

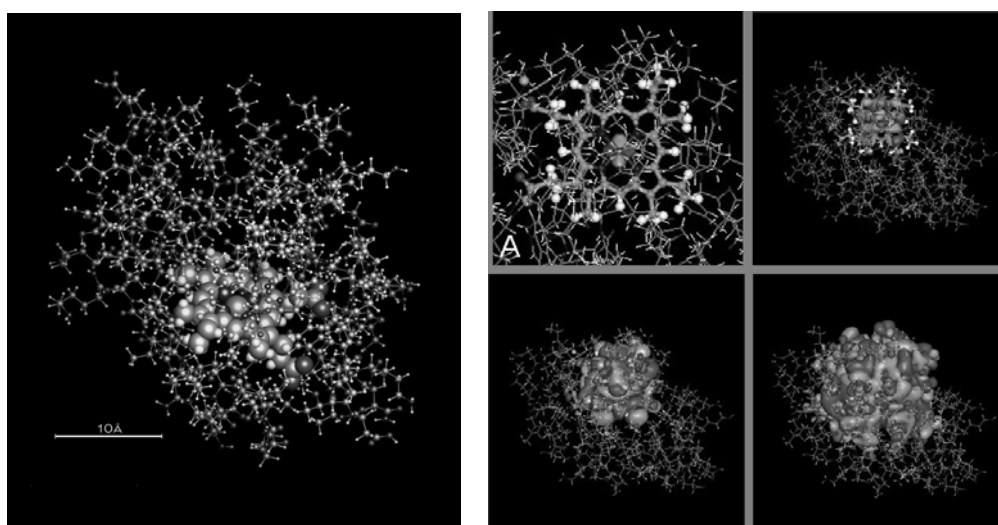
# 密度汎関数法によるタンパク質全電子シミュレーション

佐藤 文俊

東京大学生産技術研究所

## 1. はじめに

タンパク質はミクロな世界で生命活動を支える重要な生体分子である。栄養となるもの、構造を形成するもの、運動に関与するもの、免疫や触媒として働くものなど、実に多様な機能を持つ。図1はシトクロム *c* という約 2000 原子からなるタンパク質の構造である。1つの電子を分子間でやり取りするのが役目であるが、ご覧の通りその構造はたいへん複雑である。



第1図: シトクロム *c* の構造(左)と最高被占分子軌道(右)。

(左)ミトコンドリアの中で電子を運ぶタンパク質。アミノ酸残基104個にヘム分子が結合し、原子数は1738、鉄2価のときの電子数は6586である。下部のスペースフィルモデルで示す分子がヘム。その中央の大きな球が鉄イオンを表し、ここを中心に電子のやり取りを行う。(右)  $3d_x^2-y^2$  軌道を主成分とする分子軌道の等値面表示<sup>1</sup>。等値面の関数値は左上, 右上, 左下, 右下の順に  $\pm 0.05$ ,  $\pm 0.005$ ,  $\pm 0.0005$ ,  $\pm 0.00005$ 。濃淡の差はプラス, マイナスを表す。左上Aはスケールを2倍にして描いている。

電子をやり取りすることができる化学分子ならば様々あるのに、なぜ生体はこのような巨大な分子を使うのだろうか、物理化学的な視点から簡単に考察してみよう。

シトクロム *c* は活性中心にヘム(鉄-ポルフィリン)という分子を持っている。仮にもし1つの電子をやり取りする機能を鉄イオンで行うならば 1000 kcal/mol も必要となってしまう。続いて、鉄-ポルフィリンならば 100 kcal/mol で行えるが、まだまだ要するエネルギーは大きい。しかし、これにペプチド鎖が結合したシトクロム *c* ならば 10 kcal/mol 以下になる。電子の授受の前後で、広範囲にわたり少しずつ電子分布を変えることによって、エネルギー変化を小さく

<sup>1</sup> F. Sato, T. Yoshihiro, M. Era, H. Kashiwagi, Chem. Phys. Lett., 341 (2001) p. 645; *ibid.*, 392 (2004) p. 565.

くしていると考えられる。

反応の方向も重要である。シトクロム *c* が活躍するミトコンドリア内膜では、望ましい順番に反応が効率よく進むよう、また逆反応を抑えるように酸化還元電位が順次異なる物質が配置されている。電子のやり取りを行う物質間で形状や静電場などの相補性を利用している場合もある。酵素タンパク質が行っている触媒機能の戦略も基本的にこれと同じである。

以上のように、なぜ生体はタンパク質を働き手に使うのかは、生理温度という穏やかな環境下のもと、わずかなエネルギーで効率よく、望ましい方向に反応を進めるよう最適な場を与えるためと考えると良さそうだ。したがって、複雑な構造を持っていることこそがタンパク質の多様で最適な機能発現を支えているわけで、これらは密接不可分の関係にある。

自然は実に精巧な仕組みを私たちに授けた。しかし、逆にいえば、タンパク質の反応を解析するためには、タンパク質の全構造を考慮に入れ、kcal/mol の精度を保証する手段が必要ということになる。そのうえ、タンパク質の約 3 分の 1 はアミノ酸が連なってできたペプチド鎖以外に、シトクロム *c* のように活性中心や構造安定化モチーフとして金属イオンも利用している。金属イオンとアミノ酸とを同程度の精度で扱うことができ、電子の移動や、触媒反応といった電子が直接かかわる現象を解析する必要がある。

分子が持つ電子の働きを解析する手段は、量子力学の原理を原子分子に適用する量子化学である。その計算方法は分子軌道法と呼ばれ、分子中の各電子の空間分布(分子軌道)を求めるものである。その形と軌道のエネルギー、各種演算子の期待値の計算から分子の全エネルギーや様々な物理量を見積もり、分子物性を解析することができる。

分子軌道計算には Gaussian<sup>2</sup> や GAMESS<sup>3</sup> といったパッケージが整備されており、多くの化学分子で精度の検証が行われていて、kcal/mol 単位でエネルギーが、数%の誤差で物理量の議論ができるようになった。生体分子もこれらの方法を使用して解析できれば、生体反応を解明するための極めて強力な武器となるであろう。

問題は、分子軌道法の計算量は膨大であるため、タンパク質のように数千~数十万原子からなる分子の計算が大変困難となることである。これに加えて、分子軌道法には収束という計算機の能力とはまた別の問題も解決しなければならない。

これらの困難解決に成功したのが、以下の 3 つの方法である。1 つは、一般化学分子の計算で使用される標準的なカノニカル波動関数計算法を高効率に並列化することで、タンパク質分子を丸ごと計算させる、いわゆる力技である<sup>4,5</sup>。2 つ目は、北浦が開発したエネルギー展開式を利用したフラグメント分子軌道法<sup>6</sup>で、生体分子に適した強力かつ有効な方法である。3 つ目は、量子力学(QM)計算と古典分子力学(MM)計算を組み合わせる QM/MM 法で、古くから生体分子に適用され<sup>7</sup>かつ最も研究が進んでいる。ちなみに、諸熊らの ONIOM 法<sup>8</sup>も QM/MM 法を応用した

---

<sup>2</sup> <http://www.gaussian.com/>

<sup>3</sup> <http://www.msg.chem.iastate.edu/gamess/>

<sup>4</sup> F. Sato Y. Shigemitsu, I. Okazaki, S. Yahiro, M. Fukue, S. Kozuru, H. Kashiwagi, Int. J. Quant. Chem., 63 (1997) p. 245.

<sup>5</sup> K. Tsuda, H. Kaneko, J. Shimada, T. Takada, Comp. Phys. Comm., 142 (2001) p. 140.

<sup>6</sup> K. Kitaura, T. Sawai, T. Asada, T. Nakano, M. Uebayasi, Chem. Phys. Lett., 312 (1999) p. 319.

<sup>7</sup> A. Warshel, M. Levitt, J. Mol. Biol., 103 (1976) p. 227.

