (PDB ID=1LZT、アミノ酸 129 残基、1961 原子) と 40 個のアデニン-チミン対で作られたモデル DNA 分子である。Lysozyme は脊椎動物の細胞や分泌物中に広く分布し、植物性細菌の細胞壁を破壊する、殺菌作用のある酵素である。機能がよく知られており、タンパク質の中では比較的小規模ではあるが、それでも従来の MO 法では計算が非常に困難な大きさを持つ分子である。モデル DNA 分子は Lysozyme 分子のほぼ半分のサイズの系である。

Lysozyme 分子では、まず Protein Data Bank (PDB) の構造に水素を付加し、対イオンとしての Cl⁻ イオンと水分子を加えて古典分子動力学計算を行って平衡化した溶媒和構造を得た。このとき、ポテンシャルは Amber ポテンシャルを仮定した。それぞれの溶媒和構造を用いて Lysozyme 及び対イオンから最近接距離が 3.5, 5.0, 10.0Å 以内の水分子のみを取り込んで FMO-MO 計算に用いる構造を作成した。溶媒和構造に含まれた水分子の数はそれぞれ 365, 713, 2096 分子であり、Lysozyme を含めた全原子数はそれぞれ 3062, 4109, 8258 原子となった。また FMO/HF/STO-3G 基底関数の数は最大で 20758 であった。また Lysozyme のみの系の計算も行った。Figure 1 に Lysozyme に対イオン及び 5.0Å 以内の水 713 分子を加えた構造を示した。

モデル DNA 分子の構造は、アデニンとチミンの核酸対 40 を用い人工的に作成した。Insight II[19] によって作成した標準的な B 型二重螺旋構造に、対イオンとして Na⁺ を付加し、対イオンの位置を古典力場 charmm22[20] を用いて最適化した。

Table 2. E-time of FMO-MO calculation for Lysozyme and 2096 H₂O system

	CPU	#proc	E-time
FMO Calc.	Opteron ^a	128	1.7h
Fock Matrix Gen.	Opteron ^a	256	3.9h
Eigen Value & Vector			
Cholesky decomposition+			
Householder-bisec	Opteron ^a	1	4.1h
Sakurai-Sugiura	Xeon ^b	1	44.5m
	Xeon ^b	128	1.9m

^aOpteron :model 246, 2.0GHz, ^bXeon :3.06GHz

Table 3. E-time of FMO-MO calculation for a model DNA system

	CPU	#proc	E-time
FMO Calc.	Opteron ^a	640	19.7m
Fock Matrix Gen.	Opteron ^a	256	30.7m
Eigen Value & Vector			
Sakurai-Sugiura	Opteron ^a	16	5.7m

^aOpteron :model 246, 2.0GHz

C2 の対称性をわずかに崩してある。全原子数は 2636、基底関数は 10108 (FMO/HF/STO-3G) であり、Lysozyme と水分子 2096 分子系の約半分の規模である。

4 計算結果

4.1 FMO-MO 計算の経過時間

Lysozyme 分子と水分子 2096 分子系 (20758 基底関数) の FMO-MO 計算の経過時間を Table 2 に示す。FMO 計算に Opteron 128CPU で 2 時間、全系の Fock 行列生成に Opteron 256CPU で 4 時間ほどかかっている。固有値問題を従来法で Opteron 1CPU で行うとさらに 4 時間必要である。これは、ほぼフォック行列生成と同じ経過時間であり、従来はフォック行列生成の陰に隠れていた固有値の計算時間が、行列生成の並列処理による高効率化によって顕在したことに他ならない。幸いなことに必要最小限の固有値、固有ベクトルの計算が可能で、高度な並列処理が可能な櫻井-杉浦法を用いると 1CPU で 1 時間、128CPU ではほんの 2

