

《特 集》

タンパク質の第一原理電子状態計算 —生体化学反応の理論的解析を目指して—

中 野 達 也*・高 島 一**・長 嶋 雲 兵***

ABSTRACT Information of electronic state of biomolecules such as protein is indispensable to understanding of property of biomolecules and mechanism of biochemical reactions. Molecular orbital calculation is applied to simulate its electronic states. However molecular orbital calculation for large scale system as biosystem is very difficult to execute even on supercomputer because the amount of calculation is proportionally increased with the fourth power of system size. In order to realize low cost and personal execution of ab initio MO, we are developing a special purpose computer for molecular orbital calculation: MOEngine to increase calculation power. A novel method: the fragment MO method was also developed to decrease the amount of computation inn MO calculation. We report that the calculation of electronic state of biomolecule will be inexpensively and personally executable using MOEngine and the fragment MO method showing some examples.

1. はじめに

ポストゲノムの時代を迎え、様々な生体化学反応の理解の重要性が高まっている。特にタンパク質の構造だけでなくその機能を解析する上で、タンパク質の電子状態や酵素との化学反応機構の理解は欠くことができない。それには、量子力学的電子状態計算を基盤とする分子科学的シミュレーションが不可欠である。物質の電子状態の理論的計算は、分子軌道法として知られている方法を用いて行われている。

1929年 Dirac¹⁾が『量子力学の一般原理は、今やほぼ完成し、残っている不備な点と言えば、相対論の考えと正確に合致させることである。不都合が生じるのは高速粒子が関与する場合であり、従って原子分子の構造や普通の化学反応については重要ではない。実際、こうした問題では、質量の相対論的速度依存性を無視し、電子や原子核の間に働く力として Coulomb 力だけを仮定しても、だいたい十分正確なのである。要す

るに、物理学の大部分と化学の全体の数学的理論に必要な基礎的物理法則は完全にわかっているということであり、困難は、ただ、これら法則を厳密に適用すると複雑すぎて解ける望みのない方程式に行きついてしまうことである。したがって、量子力学を応用するための実用的な近似方法を発展させ、過度の計算を行うことなく複雑な原子集合体の主だった性質を説明できるようにすることが望ましい。』と述べたが、分子軌道法は量子力学の完成とともに「量子力学を応用するための実用的な近似方法」として広く原子分子の電子状態の計算に用いられてきている。1999年には、J. A. Pople が実験データから導き出される経験的パラメータを一切用いない非経験的分子軌道法の普及の功績によりノーベル賞を受けた。非経験的分子軌道法は21世紀の生物科学や新規物質探索にとって強力なシミュレーション技法として、さらなる発展が求められている技術である。

経験的パラメータを一切用いない非経験的分子軌道法は、化学反応一般を取り扱うことが可能であるため、さまざまな機能性分子の設計や開発に対して最も基本的でかつ重要なシミュレーション手法である。しかし、その計算量はもっとも単純なハートリーフォック計算の場合でも、系のサイズ（正確には用いる基底関数）の4乗に比例するため、生体内や固体表面での化学反応解析等の大規模系の電子状態計算には膨大な

The first principle electronic state calculation for protein -for theoretical analysis of biochemical reaction-. By Tatsuya Nakano (National Institute of Health Science), Hajime Takashima (Taisho Pharmaceutical Co., Ltd.) and Umpei Nagashima (National Institute for Advanced Interdisciplinary Research).

*国立医薬品食品衛生研究所

**大正製薬㈱

***通産省工業技術院産業技術融合領域研究所

計算コストが必要になる。現実的には、非経験的分子軌道法は、シミュレーションしたい大きさの大規模分子系ではなく、それを簡素化した小規模モデル分子への適用がせいぜいである。1990年代前半までは数百基底（数十原子）の分子軌道計算でさえ、その当時のスーパーコンピュータシステムを必要としていた。そのため、現在でも分子設計に必要な情報を十分な精度で得ることができないでいる。

1990年代後半における計算機性能の急速な進歩、特にスーパーコンピュータや高性能ワークステーションクラスタのような大規模並列計算機の著しい進歩により、ようやく生体分子などの高分子を意識した1,000基底を超える分子軌道計算^{2~6)}が行われつつある。とはいえ、これらの計算機は高価であり、かつ設備の維持管理も大変であるため、研究者が研究室レベルで「現実を反映した大規模分子系」の分子軌道計算を実現することは容易なことではない。「現実を反映した大規模分子系」の分子軌道計算を、「低コスト＝パーソナルユース」で実現するためには、計算量を軽減するための近似法を取り入れ、さらに計算機の性能を飛躍的に向上させることが必要である。

計算機の性能を飛躍的に向上させるための試みは、非経験的分子軌道法の持つ高い並列性を背景にして、パソコンなどによる並列分散処理システム（PCクラスタ）の構築⁷⁾や専用計算機の開発等が進められてきている。現在我々のグループも大規模分子の非経験的分子軌道計算専用機として、分散小メモリ型並列アーキテクチャである分子軌道計算専用並列計算機 MOE の開発^{8~10)}に取り組んでいる。しかしながら、そのような専用計算機を用いたとしても生体分子や物質材料等の大規模系の分子軌道計算を行うことは依然として難しい。

他方、計算量を軽減するためには2つの方法が考えられる。一つは経験的または半経験的分子軌道法のように、必要な計算をまともにする代わりに実測値から導き出されるパラメータを用いて計算量を削減する方法である。この方法は演算量が大幅に軽減されるにも関わらず、炭化水素などの系では計算される物理量が実測値をよく再現することから、MOPAC¹¹⁾のように非常に広範囲に利用されている。有機化学反応におけるフロンティア軌道理論や遷移状態理論などの普及は、経験的または半経験的分子軌道法なしには考えられなかった。しかしながら、経験的または半経験的分子軌道法は、生体反応のように実測が難しい系や金属のように周辺環境によって様々な状態を容易にとる

系には利用できないことや、計算結果の信頼性にばらつきがあることがよく知られている。

他は、実測値をパラメータとして用いる代わりに、本論文で概要を説明するフラグメント MO 法^{12~16)}のように、ハミルトニアンに大胆な近似を導入することで計算量を軽減する方法である。この方法の演算量は、構成原子数（正確には基底関数）のほぼ2乗に比例する演算量となる。このように近似ハミルトニアンを用いる計算法は、計算量や結果の信頼性などの面から非経験的分子軌道法と半経験的分子軌道法の間位置づけられている。このような方法は経験的パラメータを陽に含まないことから、第一原理分子軌道計算または第一原理計算と呼ばれている。