<sup>8</sup> M. Svensson, S. Humbel, R. D. J. Froese, T. Mastubara, S. Sieber, K. Morokuma, J. Phys. Chem., 100 (1996) p. 19357.

外挿法である。本拙文は、1 番目の方法の立場から特集号の趣旨に応えようというものである。

生体分子の量子化学計算は、基礎研究のみならず様々な応用も期待されている。ただし、効果的な高速化技法が開発されてはいるものの依然計算量は莫大である。本格的な実用にはペタスケールあるいはその先のエクサスケール規模の計算機が必要となるだろう。

## 2. 密度汎関数法

分子軌道法は、Hartree-Fock 方程式を基礎方程式とした (*ab initio*) Hartree-Fock 法およびその拡張法や Kohn-Sham 方程式を基礎方程式とした密度汎関数法によるカノニカル(正準)分子軌道法が標準法である。本章では密度汎関数法の基礎について解説する。多原子系に広く適用され、種々の分子で計算精度の検証が行われている堅固な理論で、現在、一般的な化学分子向けの主要な分子軌道計算パッケージで採用されている。加えて、タンパク質のような巨大分子の量子化学計算と高い計算精度という、相反する要求を両立させる方法として注目されている。実際の計算アルゴリズムとタンパク質への適用技法についてはそれぞれ3章と4章で説明する。

### 2. 1. 密度汎関数法の概要

対象が水素原子であってもタンパク質であっても、原子分子の持つ電子の研究には量子力学が必要であり、その基礎方程式は Schrödinger 方程式である。本稿では、1900 年に始まった量子力学から密度汎関数法にいたる理論を解説するスペースはないため、簡単な説明を述べたのち、密度汎関数法の基礎方程式、計算に用いる方程式を登場させる。これらの詳細に興味がある方は、文末の参考文献に挙げた名著を参照願いたい(筆者らの拙書はご愛敬)。

Schrödinger 方程式は物質などの状態の全波動関数を求める方程式である。原子分子の全電子波動関数を求める式では、電磁波などが外から加えられていない場合、座標と時間の変数分離が達成され、全波動関数が時間変化する式と時間に依存しない方程式が得られる。本稿では主に時間に依存しない方程式を対象に説明する。

続いて、原子分子の物理化学では、電子の質量は1番軽い水素原子核の質量に比べても1800倍軽いという事実を利用して、電子状態は原子核の動きに直ちに追従するという近似が用いられる。これは Born-Oppenheimer 近似と呼ばれる大変重要な仮定である。オングストローム ( $\text{Å} = 0.1 \text{ nm}$ ) 領域の分子を扱う理論研究において、この近似の恩恵は絶大である。分子の電子状態を研究する際は、核の運動を無視し、核座標を固定して電子の振る舞いを計算する。核の運動を議論する際は、核の運動に比べて電子は速やかに平衡分布に達するので、核の運動のみを取り扱う。Born-Oppenheimer 近似はこのような研究方法の基盤となっている。

しかし、Schrödinger 方程式は1電子や非常に特殊な2電子系以外は解析的に解くことができない。そのため、Hartree-Fock 法をはじめとして様々な方法が考案・研究されてきた。密度汎関数法はそれらの方法の1つで、①基底状態のエネルギーが電子密度の汎関数として記述できること、②電子密度を試行関数とする変分原理を満足すること、から導かれた Schrödinger 方程式と同等の解法である。なお、これらの原理は、時間に依存しない Schrödinger 方程式に適用させたのが Hohenberg-Kohn の定理<sup>9</sup>、のちに時間に依存する方程式に適用させたのが Runge-Gross の定理<sup>10</sup>と呼ばれている。

<sup>9</sup> P. Hohenberg, W. Kohn, Phys. Rev., B, 136 (1964) p. 864.

<sup>10</sup> E. Runge, E. K. U. Gross, Phys. Rev. Lett., 52 (1984) p. 997.

密度汎関数法はもう 1 つの標準法である Hartree-Fock 法に対する近似や拡張ではなく、これとは独立した体系である。Hartree-Fock 法の基礎方程式である Hartree-Fock 方程式は、波動関数の汎関数で表された全エネルギーを軌道の規格直交条件のもと、変分することで導かれる。一方、密度汎関数法の骨子は、全エネルギー  $E$  が電子密度  $\rho(\mathbf{r})$  の変分可能な汎関数  $E[\rho]$  として与えられるとき、次の変分

$$\delta\left\{E[\rho] - \mu\left(\int \rho(\mathbf{r})d\mathbf{r} - N_{elec}\right)\right\} = 0 \quad (2.1)$$

を求めることである。ここで、 $\mathbf{r}$  は電子の座標である。 $\mu$  は電子密度の全空間による積分が全電子数  $N_{elec}$  となる拘束条件を付加するためのラグランジェの未定乗数であり、(2.1)式から

$$\mu = \frac{\delta E[\rho]}{\delta \rho} \quad (2.2)$$

と表される化学ポテンシャルとなる。そのため、まず全エネルギーを電子密度  $\rho$  の汎関数で表す必要があるが、これを保障するのが Hohenberg-Kohn の定理である。

ところで、波動関数  $\Psi$  を汎関数とした変分原理で導かれた Hartree-Fock 系の方法では、分子が作る外場ポテンシャル  $v(\mathbf{r})$  が与えられると、その解として  $\Psi$  が一意的に決まる。それから電子密度  $\rho(\mathbf{r})$  が一意的に決定される。すなわち、 $v \rightarrow \Psi \rightarrow \rho$  の関係にある。ここで、縮退がなければ、 $\Psi \rightarrow v$  も成り立つ。一方、密度汎関数の定理は  $v(\mathbf{r})$  が  $\rho(\mathbf{r})$  の 1 価関数であることを保障する。これら全ての結果から、電子密度  $\rho(\mathbf{r})$  と波動関数  $\Psi$  は  $v(\mathbf{r})$  を通して 1 対 1 に対応することになる。

$$\rho \leftrightarrow v \leftrightarrow \Psi \quad (2.3)$$

これは真の電子密度で成り立つ。電子密度は波動関数同様、一義的汎関数である。ということは、例えば Hartree-Fock 法では全ての軌道を求める必要があるため、軌道数本分の方程式を連立して解かなければならないが、密度汎関数原理によれば、電子密度による正確な全エネルギーの表現が得られれば、電子密度に関する(2.2)式のオイラー方程式 1 本で済むことになる。

電子密度もエネルギーも観測量である。そのため、電子密度とエネルギーの対応表が十分にそろえば、十分精密なエネルギー汎関数を構築できる可能性がある。初期の密度汎関数法は、 $E[\rho]$  の記述に明示的な近似汎関数を導入し、著しく簡略化された方程式の導出に関心が持たれた。しかしながら、分子の化学的性質の記述に耐えられるような普遍的なエネルギー汎関数は、今のところ発見されていない。代わりに、その精神は次に述べる Kohn-Sham 法の枠組み中で、実用的な交換相関汎関数を得る問題へと引き継がれ、現在も精力的に研究が行われている。

## Kohn-Sham 方程式

Kohn と Sham は汎関数の記述が特にネックとなった運動エネルギー汎関数の表現を放棄し、Hartree-Fock 法と同様の手法を採用した。具体的には、単一のスレーター行列式による波動関数を導入して、多電子問題を有効 1 電子問題に帰着させた<sup>11</sup>。この試みは見事に成功し、Hartree-Fock 法と同様に軌道数本分の方程式を連立して解かなければならない代わりに、密度汎関数法の精度は驚異的に向上し、多原子分子を解くための現実的な方法となった。これを Kohn-Sham 法と呼ぶ。現在では、密度汎関数法といえば Kohn-Sham 法のことを指し示す言葉と

<sup>11</sup> W. Kohn, L. J. Sham, Phys. Rev., A140 (1965) p.1133.