分ほどで終了する。

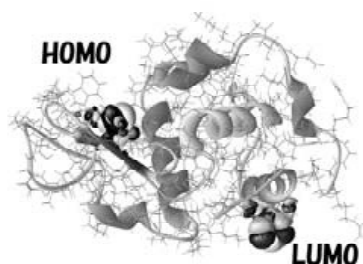
2 万次元ほどの分子軌道計算がたかだか Opteron 256CPU のクラスタを用いると 6 時間ほどで実行できることが判る。残念ながらこのサイズの系はもはや 1CPU で実行される従来法のプログラムでは様々な制限のために解くことができないサイズである。

Table 3 には、モデル DNA の経過時間を示した。Fock 行列の計算は 256CPU を用いて、ほぼ 30 分で終了しており、Lysozyme と水分子 2096 分子系の計算のほぼ 8 分の 1 の計算時間となっている。この程度の系の MO 計算が Opteron 640CPU のクラスタで 1 時間程度で実行可能である。

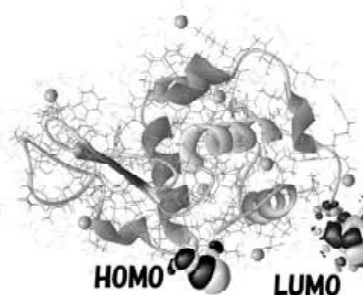
4.2 分子軌道と溶媒和

Figure 2 に FMO-MO 計算によって得られた Lysozyme の HOMO 及び LUMO を示した。SCF-MO と FMO-MO のように Virial 定理を満たさない MO は、軌道エネルギーの値が違うが、ノード関係や空間的広がりは全く同じである。水分子を含めることによって、HOMO や LUMO の位置が大きく異なってくるのがわかる。水を多く取り込むことで巨大分子の HOMO や LUMO のエネルギー準位の付近には、多くの MO ができてきており、そのため、HOMO や LUMO のみの軌道エネルギー、及び一つ溶媒和構造からその位置や安定化の議論をすることは非常に危険である。溶媒分子を陽に考慮する Super Molecule 法を用いて水和構造を見る場合には、十分な溶媒分子を考慮することが必要であることを伺わせる。Figure 2 は、2007 年現在世界最大の系の MO である。

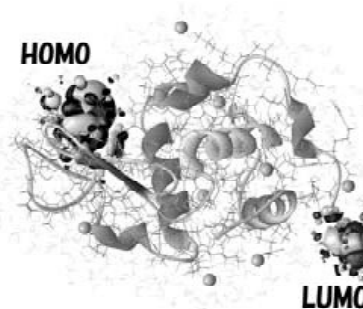
モデル DNA の HOMO と LUMO を Figure 3 に示した。この系では、HOMO が DNA 鎖の中央に大きな振幅を持ち、LUMO が DNA 鎖端に大きな振幅を持つ。 C_2 の対称性が完全に満足されれば、LUMO は DNA 鎖の両端に大きな振幅を持つが、わずかに C_2 の対称性を崩してあるので、片方によっている。ここでは図示しないが、LUMO の一つ上の LUMO+1 は、LUMO とは反対側の DNA 鎖端に大きな振幅を持つ。



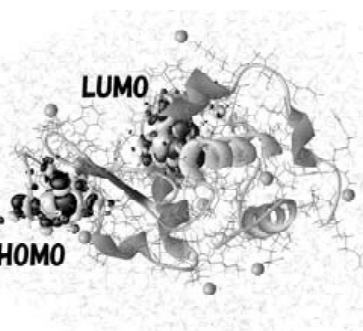
(a) HOMO-LUMO of Lysozyme



(b) HOMO-LUMO of Lysozyme with 365 H₂O (<3.5)



(c) HOMO-LUMO of Lysozyme with 713 H₂O (<5.0)



(d) HOMO-LUMO of Lysozyme with 2096 H₂O (<10.0)