我々はタンパク質などの大規模分子系の電子状態計算とそれに続く分子動力的シミュレーションを可能とするため、専用分散並列計算機を開発して計算速度の向上をはかり、さらに第一原理分子軌道計算法の一つであるフラグメント MO 法を用いて計算量を軽減することで、タンパク質の様な「現実を反映した大規模分子系」の分子軌道計算を「低コスト＝パーソナルユース」で実現するシステムの開発を行っている。本稿ではその概要と成果を報告する。

2. 分子軌道計算専用機 MOE (Molecular Orbital Engine) の開発

本節では、非経験的分子軌道法について概略し、つづいて現在開発中の分子軌道計算専用機 MOE の開発について解説する。

2.1 非経験的分子軌道法の概要—閉殻系の Hartree-Fock 法—

現在広く用いられている非経験的分子軌道法の出発点は Born-Oppenheimer 近似を仮定した、電子に関する Schrödinger 方程式である。原子核の種類や位置は、パラメータとして取り扱われる。

$$H\Psi = E\Psi \quad (2-1)$$

$$H = -\sum_i \frac{1}{2} \nabla_i^2 - \sum_i \sum_A \frac{Z_A}{r_{Ai}} + \sum_{i \neq j} \frac{1}{r_{ij}} + \sum_{A \neq B} \frac{Z_A Z_B}{R_{AB}} \quad (2-2)$$

ここで、 H はハミルトニアン、 i, j は電子の番号、 A, B は原子核の番号、 Z_A は原子核 A の電荷（原子番号）を表す。式(2-2)は第1項から順に電子の運動エネルギー、原子核による引力エネルギー、電子間クーロン反発エネルギー、原子核間の反発エネルギーを表す。また、式(2-1)の E （エネルギー固有値）は分子の全エネルギーを、 Ψ は分子内の全ての電子状態を表す

全波動関数である。なお、単位は全て原子単位 (atomic unit) である。

Schrödinger 方程式は多体の線形方程式であり解析的に解くのは不可能なため、これを非線形の一体問題に近似して計算する。簡単のため、以下では全て電子数 $2n$ の閉殻系を考える。(閉殻系の Hartree-Fock 法) まず、全波動関数 Ψ を次のようなスレーター行列式で表現する。

$$\Psi = |\phi(1)\phi(2)\cdots\phi(2n)\rangle \quad (2-3)$$

$\phi(1), \phi(2), \dots, \phi(2n)$ は、 $1, 2, \dots, 2n$ 番目の電子の一電子分子軌道である。さらに、 a 番目の分子軌道 ϕ_a を式(2-3)のように原子軌道 χ の一次結合で表す。

$$\phi_a = \sum_{I=1}^N C_{aI} \chi_I \quad (2-4)$$

N は原子軌道の総数である。原子軌道の組は基底関数系 (basis set) と呼ばれ、分子軌道計算を行なう際に入力データとして与えるものである。基底関数は、その選び方により計算の精度が左右される重要な因子であり、通常 $\exp(-\zeta^2)$ で表される GTO (Gaussian Type Orbital) の線形結合で表現される。式(2-4)のような近似を、LCAO 近似 (Linear Combination of Atomic Orbital) と呼ぶ。式(2-2)～式(2-4)を式(2-1)に代入し、式(2-1)の全エネルギー E が最小 (極小) となるような Ψ 、すなわち式(2-4)の係数行列 C を求める手法の一つとして、Hartree-Fock-Roothaan の変分法¹⁷⁾が知られている。導出過程は成書¹⁸⁾にゆずり、変分法を用いた結果として次のような式が得られる。

$$\sum_{I=1}^N F_{IJ} C_{aI} = \sum_{I=1}^N S_{IJ} C_{aI} \epsilon_a \quad (2-5)$$

ϵ_a は a 番目の軌道の軌道エネルギーである。このように、フォック行列 F を対角化することにより、固有ベクトルである係数行列 C と固有値であるエネルギー行列 ϵ を得る。フォック行列 F 及び重なり行列 S の各要素は、それぞれ以下の様に表される。

$$F_{IJ} = T_{IJ} + V_{IJ} + G_{IJ} \quad \text{フォック行列} \quad (2-6)$$

$$S_{IJ} = \int \chi_I(1) \chi_J(1) dr_1 \quad \text{重なり行列} \quad (2-7)$$

$$T_{IJ} = -\frac{1}{2} \int \chi_I(1) \chi_J(2) dr_1 \quad \text{運動エネルギー積分} \quad (2-7)$$

$$V_{IJ} = -\sum_A \int \chi_I(1) \frac{Z_A}{r_{1A}} \chi_J(1) dr_1 \quad \text{核引力積分} \quad (2-9)$$

$$G_{IJ} = \sum_{K=1}^N \sum_{L=1}^N P_{KL} \left\{ (IJ, KL) - \frac{1}{2} (IK, JL) \right\} \quad \text{フォック行列(二電子項)} \quad (2-10)$$

$$P_{IJ} = 2 \sum_a^{\text{occ}} C_{aI} C_{aJ} \quad \text{密度行列} \quad (2-11)$$

$$(IJ, KL) = \iint \chi_I(1) \chi_J(1) \frac{1}{r_{12}} \chi_K(2) \chi_L(2) dr_1 dr_2 \quad \text{二電子積分} \quad (2-12)$$

ここで、式(2-10)の第一項はクーロン項、第二項は交換項と呼ばれている。式(2-12)の添字 1 と 2 は、それぞれ電子 1 と電子 2 を表す。

二電子積分は、その式の形から分かる通り、 I と J , K と L , IJ と KL の入れ替えに対する対称性を有する。

$$\begin{aligned} (IJ, KL) &= (IJ, LK) = (JI, KL) = (JI, LK) \\ &= (KL, IJ) = (KL, JI) = (LK, IJ) = (LK, JI) \end{aligned} \quad (2-13)$$

二電子積分 (IJ, KL) の添え字が 4 つあるため、非経験的分子軌道法の演算量は原子数 (正確には基底関数の数) の 4 乗に比例するということになる。この積分は添え字の順序に独立に計算が可能のため、分子軌道計算は高度の並列性を持つ。

エネルギー E 及び係数行列 C を求めるには、式(2-6)～式(2-12)を用いてフォック行列を求めて式(2-5)を解けばよい。しかし、フォック行列にはあらかじめ答えとなるべき係数行列 C が入っている。そこで、実際の計算手順としては、図 1 に示すように、係数行列 C の初期値を適当に与えて、式(2-5)及び式(2-6)の計算を、係数行列 C (正確には密度行列 P) が収束するまで繰り返すことになる。収束後得られた状態は系内部で矛盾のない場 (Self Consistent Field) を作る。