なっている。

それでは、早速 Kohn-Sham 法の支配方程式を見てみよう。本稿では、 $\hbar, e^2, m$  の物理定数値をそれぞれ 1 とする Hartree 原子単位系 au を用いる。この単位系を用いると式がすっきりするため、分子軌道計算ではこの単位を使うことが多い。ただし、実際の計算においては以下のことを念頭に置く必要がある。1 au の距離はボーア半径となり約 0.529 Å である。また、1 au のエネルギーは 27.21 eV, 627.5 kcal/mol であり、タンパク質の反応エネルギーに比べて非常に大きい。分子軌道計算の結果解析の際には十分に注意すべきである。

エネルギー汎関数  $E[\rho]$  は運動エネルギー  $T[\rho]$  と核電子間のクーロン引力  $V_{ne}[\rho]$ 、電子間クーロン相互作用  $V_{ee}[\rho]$  の和である。ただし、

$$V_{ne}[\rho] = - \sum_A^{N_{atom}} \int \frac{\rho(\mathbf{r}) Z_A}{|\mathbf{r} - \mathbf{R}_A|} d\mathbf{r} \quad (2.4)$$

である。 $Z_A$ 、 $\mathbf{R}_A$  はそれぞれ原子核  $A$  の電荷、座標である。ここで、単一のスレーター行列式が与える運動エネルギー  $T_s[\rho]$ 、古典的な電子間クーロン相互作用を  $J[\rho]$  とし、

$$E_{XC}[\rho] \equiv (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) \quad (2.5)$$

真の運動エネルギーと近似運動エネルギーとの差、および真の電子間クーロン反発エネルギーと古典の電子間クーロンエネルギーとの差の和を定義する。すなわち、それぞれの項の不明分を押し込めたものである。特に、電子間クーロン反発エネルギーの差が与える物理的意味から、

$E_{XC}[\rho]$  は便宜上、交換相関エネルギーと呼ばれている。

(2.5) 式を用いると、(2.2) 式のオイラー方程式は以下と書ける。

$$\mu = v_{eff}(\mathbf{r}) + \frac{\delta T_s[\rho]}{\delta \rho} \quad (2.6a)$$

$$v_{eff}(\mathbf{r}) = - \sum_A^{N_{atom}} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mu_{XC}(\mathbf{r}) \quad (2.6b)$$

ここで、 $\mu_{XC}(\mathbf{r})$  は交換相関ポテンシャルと呼ばれ、 $E_{XC}[\rho]$  が決まれば

$$\mu_{XC}(\mathbf{r}) = \frac{\delta E_{XC}[\rho]}{\delta \rho} \quad (2.7)$$

より直ちに求まる。

単一のスレーター行列式は相互作用のない系における全波動関数を与える。そのため、単一のスレーター行列式の軌道  $\phi_i(\mathbf{r})$  で定義した電子密度

$$\rho(\mathbf{r}) = \sum_{i=1}^{N_{elec}} |\phi_i(\mathbf{r})|^2 \quad (2.8)$$

が厳密な電子密度と等しいとすれば、(2.6a) 式は、(2.6b) 式で表されたポテンシャル  $v_{eff}(\mathbf{r})$  のもとで、電子どうしが互いに相互作用無く運動する ( $\delta T_s[\rho]/\delta \rho$ )、仮想的な電子系の電子密度

を求める方程式に相当する。このように Kohn と Sham は、真のポテンシャルのもとで相互作用する電子系の問題を、(2.6b)式の $v_{eff}(\mathbf{r})$ のもとで電子間の相互作用のない系を解く形式にすり替えたのである。

Kohn-Sham 方程式は、スレーター行列式の軌道の規格直交条件のもとでエルミート性を利用して、正準 Hartree-Fock 方程式と似た 1 電子方程式として与えられる。

$$\left[ -\frac{1}{2}\Delta - \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mu_{XC} \right] \phi_i(\mathbf{r}) = \varepsilon_i \phi_i(\mathbf{r}) \quad (2.9a)$$

これを制限付き閉殻 Kohn-Sham 方程式と呼ぶ。 $\alpha, \beta$  スピンを分離する場合は以下となる。

$$\left[ -\frac{1}{2}\Delta - \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d\mathbf{r}' + \mu_{XC}^\sigma \right] \phi_i^\sigma(\mathbf{r}) = \varepsilon_i^\sigma \phi_i^\sigma(\mathbf{r}) \quad (2.9b)$$

これを非制限 Kohn-Sham 方程式と呼ぶ。ここで、 $\Delta$  はラプラス演算子、 $\sigma = \alpha, \beta$  である。(2.9)式の左辺の括弧内は Kohn-Sham 演算子と呼ばれ、右辺の $\varepsilon_i$ は Kohn-Sham 軌道エネルギーである。Kohn-Sham 法に基づく分子軌道法は、これら正準 Kohn-Sham 方程式を解く方法である。

## 交換相関エネルギー

Kohn-Sham 方程式と Hartree-Fock 方程式は大変よく似ており、実際(2.9)式において交換相関ポテンシャル演算子 $\mu_{XC}$ を Fock 交換演算子に置き換えたものが Hartree-Fock 方程式である。

(2.7)式で $\mu_{XC}$ を決める $E_{XC}[\rho]$ は(2.5)式で定義されるように、真の運動エネルギーとの差も繰り込まれているものの、本質的には電子間クーロン相互作用における量子効果が生むエネルギーの総和である。その1つは Pauli の排他原理から要請される電子の交換エネルギーである。これとは別のエネルギーの利得がある。電子はマイナスの電荷を持っているため、電子が互いに接近するとエネルギーが高くなる。そこで平均分布は同じでも互いに避けあう方が安定になる。これを電子相関効果と呼ぶ。Hartree-Fock 法では正確な交換エネルギーを与えるが、異なるスピンを持った電子間のクーロン反発が考慮されていない。したがって、この電子相関効果と運動エネルギー補正を正確に見積もることができれば、Kohn-Sham 方程式は基底状態のシュレディンガー方程式と等価になる。もっとも、先に述べたとおり、 $E_{XC}[\rho]$ の普遍的な汎関数はわかっていない。これが Kohn-Sham 法の最大の特徴でもあり同時に欠点でもある。原理的には、 $E_{XC}[\rho]$ の記述が適切であるほど Kohn-Sham 方程式の精度が向上する。

Hartree-Fock 法の枠組みで不足している電子相関を取り込む場合には、複数のスレーター行列式を用いた系で計算するか、摂動展開でその効果を取り込む。これらの方法では計算量が飛躍的に増えてしまううえに、相関効果の取り込まれ方も非常にゆっくりとしている。Kohn-Sham 法の枠組みでは、 $E_{XC}[\rho]$ の記述の中に相関エネルギーを取り込むことができるため、本質的な

計算量を格段に増やさずに済む可能性が高い。そのため、現在も精力的に新しい交換相関汎関数が提案されている。実際に、Hartree-Fock 法の拡張法よりも遥かに少ない計算時間で、それらと対等またはそれ以上の計算精度を得るまでになっている。今や Kohn-Sham 法は一般化学分子の標準的な分子軌道計算法として認められているが、タンパク質のような巨大で豊富な電子を持つ系においてこそ、まさにその威力が発揮される計算方法である。

## 2. 2. Kohn-Sham-Roothaan 方程式

Kohn-Sham 方程式の多原子分子の解法には、Hartree-Fock 方程式の解法で開発された Roothaan 法<sup>12</sup>が適用される。分子軌道を基底関数の線形結合で表現するというもので、 $i$  番目の分子軌道  $\phi_i(\mathbf{r})$  を次のように既知の関数系 ( $\mathbf{g}_p(\mathbf{r}): p = 1, 2, \dots, N_{bf}$ ) によって展開する。

$$\phi_i(\mathbf{r}) = \sum_p^{N_{bf}} C_{pi} \mathbf{g}_p(\mathbf{r}) \quad (2.10)$$

ここで、 $\mathbf{g}_p(\mathbf{r})$  を基底関数と呼ぶ。基底関数はスレーター型やガウス型の関数、さらには平面波関数を用いるものもある。いずれにせよ、実際の計算では、(2.10) 式の展開は有限である。