Figure 2. HOMO-LUMO of Lysozyme with H₂O

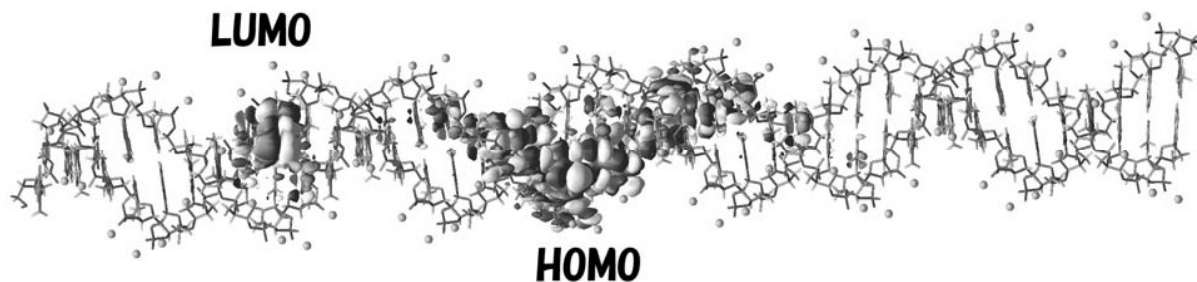


Figure 3. HOMO-LUMO of model DNA

4.3 軌道エネルギーレベルと溶媒和

FMO-MO 法によって得られた一つの溶媒和構造におけるそれぞれのモデルでの Lysozyme の HOMO-LUMO 近傍の軌道エネルギー分布を Figure 4 に示した。占有軌道と非占有軌道、それぞれ 40 本ずつの軌道エネルギーを示し、特に HOMO 近傍 20 本と LUMO 近傍 20 本、合計 40 本を黒で示した。まず、Lysozyme のみ（真空中のモデル）の場合には、HOMO-LUMO ギャップが -0.2 hartree 近辺なのに対して、対イオンを取り込むことにより、モデルの全系の電荷が中和され、HOMO-LUMO ギャップが 0.0 hartree の位置へ移動した。また、 3.5\AA 以内の結果では、HOMO 近傍の 3 本の MO が非常に高い軌道エネルギーを示しているが、これは対イオンの Cl^- イオン上に局在化しており、対イオンへの溶媒和が不十分であることによる結果である。溶媒和の定量的議論には十分な量の溶媒分子を考慮する必要がある。

HOMO 及び LUMO 近傍の 20 本の MO が Lysozyme のみの場合にはそれぞれ $0.130, 0.091$ hartree のエネルギー幅の間に分布しているのに対し、 10.0\AA 以内の水分子を取り込んだ系では、それらがそれぞれ $0.062, 0.044$ hartree とより狭い領域へ分布していることが分かる。

また HOMO-LUMO エネルギーギャップも Lysozyme のみの場合には 0.125 hartree なのに対し、 10.0\AA 以内の水分子を取り込んだ系では 0.376 hartree と大きく、広がっている。これは、対イオン及び水分子を含まない系では、荷電アミノ酸が溶媒和による安定化を受けないため、不安定な軌道エネルギーを持っているためである。

HOMO 及び LUMO 近傍の軌道エネルギーが密になっているのは、Lysozyme の MO に加えて、溶媒分子や対イオンの MO も加わったことも一因である。

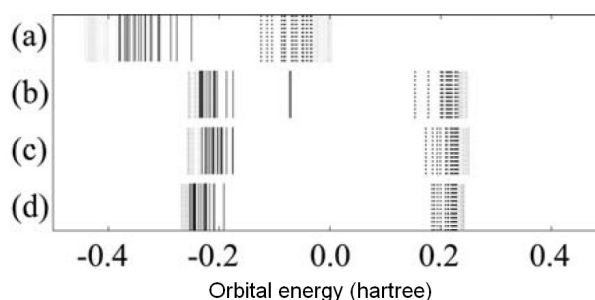


Figure 4. MO energy level close to HOMO-LUMO region. (a) Lysozyme, (b) Lysozyme with 365 H_2O ($<3.5\text{\AA}$), (c) Lysozyme with 713 H_2O ($<5.0\text{\AA}$), (d) Lysozyme with 2096 H_2O ($<10.0\text{\AA}$)