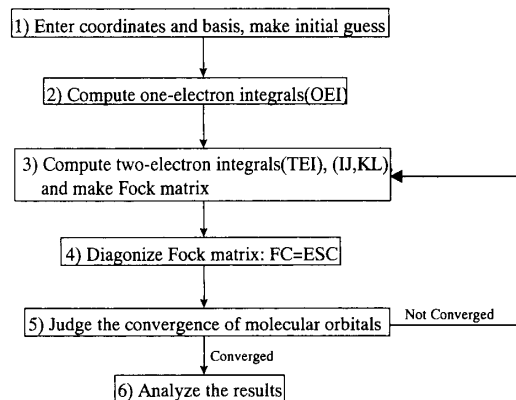


図 1 Typical flow charts for an ab initio molecular orbital calculation.

ていると考えられるので、この計算方法はSCF法とも呼ばれる。

二電子積分はSCF繰り返し計算のたびに用いられるが、その値は常に同じである。そこで、従来のアルゴリズムでは最初に計算したものをディスクに保存しておき、積分が必要になるたびにディスクから読み出す方法（ディスクストレージ法）が主流であった。しかし、二電子積分の数は原理上基底数 N の4乗に比例して増加するためディスク容量が制限される事と、計算機の発達により演算時間とディスクアクセス時間が拮抗してきたため、現在では、SCF計算で積分が必要になるたびに二電子積分を計算し直し、求まった二電子積分を直ちにフォック行列に足し込む直接法¹⁹⁾が主流である。

いくつかのペプチド分子について実際の非経験的分子軌道計算を行った時の各ステップ別の計算時間結果を表1に示す。計算はSGI-CRAY社製R10000を用いて行い、基底関数は4-31Gを用いた。表中のG, A, Q, M, YはそれぞれGly, Ala, Gln, Met, Tyrを表す。表1に示すように全計算時間の実に95~99%以上を占めるのが図1に示したステップ³⁾、すなわち式(2-12)で二電子積分を求めて式(2-10)を計算するステップである。非経験的分子軌道計算の超高速化を実現するためには、いかに二電子積分を高速に計算するかが重要であるかが分かる。さらに、基底関数の数が多くなる（系のサイズが大きくなるか、もしくは精度の良い基底関数を使う）に従って二電子積分の計算時間の割合が大きくなる傾向が強まる。

次節では、計算性能向上のための分子軌道計算専用機の開発について説明する。

2.2 分子軌道計算機 MOE のシステム構成と性能

分子軌道計算は、並列性が極めて高く、またに電子積分計算は演算パターンが限られているため、そのパターンに特化した演算装置をもつ専用計算機が有効である。そのため、我々は、並列分子軌道計算機

MOE^{8~11)}の開発を続けている。MOのシステム構成を図2に示す。MOEはPCやWSをホストとし、それにIEEE1394(200 MB/sec.)によって結合されたアタッチドプロセッサ群よりなる。アタッチドプロセッサ群は1ボードあたり6の専用LSIが搭載され、ボードないのプロセッサユニットPUはPPRMLink²⁰⁾によって結合されている。6のPUのうち、1つはネットワークの管理を行い、他の5つは図1に示す3)二電子積分の生成、4)Fock行列要素の生成を行う。二電子積分の計算法は小原法²¹⁾である。浮動小数点演算は64ビットではなく、10000軌道以上の大規模系の分子軌道計算でも結果が化学的精度を失わないよう精度を向上させるために73ビットを用いている^{22,23)}。このように計算に必要な精度の演算が可能となるように演算器を開発できることが、専用計算機開発の一つのメリットである。

図3に演算性能の見積もりを示した。実線は100台のSGI-CRAY社製R10000をクラスタ構成で用いた場合であり、破線は100PU構成のMOEを用いた場合である。左の図は、有機分子の分子軌道計算に広く使われている6-31G**という基底関数を用いた場合であり、右側の図は、さらに計算精度を上げるために基底関数を拡張した場合である。これをみると、MOEは約20倍の性能を持つことが判る。また、計算精度の高い拡張基底の場合にその威力を発揮する。

次節では、計算性能の向上ばかりでなく、さらに専用計算機上での演算量削減の方法：積分カットオフについて説明する。

2.3 大規模系の分子軌道計算のために一積分のカットオフ

実際に非経験的分子軌道計算を行なう際、特に大規模系の計算において計算時間を短縮するために、実際に計算する二電子積分の個数を少なくすることを行う。この手段としては、式(2-13)で示した二電子積分の対称性を利用しユニークな値を持つ二電子積分のみを計算する方法があるが、それに加え、値が非常に小さいことをあらかじめ予測して、二電子積分の計算を積極的に排除する方法（カットオフ）も広く用いられる。大規模系における積分カットオフの効果は非常に大きい。

積分カットオフの物理的な意味としては、軌道 I と軌道 J の重なり部分にある電子1の電子密度や軌道 K と軌道 L の重なり部分にある電子2の電子密度が十分に小さければ、これらで表現される積分自体も十分に小さいであろうからその計算を省略する事ができ

表1 Computing times in every SCF calculation steps for some peptide molecules (sec).

	G	GA	GAQ	GAQM	GAQMY
Number of dimensions	55	110	207	316	427
Initial guess	0.1	0.7	4.4	15.0	43.7
One Electron Integral	0.1	0.5	2.3	7.8	14.4
Two Electron Integral and Fock Matrix	21.9 (96.9%)	268.5 (98.8%)	2241.3 (99.1%)	10994.7 (99.4%)	32624.7 (99.4%)
Diagonalization	0.1	1.0	6.9	28.4	103.3
Properties	0.1	0.5	2.8	9.7	24.2
Others	0.3	0.6	3.0	9.5	25.1
Total	22.6	271.8	2260.7	11065.1	32835.4