(2.10) 式の分子軌道  $\phi_i(\mathbf{r})$  の展開式では基底関数は明確に定義されているため、微積分方程式 (2.9) 式を解く問題は未知である展開係数  $C_{pi}$  と  $\varepsilon_i$  を決める問題になる。これが Roothaan 法最大の特徴である。ちなみに展開係数は当初の発想から LCAO 係数 (Linear Combination of Atomic Orbitals; 原子軌道の線形結合係数) と呼ばれている。

Kohn-Sham-Roothaan 方程式の導出は、軌道を (2.10) 式で展開し、(2.9) 式の左から  $\langle \mathbf{g}_p(\mathbf{r}) |$  を演算する。閉殻系の (2.9a) 式の行列表現は

$$\mathbf{FC} = \mathbf{SC}\boldsymbol{\varepsilon} \quad (2.11)$$

である。 $\mathbf{C}, \boldsymbol{\varepsilon}$  は  $\mathbf{F}$  を対角化したときに求まる固有ベクトルの行列と固有値であり、 $\boldsymbol{\varepsilon}$  は対角行列である。 $\mathbf{S}$  は重なり積分の行列である。上の式を行列要素で表すと次のようになる。

$$\sum_q^{N_{bf}} F_{pq} C_{qi} = \varepsilon_i \sum_q^{N_{bf}} S_{pq} C_{qi} \quad (2.12a)$$

$$F_{pq} = h_{pq} + \sum_{rs} P_{rs} \langle pq | rs \rangle + \mu_{pq}^{XC} \quad (2.12b)$$

ただし、

$$h_{pq} = \int \mathbf{g}_p^*(\mathbf{r}) \left[ -\frac{1}{2} \Delta - \sum_A^{N_{atom}} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} \right] \mathbf{g}_q(\mathbf{r}) d\mathbf{r} \quad (2.13)$$

<sup>12</sup> C. C. J. Roothaan, Rev. Mod. Phys., 23 (1951) p. 69.

$$\langle pq|rs\rangle \equiv \iint g_p^*(\mathbf{r})g_q(\mathbf{r})\frac{1}{|\mathbf{r}-\mathbf{r}'|}g_r^*(\mathbf{r}')g_s(\mathbf{r}')d\mathbf{r}d\mathbf{r}' \quad (2.14)$$

$$P_{rs} = 2\sum_j C_{rj}^*C_{sj} \quad (2.15)$$

$$S_{pq} = \int g_p^*(\mathbf{r})g_q(\mathbf{r})d\mathbf{r} \quad (2.16)$$

$$\mu_{pq}^{XC} = \int g_p^*(\mathbf{r})\mu_{XC}(\mathbf{r})g_q(\mathbf{r})d\mathbf{r} \quad (2.17)$$

である。 $\mathbf{F}$ はKohn-Sham行列と呼ばれる。 $\mathbf{h}$ ,  $\langle pq|rs\rangle$ ,  $\mathbf{S}$ および $\mathbf{P}$ は, それぞれ1電子積分, 2電子積分, 重なり積分および密度行列と呼ばれ, Hartree-Fock-Roothaan方程式と共通である。Kohn-Sham-Roothaan方程式は, (2.17)式の交換相関積分演算以外は, Hartree-Fock-Roothaan方程式と同じ分子積分演算と一般行列演算からなる。解法そのものもHartree-Fock-Roothaan方程式と共通するものが多い(3.1節参照)。逆に言えば, 化学の世界で培われてきたHartree-Fock-Roothaan方程式を解くための様々な技術が応用できる。

Kohn-Sham-Roothaan方程式の最大の特徴は(2.17)式の $\mu^{XC}$ である。Hartree-Fock-Roothaan方程式では, これに対応する項は(2.14)式と同じ形の分子積分を用いて顕に書き下すことができる。現在提案されている $\mu_{XC}(\mathbf{r})$ は一般に有理関数であるため,  $\mu^{XC}$ の行列要素の計算には分子積分演算とは異なる数値積分計算を用いることになる(3.2節参照)。

このように, 交換相関項の計算に特殊な技術や時間が必要となるものの, Kohn-Sham-Roothaan法はHartree-Fock-Roothaan法と本質的に同程度の計算量で飛躍的に計算精度が増すことができる。これこそが, 小分子のみならず巨大分子であるタンパク質の量子化学計算にとって大変魅力的である所以である。

### 3. Kohn-Sham-Roothaan法の解法概要

この章では, 2章で示したKohn-Sham-Roothaan方程式を解く基本アルゴリズムと, 巨大なタンパク質の計算に欠かせない高速化と大量データの扱いに関する基礎技術を概説する。なお, これらの方法は1章で紹介したフラグメント分子軌道法やQM/MM法にも有効で, 大きなフラグメントやQM領域を計算する方法に適用できる。

#### 3. 1. Kohn-Sham-Roothaan法のアルゴリズム

まず本節では, タンパク質の座標が固定され, スピンの異なる電子が2個ずつ分子軌道に入る閉殻系の基底状態の計算を例に, もっとも基本的なKohn-Sham-Roothaan法の自己無撞着(SCF)計算アルゴリズムを説明する。一部の計算を除けばHartree-Fock-Roothaan法と共通である。また, 閉殻系の計算は $\alpha$ スピンと $\beta$ スピンでそれぞれ行えばよい。

解くべき方程式は(2.11)式である。基底関数 $g_p(\mathbf{r})$ は普通規格化されているが, 直交はしていないため, (2.16)式の $\mathbf{S}$ は単位行列ではない。そこで,  $\mathbf{S}$ を対角化するユニタリ行列 $\mathbf{U}$ を



使って、直交変換行列  $\mathbf{X}$  を作る。

$$\mathbf{X} = \mathbf{U}\mathbf{s}^{-\frac{1}{2}} \quad (3.1a)$$

$$\mathbf{U}^\dagger \mathbf{S}\mathbf{U} = \mathbf{s} \quad (3.1b)$$

ここで、 $\mathbf{s}$  は  $\mathbf{S}$  の固有値を対角要素に持つ対角行列である。 $-1/2$  乗は対角要素の平方根の逆数をとることを意味する。この  $\mathbf{X}$  を用いて (2.11) 式を変形する。

$$\mathbf{F}'\mathbf{C}' = \mathbf{C}'\boldsymbol{\varepsilon} \quad (3.2a)$$

$$\mathbf{F}' = \mathbf{X}^\dagger \mathbf{F}\mathbf{X} \quad (3.2b)$$

$$\mathbf{C}' = \mathbf{X}^{-1}\mathbf{C} \quad (3.2c)$$

これで (2.11) 式は通常の固有値問題 (3.2a) 式になった。得られた  $\mathbf{C}'$  から  $\mathbf{C}$  を逆変換  $\mathbf{X}\mathbf{C}'$  で計算し、(2.15) 式で  $\mathbf{P}$  を計算しなおして全エネルギーが求まる。

$$E = \sum_{pq} P_{pq} h_{pq} + \frac{1}{2} \sum_{pqrs} P_{pq} P_{rs} \langle pq | rs \rangle + E_{XC} + \sum_{A=1}^{N_{atom}} \sum_{B>A}^{N_{atom}} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|} \quad (3.3)$$

SCF 計算では、適当な波動関数や電子密度の初期値から計算を始め、(2.12) から (3.3) 式を繰り返し計算し、全エネルギーや各行列要素などの差の絶対値が望みの閾値以下になったときに計算を終了する。これを収束と呼ぶ。

化学分子の Roothaan 法で使用する基底関数は、いくつかのガウス型関数の和で表される実数型の縮約ガウス型関数が一般的である。この場合、これまでの式で複素共役を意味する\* (アスタリスク) は取ってかまわない。

Kohn-Sham-Roothaan 方程式の計算手順を、図 2 にしたがって説明する。なお、Hartree-Fock-Roothaan 法と異なるところに図では網掛けを、本文中では○印をつけている。