5 まとめ

まず小さなタンパク質である Lysozyme とモデル DNA 分子の分子軌道を、我々が開発している FMO-MO 法を用いて 640CPU 程度のクラスタ計算機で求め、その経過時間と溶媒和について報告した。

その結果、これまでは計算が困難であった、基底関数が 20,000 を越えるような巨大分子に対する MO 計算が、数時間程度という非常に短い時間で行うことを示すことができた。タンパク質や DNA などの生体分子の反応に対して、分子軌道計算による理論的な反応性予測が、FMO-MO 法を用いることで比較的短時間に行える可能性があることがわかった。

また、従来の MO 計算では 2 電子積分計算 (Fock 行列作成) の計算量に紛れて見えていなかった一般化固有値、固有ベクトルの計算時間が、分子サイズが大きくなると大きなウェイトを占めることがわかった。今回の Lysozyme に対する計算では、無視できない計算

時間を固有値、固有ベクトルの計算に費やしており、並列化効率の高いFMO計算、ならびにFock行列作成のためにより多くの計算機資源を使った場合には、さらに固有値、固有ベクトルの計算時間の割合が増加するはずである。したがって、高速なFMO-MO計算のためには、Fock行列作成の高速化のみならず、一般化固有値問題を高速に解く手法が求められており、櫻井 杉浦法の一層の改良を進めている。

溶媒分子の取り扱い方により、HOMO-LUMOの位置が大きく変わり、溶媒和の定量的議論には十分な量の溶媒分子を考慮する必要があることが判った。より大規模な系の計算の必要性が増している。

本研究は独立行政法人科学技術振興機構 (JST) が行う戦略的創造研究推進事業 (CREST) の研究領域「シミュレーション技術の革新と実用化基盤の構築」の研究プロジェクト「グリッド技術を用いた大規模分子シミュレーションプログラムの開発」の支援による。また、すべての計算は独立行政法人産業技術総合研究所 (AIST) の AIST super cluster を用いて行った。

参考文献

- [1] References in, A. Szabo and N. S. Ostlund, MODERN QUANTUM CHEMISTRY (Introduction to Advanced Electronic Structure Theory), Dover Publications Inc., Mineola, New York.
- [2] References in, 米澤貞次郎、永田親義、加藤博史、今村詮、諸熊奎治, 三訂 量子化学入門(上), 化学同人.
- [3] K. Kitaura, T. Sawai, T. Asada, T. Nakano, and M. Uebayashi, *Chem. Phys. Lett.*, **312**, 319 (1999).
- [4] K. Kitaura, E. Ikeo, T. Asada, T. Nakano, and M. Uebayashi, *Chem. Phys. Lett.*, **313**, 701 (1999).
- [5] T. Nakano, T. Kaminuma, T. Sato, Y. Akiyama, M. Uebayashi, and K. Kitaura, *Chem. Phys. Lett.*, **318**, 614 (2000).
- [6] K. Fukuzawa, K. Kitaura, K. Nakata, T. Kaminuma, and T. Nakano, *Pure Appl. Chem.*, **75**, 2405 (2003).
- [7] T. Ikegami, T. Ishida, D. G. Fedorov, K. Kitaura, Y. Inadomi, H. Umeda, M. Yokokawa, and S. Sekiguchi, *Proceedings of International Conference for Supercomputing and Networks, SC2005*, Seattle, USA, November, 2005.
- [8] Y. Inadomi, T. Nakano, K. Kitaura and U. Nagashima, *Chem. Phys. Lett.*, **364**, 139 (2002).
- [9] T. Sakurai and H. Sugiura, *J. Comput. Appl. Math.*, **159**, 119 (2003).
- [10] H. Sekino, Y. Sengoku, S. Sugiki, and N. Kurita, *Chem. Phys. Lett.*, **378**, 589 (2003).
- [11] AIST Super Cluster homepage
<http://unit.aist.go.jp/tacc/en/supercluster.html>
- [12] ABINIT-MP homepage,
<http://moldb.nihs.go.jp/abinitmp/>
- [13] Y. Inadomi, T. Nakano, K. Kitaura, U. Nagashima, High- Performance Computing and Networking, 9th international conference, HPCN Europe 2001, Amsterdam, The Netherlands, June 2001.
- [14] MPI: A Message-Passing Interface Standard, MPI forum, 1995.
- [15] Intel Fortran Compiler version 8.0: Copyright © 2001-2004 Intel Corporation.
- [16] MPICH homepage,
<http://www-unix.mcs.anl.gov/mpi/mpich/>
- [17] PGI Workstation 5.1: Copyright © 2003, The Portland Group™ Compiler Technology, STMicroelectronics, Inc.
- [18] Score homepage
<http://www.pccluster.org/index.html.en>
- [19] Insight II homepage
<http://www.accelrys.com/products/insight/>
- [20] J. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. W. E. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorcikiewicz-Kuczera, D. Yin, M. Karplus, *J. Phys. Chem. B*, **102**, 3586 (1998).