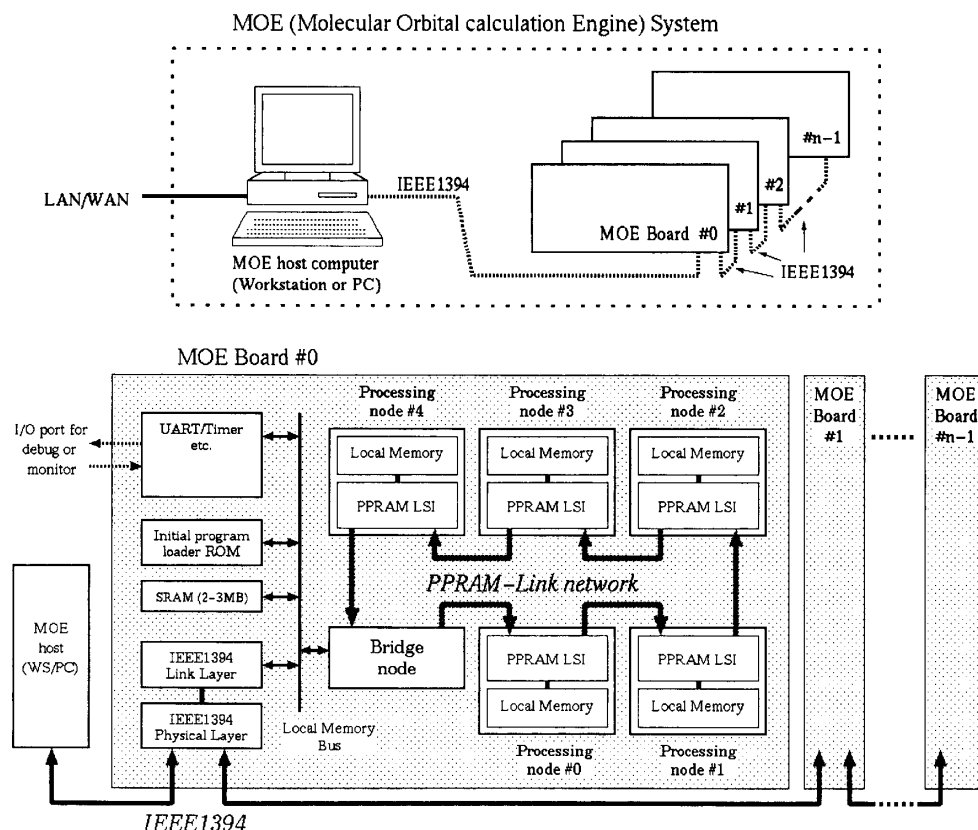


図2 MOEngine, a special purpose machine for accelerating molecular orbital calculations.

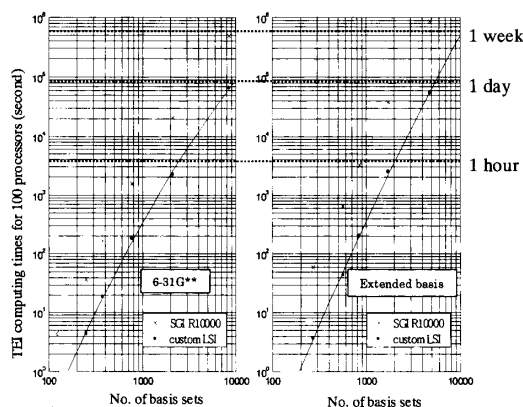


図3 Computing time for all ERI one times (seconds). Comparison between one MOEngine (100LSI) and 100 workstations

る, という風に解釈できる。

我々は, 各専用計算機 MOEngine のような各プロセッサのメモリが小容量 (数 MByte) しかない分散メモリ型並列アーキテクチャのもとで大規模フォック行列を高効率に計算するための新規アルゴリズムを考案した^{24,25)}。このアルゴリズムを用いると, 大規模系

では N を基底関数の数とすると, 演算量が $O(N^4)$ ではなく $O(N^2)$ となり, 通信量が $O(N)$ となる。

全エネルギー値に十分小さな影響しか与えない積分カットオフ閾値を 10^{-15} とした場合に, 実際の大規模分子でどの程度カットオフが行われるのかについて具体的にしめそう。二電子積分 (IJ, KL) は N^4 個あるが, カットオフを考慮した後に計算すべき二電子積分の数は,

$$(\alpha N \times N)^2 = (MN)^2 \quad (2-14)$$

になる。ここで α は, カットオフをおこなったときの実質的に生き残る軌道数の比率, つまりカットオフ生き残り率である。小分子から蛋白質程度の大きさまでさまざまなサイズの分子で, カットオフ生き残り率 α および生き残った基底数 M を見積もった結果を図4及び表2に示す。計算に用いた分子は, H_2O , ベンゼン (C_6H_6), アミノ酸である Arginine, GAQ (Gly-Ala-Gln), GAQMY (Gly-Ala-Gln-Met-Tyr), スクレオチド d(CGC), Gramicidin A 二量体 (30残基), BPTI (Bovine Pancreatic Trypsin Inhibitor, 57残基), Human Immunoglobulin IGG1 (99残基), Lysozyme (129残基), の10種類である。基底関数による違いを

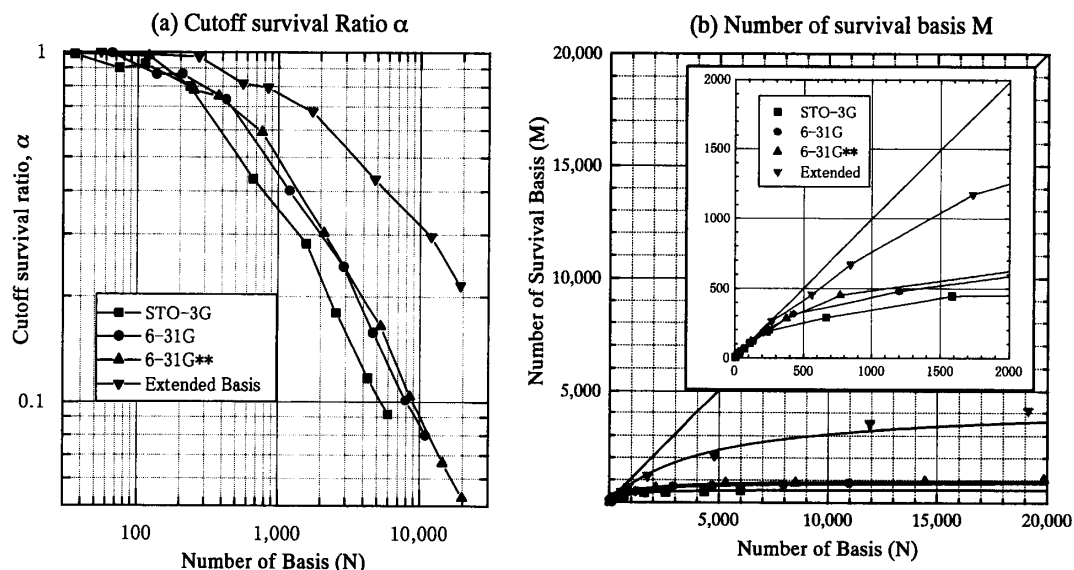


図4 Effects of Cutoff. (a) Cutoff survival Ratio α . (b) Number of survival basis M.