1) 分子構造、基底関数、計算条件を読み込む。タンパク質の分子軌道の計算を行うにあたり、タンパク質分子の座標や基底関数の組、電子の配置や計算の手法、用いる交換相関汎関数といった情報が必要であり、これを初めに読み込む。  
続いて、SCF 計算のための準備計算を行う。SCF の繰り返し計算を通して変化しない物理量をあらかじめ計算しておく。

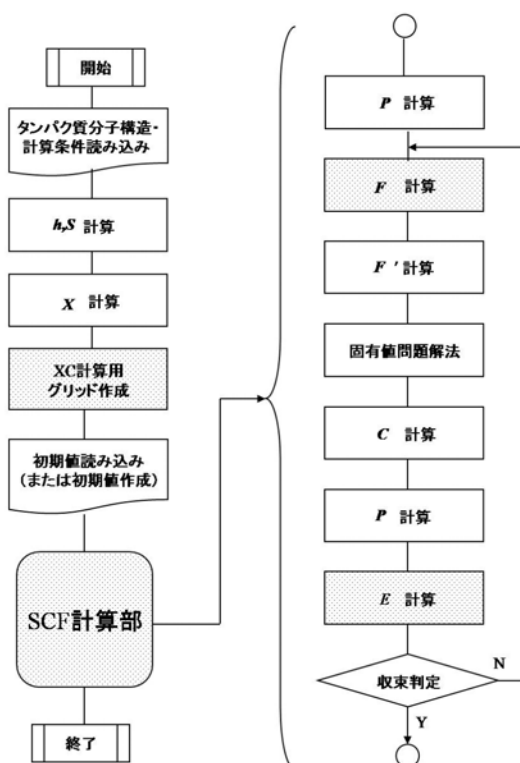
- 2) (2.13) 式の 1 電子積分  $\mathbf{h}$  と (2.16) 式の重なり積分  $\mathbf{S}$  を計算する。
- 3) (3.1a) 式で直交規格化条件を満たすよう基底関数を変換する行列  $\mathbf{X}$  を計算する。この際、基底の線形従属性を取り除くという作業も行っておく。
- 4) (2.17) 式の交換相関項の積分に用いるグリッドとその重みを生成する(○)。
- 5) SCF 計算の初期値を作成する。別に作成したものを読み込んでもよい。LCAO 係数  $\mathbf{C}$ 、密度行列  $\mathbf{P}$  などを与えることができる。初期値作成法に関しては 4 章を参照。

以降が、SCF 計算における繰り返し計算部分である。様々な計算ルーチンが逐次呼び出され計算が実行される。収束条件が満たされるまで計算が繰り返される。

- 6) (2.15) 式から電子密度行列  $\mathbf{P}$  を計算する。電子密度行列が初期値として与えられている場合はスキップする。
- 7) 2 電子積分を計算しながら、(2.12) 式の Kohn-Sham 行列  $\mathbf{F}$  を計算する(○)。
- 8) (3.2b) 式により Kohn-Sham 行列を規格直交基底  $\mathbf{F}'$  に変換する。
- 9) (3.2a) 式の固有値問題を解く。固有ベクトルとして直交基底の Kohn-Sham LCAO 係数  $\mathbf{C}'$ 、

固有値として Kohn-Sham 軌道エネルギー  $\epsilon$  を得る。

- 10)  $C = XC'$  により, LCAO 行列を本来の非直交基底に変換する。
- 11) (2.15) 式で新しい電子密度行列を計算する。
- 12) (3.3) 式の全エネルギー (○) や各原子の電荷などの物理量を計算する。
- 13) SCF 計算の収束条件を満たしたかどうかを判定する。1 回前の SCF で計算された密度行列の全要素と全エネルギーを今回のものと比較して, その差の絶対値が閾値以下になった場合を収束とみなす。収束したら繰り返し計算を終了する。収束条件を満たさない場合は, 4 節の収束法を施して, 7) に戻る。



第 2 図: Kohn-Sham-Roothaan 方程式の計算手順[1]。

Hartree-Fock-Roothaan 法と異なるところに網掛けしている。

以上が Kohn-Sham-Roothaan 方程式の SCF 計算手順である。○印をつけた Hartree-Fock-Roothaan 法との差異はわずかで, 化学の世界で蓄積してきた分子軌道法の技術が利用しやすい形になっている。

なお, SCF 計算は必ず収束するとは限らない。発散したり, 振動したり, また収束しない場合もあるので, 実際の計算には注意が必要である。これはタンパク質の SCF 計算では特に深刻な問題である。この問題への対策は 4 章で説明する。

### 3. 2. 律速計算のアルゴリズム概要

続いて, 上記の各ルーチンのうち, 律速となる計算のアルゴリズムを概説しておこう。これらは多岐にわたるうえ, 計算機の進歩とともに次々と新しい方法が開発されている。現在もホットなテーマである。本稿ではそれぞれの標準法を簡単に紹介する。

## 分子積分

分子積分は Roothaan 分子軌道法において最も計算律速となる計算ステップである。分子積分計算の問題点は大きく分けて 3 つある。

第 1 に計算自体が難しいことである。原子軌道型の基底関数には軌道のすその振る舞いからスレーター型の関数 ( $\exp(-\alpha|\mathbf{r}-\mathbf{R}|)$ ) が適しているが、この関数を用いると分子積分の計算に多くの行程が必要になる。そこで、Boys は基底関数にガウス型関数 ( $\exp(-\alpha|\mathbf{r}-\mathbf{R}|^2)$ ) を使用することを提案した。軌道の良好な表現を得るためには、スレーター型の関数に比べて多くの数の関数を使用しなければならないが、分子積分をスレーター型に比べ少ない行程で解析的に計算することができる。現在では、ほとんどの化学分子用分子軌道法プログラムはガウス型基底関数を採用している。

第 2 にデータ量が膨大になることである。中でも、4 つのインデックスを持つ 2 電子積分  $\langle pq|rs\rangle$  のサイズは倍精度浮動小数点型で  $8(\text{Byte}) \times N_{bf}^4 / 8 = N_{bf}^4 (\text{Byte})$  になる。ここで、1/8 は分子積分の対称性を考慮した比率である。基底関数が 1 万で  $10^{16}$  Byte, 10 万では  $10^{20}$  Byte になる。保持するには大変な量である。

第 3 に同様に計算量も膨大になることである。2 電子積分の計算量は  $N_{bf}^4$  に比例する。個々の分子積分を高速に計算することができても、分子サイズが大きくなるとすぐに計算できる限界に達してしまう。分子サイズ依存性を低下させるような計算量の削減方法が必要になる。

このような問題点の解決に用いられる現在の標準法を以下に簡単に説明しよう。まず第 1 の問題には、ガウス型関数の分子積分では、漸化式による高速計算法が用いられている。ガウス型関数のもつ基本的な性質として、ガウス型関数どうしの積はやはりガウス型関数になるので非積分関数が簡単なものになる。また、ガウス型関数自体の積分も容易である。さらに、(2. 14) 式で現れるクーロン演算子  $1/|\mathbf{r}-\mathbf{r}'|$  はラプラス変換により、やはりガウス型関数の積分として取り扱うことができる。Boys の提案が受け入れられた背景にこれらの都合な性質がある。実際に使用されている縮約ガウス型関数はガウス型関数の線形結合であるから、分子積分の表現も単に展開項数が増えただけのものにすぎない。

それに加えて、少ない行程で分子積分を計算する実に巧妙な方法を編み出したのが小原と雑賀<sup>13</sup>である。ガウス型基底関数の中心座標に関して微分を計算すると、例えば

$$\frac{\partial}{\partial R_x} \exp(-\alpha(\mathbf{r}-\mathbf{R})^2) = -2\alpha(x-R_x)\exp(-\alpha(\mathbf{r}-\mathbf{R})^2) \quad (3.4a)$$

すなわち

$$\frac{\partial}{\partial R_x} g^s(\mathbf{r}, \mathbf{R}) = -2\alpha g^{p_x}(\mathbf{r}, \mathbf{R}) \quad (3.4b)$$

となる。ここで、 $g^s$  は  $s$  軌道、 $g^{p_x}$  は  $p_x$  軌道である。このように軌道角運動量の量子数が 1 つ大きい軌道が得られる。この特徴から  $s$  軌道の分子積分計算値から出発して、順次  $p$ ,  $d$  軌道

<sup>13</sup> S. Obara, A. Saika, J. Chem. Phys., 84 (1986) p. 3963.