Molecular Orbital Calculation for Large Molecule

Toshio WATANABE^{a,b}, Yuichi INADOMI^c, Takayoshi ISHIMOTO^{a,b}, Hiroaki UMEDA^{a,b},
Tetsuya SAKURAI^{a,b,d} and Umpei NAGASHIMA^{a,b*}

^aCore Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency (JST)
Kawaguchi Center Building, 4-1-8, Honcho, Kawaguchi, Saitama 332-0012 Japan

^bResearch Institute for Computational Science, National Institute of Advanced Industrial Science and Technology
1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

^cComputing & Communications Center, Kyushu University
3-8-33-710 Momochihama, Sawara-ku, Fukuoka 814-0001 Japan

^dInstitute of Information Sciences and Electronics, University of Tsukuba
1-1-1 Ten-nodai, Tsukuba, Ibaraki, Japan

**e-mail: u.nagashima@aist.go.jp*

We have been developing a computational tool to obtain the molecular orbitals for large molecules such as proteins and molecular clusters without excessive calculation costs. In our method, the entire Fock matrix is generated using density matrix obtained from the fragment molecular orbital method, which is applicable to large systems and suitable for parallel processing. To solve the large scale generalized eigenproblem, we use the Sakurai-Sugiura method. This method solves several linear equations which have large granularity and is well suited to master-worker type execution. It is sufficient for parallel processing on computers of distributed memory parallel architecture. The method is favorable for calculation of only a small number of eigenvalues and corresponding eigenvectors of a large scale matrix. Our method has high parallelization efficiency and the communication cost is negligible to the total calculation costs. Thus, this is one of the right applications for using Grid technology.

Elapsed times to obtain MOs close to HOMO-LUMO of Lysozyme (129 amino-acid residues and solvent molecules, total 8258 atoms) with FMO/HF/STO-3G (20758 basis functions) and model DNA (40 A-T base pairs, total 2636 atoms with FMO/HF/STO-3G (10108 basis functions)) are less than about 5 hours and 1 hour, respectively, on only 128 Dual Opteron cluster. The position of HOMO and LUMO in the Lysozyme and water cluster case is strongly dependent on distribution and the number of solvent water molecules. Careful treatment for solvent molecules is required to have consistent results. In the case of model DNA, HOMO and LUMO are located at the center and the terminal of the DNA chain, respectively.

Keywords: Fragment Molecular Orbital method, Large scale molecule, Parallel processing, Lysozyme, Model DNA