表2 Number of basis N and the number of cutoff survival basis M.

molecule	STO-3G		6-31G		6-31G**		Extended	
	N	M	N	M	N	M	N	M
H ₂ O	7	7	13	13	25	25	55	55
C ₆ H ₆	36	36	66	66	120	117	270	262
Arginine	74	67	136	119	250	196	560	457
GAQ	113	105	207	180	375	263	845	671
GAQMY	235	188	427	314	769	455	1733	1176
d(CGC)	668	288	1203	479	2106	635	4797	2069
Gramicidin	1587	448	2910	710	5295	869	11910	3525
BPTI	2572	462	4694	739	8504	891	19138	4121
IGG	4334	507	7940	803	14438	962	32476	4806
Lysozyme	6004	552	10967	881	19850	1044	44705	5578

調べるため、STO-3G (最小基底), 6-31G (Variance Double 基底), 6-31G** (6-31G 基底に分極関数を加えたもの), 6-311++G (3d, 2p) (分極関数や diffuse-function を多く取り込んだ拡張基底) 4 種類を用いた。これらは、順に精度が高い。積分カットオフの閾値 K は 10^{-15} を用いた。この計算で用いたプログラムは GAMESS²⁶⁾ である。

図 4(a) は、基底数 N の分子におけるカットオフ生き残り率 α (全ての添字での平均値) を両対数プロットした図である。小分子で基底数 N も小さい時 (100 以下) は、 α はおよそ 1 であり縮約基底関数で表現される積分のカットオフは殆どされない。しかしながら、分子サイズが大きくなり基底数が増えるにつれて急速に積分カットオフがなされるようになる。これは、より中心間距離が離れたガウス関数のペアが多くなるためである。10,000 次元程度になると、STO-3G, 6-31G, 6-31G** クラスの基底関数では、カットオフ生き残り率 α は 0.1 以下になっている。一方、拡張基底の場合、diffuse 関数や分極関数のような空間

的に広がった基底関数が多数用いられているため、10,000 次元でもカットオフ生き残り率 α が 0.3 程度と、他の基底関数に比べて少し大きな値である。このように、カットオフ生き残り率 α は基底関数の種類によって異なり、より精度の高く空間的に広がった軌道をもつ基底関数ほどカットオフされにくく、最小基底である STO-3G のような空間的広がりを持たない基底関数系ではカットオフされる割合が大きい (カットオフ生き残り率 α が小さい) 事が分かる。

図 4(b) は、基底数 N に対してカットオフで生存する平均基底数 M をプロットしたものである。図中の対角線は、カットオフが全くされない場合を表している。表 2 に平均基底数 M の数値を示す。基底数 N が小さい場合にはカットオフが殆どなくどの基底でも対角線上に近いが、基底数が大きくなると、精度の低い (広がった基底の少ない) 基底関数系から順番に対角線から外れ出す。そして、5,000 基底程度以上になると、カットオフ後に生存する基底数 M は拡張基底の場合を除いて殆ど増加しなくなり、ほぼ一定値になっている。タンパク質分子のように基底数 N が 10,000 を越えるような場合には、計算しなければならない二電子積分の数は $O(N^4)$ から $O(N^2)$ になる。しかも、精度の低い基底関数ほど $O(N^2)$ になりやすい傾向が見える。精度の低い基底は電子雲の歪みを表現するための空間的に広がった関数が少ないので、他の中心を持つガウス関数との重なりが小さくなりカットオフされやすいのがその理由である。

なお、ここで示したカットオフ生き残り率 α や生

き残った基底数 M は分子全体での平均値であり、原子の種類や位置、縮約シェルの種類（例えば、内殻に相当する $1s$ 軌道か、または p 軌道か）等によって、同じ分子内でもカットオフされる割合が異なる。6-31G 基底における Lysozyme 分子の場合、カットオフで生き残る基底数 M の平均値は881であるが、最も少ない場合には $M=91$ であり、逆に最も多い場合には $M=3409$ であった。このように、原子の種類や位置、軌道の種類などによって、カットオフで生き残る基底数 M はばらつきがある。

では、ある軌道に対してカットオフ操作を行なった時に生き残る軌道の空間的配置はどのようなものであろうか？ 一例として、6-31G 基底での Lysozyme 分子において、カットオフ操作で生き残る軌道の空間的広がりを示したのが図5である。分子のほぼ中央に位置する56番目 Leucine の δ 炭素原子（大きな球で表している）の $2p$ 軌道に対してカットオフを考慮した場合に、生き残る軌道が空間的にどのように分布しているかを調べた。カットオフ閾値は 10^{-15} である。生き残る軌道を一つでももつ原子を小さな球で示しているが、これらの多くは半径 $10\text{\AA}$ 以内に集まっている。しかし、原子の種類によっては、例えば $10\text{\AA}$ より離れていても生き残る軌道を持つものもあるし、逆に近くにあっても生き残る軌道を全く持たないものもある。さらに、小さな球で示されている原子であっても、その原子に含まれる全ての軌道が生き残るわけではなく「少なくとも一つの軌道が」生き残っているだけであり、多くの軌道はカットオフされている。

このように、ある添字 I の軌道に対してカットオフにより生き残る軌道 J は I の近傍の原子に集中しており、遠くの原子には殆ど存在しない。従って、蛋白質のような大規模系では、カットオフされる割合は非常に大きくなり二電子積分の数は実質上 $O(N^2)$ になる

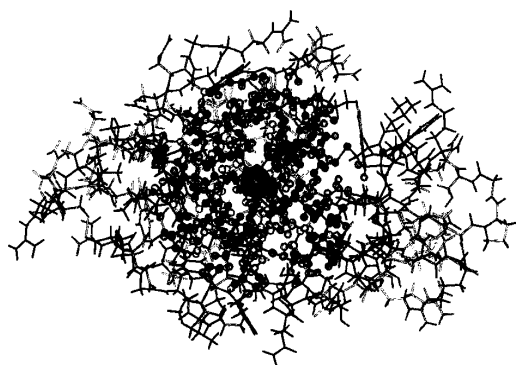


図5 Cutoff scheme in Lysozyme molecule.