の分子積分を計算してゆく漸化式が導かれた。最も複雑な 2 電子積分の漸化式でも  $d$  軌道までなら高々 8 項程度の和で表現され、演算量が少ないことが最大の利点である。

続いて、第 2 の問題、すなわち分子積分のデータ量とそれに付随する問題について検討する。従来は 2 電子積分の計算に時間がかかったため、SCF 計算の前に 1 度だけ計算してディスクなどの外部記憶に保存し、必要なときにファイルから読み出していた。これを従来法あるいはディスク法と呼ぶ。この方法では Kohn-Sham 行列要素などの計算はファイルアクセスの時間に支配され、計算できる分子サイズがディスクの容量で決められてしまう。タンパク質の計算では全ての値を保存するには膨大なディスク容量が必要になる。

そのため、ディスクに積分値を保持しない方法が考案された。分子積分を SCF 計算で必要になる度に毎回計算して、分子積分を使い捨てにする方法<sup>14</sup>である。これをダイレクト法と呼ぶ。この方法が受け入れられた背景として、小原と雑賀法のような効率のよい分子積分計算法が提案されたこと、次に述べるカットオフ法が開発されたことが挙げられる。様々な計算機メーカーが競って高速な CPU を開発していることも見逃せないだろう。現在ダイレクト法は分子軌道計算の標準になっている。しかし、今後、従来法の考え方に再帰する可能性もある。これについては本小節の最後に述べよう。

ダイレクト法では 2 電子積分が必要なときに計算され使用される。クーロン項の形は

$$U_{pq} = \sum_{rs} P_{rs} \langle pq|rs \rangle \quad (3.5)$$

である。この  $U$  であれば、 $\langle pq|rs \rangle$  の 2 電子積分を  $r,s$  あるいは  $p,q$  のインデックスで呼び出しながら、主記憶に保持した電子密度行列  $\mathbf{P}$  との積和計算で求めることができる。これらのインデックスが走る領域を固定長にして、積和が終了したら次の未計算のインデックス群に移行し、インデックスの最後に到達するまで同様の操作を行う。このようなアルゴリズムを用いれば、分子積分に要するメモリを少なく抑えることができる。この方法により、計算できる分子のサイズがメモリやディスクの容量によって大幅に限定されることがなくなった。

ダイレクト法には、SCF 計算の収束過程を利用した高速化法もある。(3.5) 式の行列要素の計算は以下のように書き換えることができる。

$$U_{pq}^{(\lambda)} = U_{pq}^{(\lambda-1)} + \sum_{rs} (P_{rs}^{(\lambda)} - P_{rs}^{(\lambda-1)}) \langle pq|rs \rangle \quad (3.6)$$

ここでインデックス  $(\lambda)$  は SCF 計算の繰り返し回数である。これらの式では、前回の SCF 計算で求められた行列要素に電子密度の差分による補正を加えている。SCF 計算が収束に近づくとつれて  $(\mathbf{P}^{(\lambda)} - \mathbf{P}^{(\lambda-1)})$  の要素の絶対値は小さくなるので、 $\langle pq|rs \rangle$  の値が小さいものは計算を省略することができる。これをアップデート法と呼ぶ。この方法は次のカットオフ法と密接な関係があるため、詳細はあとで述べるが、SCF 計算が収束に近づくとつれて計算時間が短縮する。

最後に、分子積分計算量の分子サイズ依存性を劇的に減らすカットオフ法について説明する。この方法とダイレクト法およびアップデート法を組み合わせることにより、分子積分が関与する計算を大幅に高速化することができる。

分子が大きくなればなるほど、2 電子積分の大半はゼロに近い値になる。基底関数  $p$  と  $q$  が

<sup>14</sup> J. Almlöf, K. Faegri, Jr., K. Korsell, J. Comp. Chem., 3 (1982) p.385.

離れていれば  $p$  と  $q$  の重なりは無視できるほど小さくなるからである。カットオフ法とは、分子積分の上限値を事前に簡単な式で見積もり、これが SCF 計算の収束閾値よりも十分に小さい場合は、その分子積分の計算を省略する方法である。2 電子積分は Schwarz の不等式から

$$\langle pq|rs\rangle \leq \sqrt{\langle pq|pq\rangle} \sqrt{\langle rs|rs\rangle} \quad (3.7)$$

である。したがって、 $\langle pq|pq\rangle$  あるいは  $\langle rs|rs\rangle$  の値を計算すれば、2 電子積分の上限値を求めることができる。これらは 2 つのインデックスによる計算である。この篩が理想的に働けば計算すべき分子積分の項数を  $O(N_{bf}^2)$  にまで減らすことができる。

カットオフ法をダイレクト法と組み合わせるとさらに強力になる。むしろカットオフ法のおかげでダイレクト法が有用となったといってもよい。カットオフ法の上限値の見積もりは (3.5), (3.6) 式の積和の前に行うことができる。これにより、分子積分が関与する  $U$  の計算自体が  $O(N_{bf}^2)$  に抑えられる。この操作をスクリーニングと呼ぶこともある。

このとき、 $\max|P^{(\lambda)} - P^{(\lambda-1)}| \langle rs|rs\rangle$  をカットオフの判断に使用しても、精度の問題が生じないことがわかる。 $\max|\cdot|$  は行列要素の絶対値の最大値を意味する。すなわち、アップデート法も併用することで、カットオフの効果を高めることができる。SCF 計算の収束値近傍では  $\max|P^{(\lambda)} - P^{(\lambda-1)}|$  の値は高々 SCF 収束判定の閾値になるため、 $\langle rs|rs\rangle$  のみをカットオフの判断に使うよりもさらに 3~5 桁大きい分子積分の上限値でカットすることができるからである。

ここで、従来法の考え方に回帰する可能性について述べよう。本稿で紹介した分子積分計算の標準法は目覚ましい発展を遂げている CPU の能力に頼った方法である。これらの方法においても分子積分の計算は SCF 計算の大半を占めるので、仮に収束までに SCF 繰り返し計算が 20 回必要になった場合、従来法に比べて 20 倍近く計算時間の損をしていることになる。そこで、全ての分子積分を使い捨てるのではなく、一部をディスクやメモリに保存しておく方法<sup>15</sup>が考えられている。セミダイレクト法、部分ダイレクト法と呼ばれるものである。

小原と雑賀の全ての  $s$  軌道タイプの分子積分を保持すると、これまで説明してきたダイレクト法に比べて格段に速くなることが分かっている。ただし、保持しておく分子積分の種類を増やそうとすると複雑な制御が発生し、様々な並列計算機に適合するアルゴリズムは難しいので、それほど普及しているわけではない。もし全ての分子積分を保持することができるとしたら、計算時間は短縮する。ディスクではなく主記憶に保持することができるならば、計算速度は格段に向上する。今では、 $10^{15}$  Byte の主メモリを登載した超並列計算機が誕生しつつある。これならば、タンパク質計算に必要な 2 電子積分がそっくりメモリに収まってしまいうだろう。

## 交換相関ポテンシャル

この小節では密度汎関数法特有の交換相関項の計算法について概説する。密度汎関数法にお

<sup>15</sup> R. Lindh, J. W. Krogh, M. Schütz, K. Hirao, Theor. Chem. Acc., 110 (2003) p.156.