ことが理解できる。

ここでは、詳細に説明することはしないが、大規模フォック行列を計算するための並列分散アルゴリズムでは、ホストプロセッサ間の転送量（速度）及びプロセッサエレメントのメモリ量の観点から、積分のカットオフ及びこれに伴う行列転送のカットオフを考慮する事が大事である。実際、我々が開発した新しいアルゴリズムを用いて^{24,25)}大規模分子でのカットオフの割合を調べる事と、カットオフを考慮する場合としない場合においては10倍以上転送量（速度）やメモリ量が異なることが分かった。

次節では、数百から数千の残基を持つようなタンパク質分子の分子軌道計算を実現するために、さらに大きな系に取り扱いを目指した分子軌道計算法であるフラグメント MO 法を紹介する。

3. さらに大規模な系の取り扱いのために—フラグメント MO 法—

分子軌道計算専用機を用いたとしても、タンパク質のような大規模分子系の分子軌道計算は、そのサイズの大きさからみて、まだまだ難しい。タンパク質等の生体高分子の多くは、約20種類のアミノ酸によって構成される不規則多次元高分子であり、そこにタンパク質の電子状態計算の難しさがある。北浦らの開発したフラグメント MO 法^{13~16)}は、タンパク質のような巨大分子をいくつかのフラグメントに分割し、フラグメントとフラグメントペアの計算から分子全体のエネルギーや電子密度を計算する方法である。このため計算時間が、通常の実験的分子軌道法では系の大きさの4乗に比例するのに対し、フラグメント MO 法では系の大きさのほぼ2乗に比例することになり、大規模系では計算量を大幅に短縮できる。また、ポテンシャルを完全な量子論で計算できること、任意性のあるパラメータを一切用いる必要がないこと、分子内と分子間ポテンシャルを同列に計算できるという特徴がある。

フラグメント MO 法に現れる分子積分は、分子軌道計算専用機で取り扱うことのできる積分の一部であるため、分子軌道専用計算機を用いることでさらに高速な計算の実行が可能である。以下では北浦らのフラグメント MO 法を紹介する。

3.1 フラグメント MO 法

北浦らの開発したフラグメント分子軌道法^{13~16)}では、分子の全電子エネルギー E を近似的に次のように計算する。

$$E = \sum_{I>J} E_{IJ} - (N-2) \sum_I E_I \quad (3-1)$$

ここで E_I はフラグメントの全電子エネルギー, E_{IJ} はフラグメントペアの全電子エネルギー, N はフラグメントの個数, I, J はフラグメントの番号である.

E_I を先に述べた閉殻系の非経験的分子軌道法で計算するとすると,

$$E_I = \sum_i 2h_{ii} + \frac{1}{2} \sum_{ij} \{2\langle ii|jj\rangle - \langle ij|ij\rangle\} + \sum_{j \neq I} \left\{ \sum_i V_{ji}^I + \sum_{ik} 4\langle ii|kk\rangle \right\} \quad (3-2)$$

$$V_{ji}^I = \langle i | \sum_{A \in J} \left(\frac{-Z_A}{r_A} \right) | j \rangle \equiv \langle i | V^J | j \rangle \quad (3-3)$$

ただし i, j はフラグメント I の MO, k はフラグメント J の MO である. さらに,

$$h' = h + \sum_{j \neq I} V^J \quad (3-4)$$

とまとめると, Fock 演算子 F' は,

$$F' = h' + \sum_{j \neq I} \sum_k 2\langle ii|kk\rangle \quad (3-5)$$

と表される.

静電ポテンシャル項を含む Hartree-Fock 方程式は対角化でき,

$$F' \phi_i = \epsilon_i' \phi_i \quad (3-6)$$

となる. ここで注意してほしいのは, (3-2) 式には, $\langle ii|jj\rangle$ 型の積分しか現れないことである. そのため, 先に説明した非経験的分子軌道法の「基底関数の 4 乗に比例する演算量 (二電子積分量)」に比べ, フラグメント MO 法では「基底関数の 2 乗に比例する演算量 (二電子積分量)」となり, 従来の非経験的分子軌道法に比べ, 大幅に演算量が減少する.

さらに, 核座標 R_a に関する全エネルギーを

$$E_I^a = 2 \sum_i h_{ii}^a + \frac{1}{2} \sum_{ij} \{2\langle ii|jj\rangle^a - \langle ij|ij\rangle^a\} + \sum_{j \neq I} \sum_i \sum_k 4\langle ii|kk\rangle^a \quad (3-7)$$

と書くと, E_I の核座標 R_a による微分 $\partial E_I / \partial R_a$ は,

$$\frac{\partial E_I}{\partial R_a} = E_I^a - 2 \sum_i S_{ii}^a \epsilon_i' - \sum_{j \neq I} 2S_{jk}^a \left(\sum_i 2\langle ii|kk\rangle \right) \quad (3-8)$$

で表される. $\partial E_{IJ} / \partial R_a$ についても同様である. (3-8) 式を用いると, フラグメント MO 法に基づく分子動力学シミュレーションが可能となる.

3.2 フラグメント MO 法と通常非経験的分子軌道法 (Hartree-Fock 法) の性能比較

最小基底 STO-3G を用いた Alpha-1 (PDB:3AL1) の結果を通常非経験的分子軌道計算 (Hartree-Fock 法) と比較する. FMO の計算は筑波大の菊池らが開発し, 中野が改良した ABINIT-MP²⁷⁾ を用いた. 非経験的分子軌道計算には, 分子軌道計算プログラムとして広く使われている Pople らの Gaussian98²⁸⁾ を用いた. 計算には通産省工業技術院先端情報計算センター (TACC) の SR8000 を用いた. また, Alpha-1 (PDB:3AL1) の構造は図 6 に示した.