ける交換相関項の被積分関数は複雑な有理関数  $F$  で記述されている。そのため、上記の分子積分法とは異なる計算方法が要求される。その標準法はグリッドによる数値計算法である。被積分関数が複数の原子核の近傍で大きな値を持つことから、この種の計算は3次元多中心積分と呼ばれる。このような積分には、里子によるボロノイ多面体法<sup>16</sup>を発展させた Becke のファジーセル法<sup>17</sup>が使用される。

ここで、(2.17)式の交換相関項を一般的な多中心積分として取り扱うことにする。

$$I = \int F(\mathbf{r}) d\mathbf{r} \quad (3.8)$$

通常用いられている数値積分法では、多中心積分をそれぞれの原子核を中心とする1中心積分に展開することから始める。

$$I = \sum_A \int w_A(\mathbf{r}) F(\mathbf{r}) d\mathbf{r} = \sum_A I_A \quad (3.9a)$$

$$I_A = \int w_A(\mathbf{r}) F(\mathbf{r}) d\mathbf{r} \equiv \int F_A(\mathbf{r}) d\mathbf{r} \quad (3.9b)$$

$$\sum_A w_A(\mathbf{r}) = 1 \quad (3.9c)$$

このとき導入した重み関数  $w_A(\mathbf{r})$  は原子核座標  $\mathbf{R}_A$  から作られるボロノイ多面体  $\Omega_A$  で定義するのが自然である。すなわち、 $\mathbf{R}_A$  が含まれるボロノイ多面体  $\Omega_A$  のみに値を持ち、それ以外の領域は原子核座標  $\mathbf{R}_A$  の勢力圏外とする。これが里子のボロノイ多面体法である。しかし、重み関数  $w_A(\mathbf{r})$  を階段関数で定義してしまうと  $\Omega_A$  以外からの効果を取り込めず計算精度に問題が生じるし、座標を変化させながら計算する際に連続性も失われる。そこで、Becke は  $w_A(\mathbf{r})$  に階段関数をなました滑らかな関数で定義した。これをファジーセル法と呼ぶ。

続いて、 $I_A$  は原子核座標  $\mathbf{R}_A$  を中心とした極座標系を用いて、

$$\begin{aligned} I_A &= \int F_A(\mathbf{r}) d\mathbf{r} = \int F_A(\mathbf{r}_A + \mathbf{R}_A) d\mathbf{r}_A \\ &\equiv \int f_A(\mathbf{r}_A) d\mathbf{r}_A \quad (3.10) \\ &= \iiint f_A(r_A, \theta_A, \varphi_A) r_A^2 \sin\theta_A dr_A d\theta_A d\varphi_A \end{aligned}$$

と表し、これらの積分はサンプリング点  $\{\mathbf{r}_{Ai}\}$  と重み係数  $\{C_i\}$  を用いた求積法により計算される。

すなわち、

$$I_A \cong \sum_{i=1}^{N_{Grid}} C_i f_A(\mathbf{r}_{Ai}) \quad (3.11)$$

である。ここで  $N_{Grid}$  は1中心あたりのサンプリング点の総数である。一般に、球面積分では球面調和関数展開によるガウスの求積法、動径方向の積分にはガウス・ルジャンドルやオイラー・マクローリン法などが用いられる。球面上にサンプリングした1層の点群を異なる半径  $r_A$  で何層にも配置する。これらサンプリング点をグリッドと通称するため、この方法はグリッド法と呼ばれる。したがって、積分  $I$  は(3.9a)式から、

<sup>16</sup> C. Satoko, Phys. Rev., B30 (1984) p.1754.

<sup>17</sup> A. D. Becke, J. Chem. Phys., 88 (1988) p.2547.

$$I \cong \sum_{A=1}^{N_{atom}} \sum_{i=1}^{N_{Grid}} C_i w_A(\mathbf{r}_{Ai}) f_A(\mathbf{r}_{Ai}) \quad (3.12)$$

と計算する。

この方法で、必要な精度を得るためには莫大な数のサンプリング点が必要である。精度を保ちつつなるべくサンプリング点を少なくしたグリッドが提案されており、標準グリッド<sup>18</sup>と呼ばれている。1 中心あたり 5000 点ほどのグリッドを用いると小数点以下 7~8 桁程度の計算精度があるが、これで必ずしも十分とはいえない。今後の研究で、分子積分のような解析計算ではなくとも、それと遜色ないほどに十分な精度が得られるようになれば、密度汎関数法の計算上の弱点の 1 つが克服されるだろう。

## その他の律速

残る律速計算は、行列積、対角化(計算方法によっては逆行列計算も)といった行列演算である。これらはこれまで説明してきた分子軌道計算特有の計算ではない。個々の計算機種構成でチューニングされ整備されている数値演算ライブラリを使用することが望ましい。

分子積分にしても交換相関積分にしても、カットオフ法もしくは高速多重極展開法(説明は省略)<sup>19</sup>などを使用して  $O(N_{bf}^2)$  以下の計算量となる。となると、実は  $O(N_{bf}^3)$  の計算サイズ依存性

を持つ行列演算こそが深刻で、実際、超巨大分子の計算においては逆転してしまう(次節参照)。

行列演算も SCF 計算毎に繰り返し行われる。計算機システムの能力を最大限に引き出したライブラリの整備が必須である。用途に合わせて、疎行列版と密行列版が必要となるだろう。計算機言語体系別に余計な行列の転置などの操作を要求しないことも大切である。これはコンパイラにもいえることだが、ライブラリの整備なくして高品位シミュレーションの達成はあり得ない。同様に、I/O 性能のチューニングが後回しとなることが多いことも問題である。たとえ先端的な計算機システムが登場し、そのシステム構成に適したアルゴリズムを採用したシミュレーションソフトウェアが存在していても、これらの整備の遅延のため、計算機性能を生かした画期的なシミュレーションがなかなか達成できないことをこれまでしばしば経験している。

## 3. 3. 計算量と並列計算の方法

### タンパク質 Roothaan 方程式の次元

3.2 節のアルゴリズムをもとに、計算律速となるプログラムの演算量を、それぞれ計算規模の指標である原子数、基底関数の数を用いて見積もる。Roothaan 法では非線形方程式の固有値を求める問題を行列方程式の固有値問題に置き換える。そのため、これまでの説明でも登場したように行列のサイズが計算規模を表す指標になる。

タンパク質ではアミノ酸残基の総数と分子量との間に、簡単ではあるが非常によく合う次の経験則が知られている。

$$(\text{タンパク質の分子量}) = 110 \times (\text{アミノ酸残基数}) \quad (3.13)$$

この関係式は、タンパク質は主にペプチド鎖でできていること、アミノ酸をつくる水素、炭素、窒素、酸素、硫黄といった原子の組成比率が巨大な分子で平均化されるためによく合う。

<sup>18</sup> P. M. Gill, B. G. Johnson, J. A. Pople, Chem. Phys. Lett., 209 (1993) p. 506.

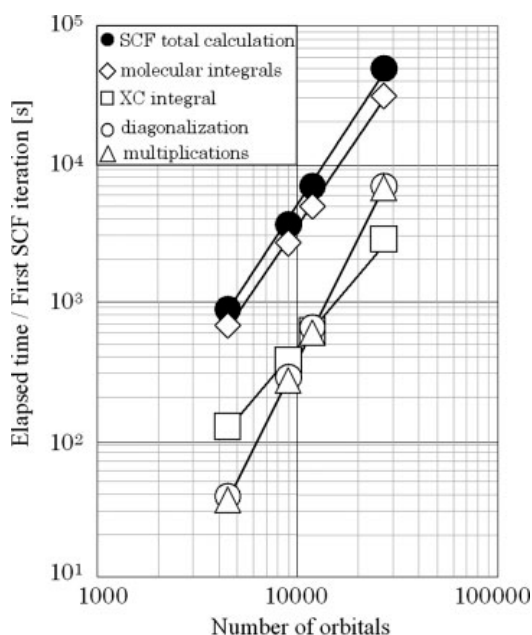
<sup>19</sup> C. A. White, M. Head-Gordon, J. Chem. Phys., 101 (1994) p. 6593.