表 3 には, 通常非経験的分子軌道法と FMO の全エネルギーの差を示した. 全エネルギー値の小数点以下 3 桁目が異なっていることがわかる. ただそれは, たかだか 1 Kcal/mol 程度でしかない. 表 4 には HitacSR8000 を用いた計算時間を示した. ABINIT-MP が利用した PU 数も同時に示した. Gaussi-



図 6 ALPHA-1 (PDB:3AL1)

表 3 Difference of Total energy between FMO and HF on ALPHA-1 (PDB:3AL1)

Chain	Number of residues	Total energy /Hartree		Difference / kcal/mol
		FMO-HF/STO-3G	HF/STO-3G	
A	13	-4855.2736778120	-4855.2754720	1.1
B	13	-4855.5681317202	-4855.5703812	1.4
A, B	26	-9711.1845407466	-9711.1884144	2.4
(A, B)-(A)-(B)		-0.3427312144	-0.3425612	-0.1

表 4 Elapsed time and ratio of ALPHA-1 (PDB:3AL1)

Chain	Number of residues	Elapsed time /Sec.		Ratio	Number of processors
		ABINIT-MP	Gaussian98	G98/ABINIT	
A	13	1556.7	757.9	0.5	16
B	13	1497.0	1893.6	1.3	16
A, B	26	1548.8	37017.6	23.9	64

表 5 Result of Estrogen receptor alpha (PDB:3ERD, 3ERT)

PDB	Complex /Hartree	Receptor /Hartree	Ligand /Hartree	ΔE /kcal/mol	Elapsed time (Complex) /second	Number of processors
3ERD	-99947.823	-99115.206	-832.572	-28.2	15283.1	512
3ERT	-100023.114	-98832.313	-1190.762	-24.2	28080.0	256

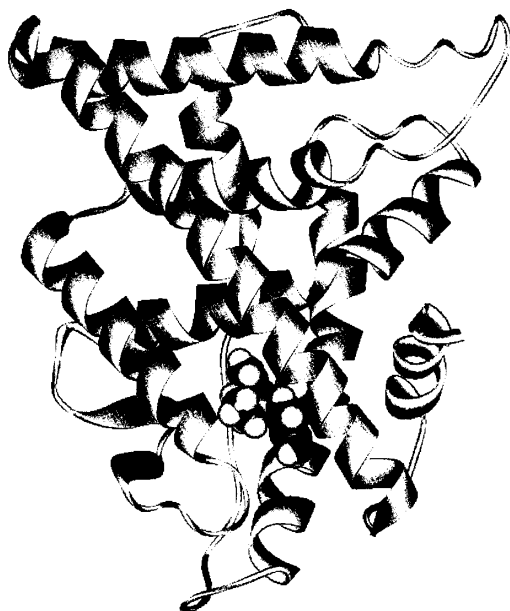


図 7 Estrogen receptor alpha (PDB:3ERD)

an98 には、最高性能が観察された 8PU を用いた。小さな系の場合は FMO は、Gaussian98 とほぼ同等かまたは遅くなるが、大規模系である会合体では、20 倍以上の高速化が達成されている。

日立 SR8000 の 512 プロセッサを用いたエストロゲン受容体 (PDB:3ERD と 3ERT) の FMO-HF/STO-3G 計算結果を表 5 に示した。その構造は図 7 と図 8 に示した。この系は、基底関数が 10000 軌道の系となってしまう、もはや通常の実験的分子軌道計算は実行できない。この計算によりレセプターとリガンドの会合エネルギーが 20 kcal/mol 程度であることが判った。また、非常に構造が似ている 3ERD と 3ERT では 3ERT のほうが 4 kcal/mol 程度会合エネルギーが小さいことが判った。これは、3ERD と 3ERT の性質の違いを議論する上で欠くことのできない情報である。

エストロゲン受容体は、通常の実験的分子軌道計算では基底関数が 10000 軌道の系となってしまう、そのような系の計算が可能な計算機および計算プログラムは、21 世紀を迎えようとする現在でも世界中のどこ



図 8 Estrogen receptor alpha (PDB:3ERT)

にもない。ところが、フラグメント MO 法を用いると高々 512 プロセッサの SR8000 を用いて約 4 時間で計算が可能となる。

フラグメント MO 法と分子軌道計算専用計算機はタンパク質の電子状態の計算とそれを用いた第一原理分子動力学シミュレーションの強力な武器となることが期待できる。タンパク質の電子状態の計算は、様々な酵素反応の反応メカニズムの解析や生体反応を模倣した安全で緩やかな反応設計を実現するために欠くことのできないシミュレーション技術である。

4. 終わりに

タンパク質等の生体分子の性質や生体反応メカニズムの理解には、電子状態の理解が不可欠である。電子状態の計算には、分子軌道法が用いられるが、近似の荒い非経験的分子軌道法でさえもほぼそのサイズの 4 乗に比例する演算量を必要とする。そのため、研究者が研究室レベルで「現実を反映した大規模分子系」の分子軌道計算を実現することは容易なことではない。そこで我々は、「現実を反映した大規模分子系」の分子軌道計算を、「低コスト＝パーソナルユース」で実現するために、計算量を軽減するための近似法を取り入れ、さらに分子軌道計算専用計算機を開発し計算機の性能を飛躍的に向上させることを試みている。本稿では専用計算機とフラグメント MO 法を用いることで、タンパク質の電子状態計算が可能となりつつあることをいくつかの例を用いて報告した。500 残基程度

の小規模なタンパク質の非経験的分子軌道計算は、もう、すぐ手が届くところにある。

計算機シミュレーションを支える計算機の演算速度は毎年約2倍ずつ伸びており²⁹⁾、2010年には Peta FLOPS, 2020年には exa FLOPS の計算機の登場が予想されている。ただ、現在の半導体技術の延長では exa FLOPS が限界であると言われているので、現実の大規模系のシミュレーションを可能とするためには、並列処理技術をふくめた計算性能の向上に加え、計算精度を損なわずに演算量の削減をはかる方法論の開発のための努力が今後とも必要である。

タンパク質のような生体分子の非経験的分子軌道法に基づく分子動力学シミュレーションは、化学物質の薬理活性や毒性の予測に不可欠な生体内反応の理解に大きな役割を果たす基本的シミュレーション技術である。現在、日本では地球シミュレータの開発が進められているが、生体や脳の時代である21世紀には、非経験的分子軌道法や流体力学をハイブリッドした方法論に基づく超並列生体シミュレータの開発が必要となるであろう。その実現はいつのことであるか判らないが、そのプロジェクトに参画する若い世代に正しいバトンを渡せるよう、今後ともさらなる非経験的分子軌道法の改良と普及に努力していくつもりである。

謝 辞

本研究に当たりいつも議論を頂いている北海道教育大学助教授小原繁博士、島根大学助教授網崎孝志博士、九州大学教授村上和彰博士、大阪府立大学教授北浦和夫博士、富士ゼロックス株式会社宮川宣明氏、稲畑深二郎氏、山田想氏、大正製薬 北村一泰博士、国立医薬品食品衛生研究所 神沼二真博士、電子技術総合研究所 秋山 泰博士、生命工学研究所 上林正巳博士、物質工学工業技術研究所主席研究官 田辺和俊博士に感謝する。

本研究の一部は、産業技術情報基盤整備研究開発「ヘテロジニアスネットワークコンピューティング環境における分子シミュレーションシステムの構築」、科学技術振興調整費総合研究「第一原理計算手法に基づくシミュレーション技術に関する研究」、文部省科学研究費試験研究 B「非経験的分子軌道計算の超高速化に基づくハードウェア及びソフトウェアの開発」、科学技術振興調整費総合研究「科学技術計算専用ロジック組込み型プラットフォーム・アーキテクチャに関する研究」、工業技術院先端情報計算センター促進課題「分子シミュレーション手法の開発に関する研究」

に基づくものである。

参 考 文 献

- 1) Dirac, P. A. M.: *Proc. Roy. Soc. (London)*, A123, 714 (1929), 訳: 藤永 茂「分子軌道法」岩波書店 (1988)
- 2) Alsenoy, C. V., Yu, C. H., Peeters, A., Martin, J. M. L., and Schafer, L.: *J. Phys. Chem.*, 102, 2246/2251 (1998)
- 3) Challacombe, M., and Schwegler, E.: *J. Chem. Phys.*, 106, 5526/5536 (1997)
- 4) Mattson, T. G.: *Parallel computing in computational chemistry*, ACS Symp. Series (1995)
- 5) McWeeny, R.: *Method of molecular quantum mechanics*, 2nd edition, Academic (1989)
- 6) Sakuma, T., Kashiwagi, H., Takada, T., and Nakamura, H.: *Int. J. Quantum Chem.*, 61, 137/152 (1997)
- 7) 分子軌道計算関係では、ベストシステムズ社 (<http://www.bestsystems.co.jp>) 等が、分子軌道計算プログラムを並列化し、それを搭載した PC クラスタを製造販売している。
- 8) 村上和彰, 小原 繁, 長嶋雲兵, 網崎孝志, 田辺和俊, 北尾 修, 北村一泰, 高島 一, 宮川宣明, 稲畑深二郎, 山田 想: *CICSJ Bulletin*, 16, 6-12 (1998)
- 9) 小原 繁, 村上和彰, 長嶋雲兵, 網崎孝志, 田辺和俊, 北尾 修, 北村一泰, 高島 一, 宮川宣明, 稲畑深二郎, 山田 想: *CICSJ Bulletin*, 16, 2/5 (1998)
- 10) 稲畑深二郎, 山田 想, 大澤 拓, 沖野晃一, 富田裕人, 橋本浩二, 早川 潔, 宮川宣明, 村上和彰: 電子情報通信学会技術報告, ICD98-21, 77/84 (1998)
- 11) Nagashima, U., Obara, S., Murakami, K., Yoshii, Y., Shirakawa, S., Amisaki, T., Kitamura, K., Takashima, H., Kitao, O., Tanabe, K.: *Comput. Materials Sci.* 14, 132/134 (1999)
- 12) Deware, M. J. S., Zuebitsch, E. G., Healy, E. F., and Stewart, J. J. P.: *J. Am. Chem. Soc.*, 107, 3902 (1985); *MOPAC93.00*, Stewart, J. J. P., Fujitsu Ltd., Tokyo, Japan, 1993. Available from Quantum Chemistry Program Exchange, University of Indiana, Bloomington, IN, USA.
- 13) Kitaura, K., Sawai, T., Asada, T., Nakano, T. and Uebayasi, M.: *Chem. Phys. Letters*, 312, 319/324 (1999)
- 14) Kitaura, K., Ikeo, E., T., Asada, T., Nakano, T. and Uebayasi, M.: *Chem. Phys. Letters*, 313, 701/706 (1999)
- 15) Nakano, T., Kaminuma, T. Sato, T., Akiyama, Y., Uebayasi, M. and Kitaura, K.: *Chem. Phys. Letters*, 318, 614/618 (2000)
- 16) 佐藤智之, 秋山 泰, 中野達也, 上林正巳, 北浦和夫: ハイパフォーマンスコンピューティングシステム, 情報処理学会誌, 41, 104/112 (2000)
- 17) Roothaan, C.C.J.: *Rev. Mod. Phys.*, 23, 69 (1951)
- 18) Szabo, A., and Ostlund, N. S.: *Modern quantum chemistry*, Macmillan (1982), 大野公明, 阪井建男, 望月裕志 訳: 新しい量子化学 (上・下), 東大出版会 (1987)
- 19) Haser, M., and Ahlrichs, R.: *J. Comp. Chem.*, 10, 104/111 (1989)
- 20) 橋本浩二, 山崎雅也, 沖野晃一, 村上和彰: 情報処理学会研究会報告, ARC-125-11 (1997); 山崎雅也, 橋本浩二, 沖野晃一, 村上和彰: 信学技報, ICD97-24 (1997)
- 21) Obara, S., Saika, A.: *J. Chem. Phys.*, 84, 3963/3972 (1986)

- 22) 高島 一, 北村一泰, 田辺和俊, 長嶋雲兵: 日本応用数理学会論文誌, **8**, 135-/151 (1998)
- 23) Takashima, H., Kitamura, K., Tanabe K., Nagashima, U.: *J. Comp. Chem.*, **20**, 443/454 (1998)
- 24) 高島 一, 山田 想, 小原 繁, 北村一泰, 稲畑深二郎, 宮川宣明, 田辺和俊, 長嶋雲兵: *J. Chem. Software*, **6**, 3, 85/104 (2000)
- 25) 高島 一, 山田 想, 小原 繁, 北村一泰, 稲畑深二郎, 宮川宣明, 田辺和俊, 長嶋雲兵: 応用数理学会論文誌, **10**, 3, 241/262 (2000)
- 26) Schmidt, M. W., Baldrige, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. J., Koseki, S., Matsunaga, N., Nguyen, K. A., Su, S., Windus, T. L., Dupuis, M., Montgomery, J. A.: *J. Comput. Chem.*, **14**, 1347/1363 (1993)
- 27) Kikuchi, O., Nakano, T. and Morihashi, K., ABINIT program, Department of Chemistry, University of Tsukuba, Tsukuba (1992): a new versin of ABINIT is now available, Kikuchi, O., Morihashi, K., Nakano, T. and Y. Inadomi, University of Tsukuba, Tsukuba (1999); T. Nakano, ABINIT-MP, National Institute of Health Sciences, Tokyo (2000)
- 28) Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M., Johnson, B. G., Robb, M.A., Cheesman, J. R., Kieth, T., Petersson, G. A., Montgomery, J.A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayara, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzalez, C., Pople, J. A., Gaussian98, Gaissian Inc., Pittsuburgh, PA.
- 29) <http://www.top500.org/>.