この経験則を発展させて、タンパク質の総アミノ酸残基数と原子数，電子数，さらには行列の次元との間の関係を導くことができる。この関係は 100 残基のタンパク質の場合を基準にすると次のようになる。

$$\begin{aligned} (\text{アミノ酸残基数}):(\text{原子数}):(\text{電子数}):(\text{行列次元数}) \\ = 100:1600:6200:10000 \end{aligned} \quad (3.14)$$

100 残基中にある 1600 個の原子のうちおよそ 800 個が水素原子，残りの 800 がその他の原子である。行列の次元数は基底関数の総数である。この値は各原子に割り当てる基底関数の種類によって大幅に増減する。上の式の 10000 という数は，大型分子によく使用されるスプリット・バレンスと呼ばれるタイプの基底関数の場合のおおよその目安の値である。

(3.14) 式は，タンパク質の量子化学計算を行う際に用いる大変役に立つ関係式である。100 残基のタンパク質であれば 1 万次元，1000 残基のタンパク質ならば 10 万次元の行列方程式を解くことになる。要素数は次元の 2 乗となるため，100 残基では 1 億要素，1000 残基では 100 億要素にもなる。このサイズの固有値問題を SCF 計算によって何回も解かなければならないので，対象となるタンパク質の計算に必要な演算量や計算時間，I/O 量の予測を欠かすことはできない。



第 3 図: タンパク質密度汎関数法の計算時間。

基底関数の総数と SCF 繰り返し計算の 1 回目の実行時間の両対数プロット<sup>20</sup>。タンパク質分子は小さいほうから，インスリン単量体，インスリン 2 量体，インターロイキン 4，インスリン 6 量体。プロットは，SCF 全体 (●)，分子積分(◇)，交換相関計算(□)，行列対角化(○)および行列積(△)で，それぞれの直線の傾きは 2.3, 2.1, 1.8, 2.9 および 2.9 である。計算には SGI Altix3700 の 16CPU を使用した。

### 律速計算の計算量

Kohn-Sham-Roothaan 法における律速計算である，分子積分，数値積分，行列対角化，行列積において，高速化を施した実際のタンパク質密度汎関数計算における分子サイズ依存性を見て

<sup>20</sup> T. Inaba, F. Sato, J. Comp. Chem., 28 (2007) p.984.



おこう。いろいろなサイズのタンパク質の密度汎関数計算を行い、その基底関数の総数  $N_{bf}$  と SCF 繰り返し計算の 1 回目にかかった実時間を両対数プロットで示したものが図 3 である。サンプルはインスリン単量体、インスリン 2 量体、インターロイキン 4、インスリン 6 量体で、基底関数の総数が約 4000 から 30000 までの本格的な計算である。交換相関ポテンシャルには SVWN を、分子積分の高速化にはカットオフ法と RI 法<sup>21</sup>(説明は省略)を採用している。SCF 計算を、分子積分、交換相関項の数値積分、行列の対角化、行列積の計算に分け、それぞれ◇、□、○、△で示した。●は SCF 計算全体である。

各プロットを結んだ直線の傾きはそれぞれ 2.1, 1.8, 2.9, 2.9 および 2.3 である。この値は、計算にかかる時間が基底関数の総数  $N_{bf}$  の何乗に依存しているかを示している。分子積分および交換相関項の計算はそれぞれ 2.1 乗と 1.8 乗で、カットオフ法による高速化の理想値である 2 乗に近い結果が得られている。一方、行列演算は行列積、行列対角化のいずれもほぼ基底関数の総数の 3 乗に依存している。基底関数の数が 30000 程度までであれば、行列演算はまだ SCF 全体の一部しか占めず、SCF 全体の計算サイズ依存性は  $O(N_{bf}^{2.3})$  に抑えられている。

しかし、タンパク質のサイズがさらに大きくなると、行列演算が計算律速となることは明らかである。立体構造が明らかにされているタンパク質の 99%以上が 1000 残基以下である。1000 残基のタンパク質の基底関数の総数は 100000 程度である。このとき、分子積分と行列演算はほぼ同程度の割合となる。

## 律速計算の並列化

これまで見てきたように、Kohn-Sham 方程式の計算時間の上限は実質  $O(N_{bf}^3)$  である。仮にアミノ酸 1 残基の計算が 1 秒で終わる計算機があった場合、その計算機で 1000 残基のタンパク質正準分子軌道計算にかかる計算時間の上限値は 11574 日 (32 年) になる。個々の計算の最適化や計算の省略で多少高速化できたとしても、今のところ 1 CPU の計算機による計算は実行不可能である。そこで必要となるのが並列計算の技術である。密度汎関数計算の律速計算は、原理的には以下のように全て並列化できる(行列演算の説明は省略)。

計算律速となる 2 電子分子積分は密度汎関数法計算の中で、(2.12b) 式の Kohn-Sham 行列の生成、(3.3) 式の全エネルギー計算で使用される。これらの共通した表現が (3.5) 式であった。これから直ちに、インデックス  $rs$  を分割して計算を各計算機に振り分けることで  $\mathbf{U}$  の並列計算を行えることがわかる<sup>22</sup>。この手法を取れば並列計算時に利用するメモリ量を削減でき、計算開始時のデータ転送量も抑えることができる。またアップデート法を採用することにより、並列計算中のデータ転送量をさらに抑えることができる。

$pq$  のインデックスで計算を各計算機に振り分ける手法もある。インデックス  $rs$  で並列化を行うよりも簡単な制御で並列化できるが、(3.5) 式の場合、 $\mathbf{P}$  の行列要素を全てのプロセッサで保持する必要がある。これに対して高島らは Hartree-Fock 計算において、カットオフ法を利

<sup>21</sup> B. I. Dunlap, J. W. D. Connolly, J. R. Sabin, J. Chem. Phys., 71 (1979) p. 3396.

<sup>22</sup> T. Yoshihiro, F. Sato, H. Kashiwagi, Chem. Phys. Lett., 346 (2001) p. 313.

用して  $\mathbf{P}$  行列を圧縮する効果的な並列化を提案し、これを RT 法<sup>23</sup>と名付けている。また、稲葉らにより RI 法の分子積分計算をプロセッサに均等分配する手段としてカットオフ法を利用する手法<sup>20</sup>が提案されている。

数値積分は(2.17)式の交換相関項の積分計算に使用される。これらの被積分関数は原子核上で大きな値を持つため、(3.9)式のように原子核を中心とする部分空間の関数  $F_A(\mathbf{r})$  に分割し、その和として計算する。したがって、これらの数値積分は原子ごとに独立して計算ができるため、並列化は容易である。原子のインデックス  $A$  で計算をプロセッサに振り分ければよい。

このような並列計算の手法はあくまでサンプルである。並列計算を行う手法は多種多様なため、用いるアルゴリズムや計算機システムに最適な並列計算手法を選択することが望ましい。

表 1: インスリン 6 量体の SCF 計算 1 回転目での各計算ルーチンにかかった実時間と並列度、並列化効率<sup>20</sup>

計算ルーチン	8CPUs		16CPUs		32CPUs		64CPUs			
	実時間 (秒)	実時間 (秒)	並列度	並列化 効率(%)	実時間 (秒)	並列度	並列化 効率(%)	実時間 (秒)	並列度	並列化 効率(%)
分子積分	59,353	31,247	15.20	94.97	15,923	29.82	93.19	8,671	54.76	85.56
数値積分	4,225	2,695	12.54	78.39	1,386	24.39	76.21	760	44.47	69.49
行列対角化	11,977	6,901	13.88	86.78	3,558	26.93	84.16	1,899	50.46	78.84
行列積	12,766	6,630	15.40	96.28	3,418	29.88	93.37	1,840	55.50	86.73
その他	3,553	2,350	12.10	75.60	2,207	12.88	40.25	756	37.60	58.75
合計	91,874	49,823	14.75	92.20	26,492	27.74	86.70	13,926	52.78	82.47

並列度: 1CPU の計算時間は 8CPU でかかった計算時間より概算した。

並列化効率 = (並列度 / CPU 数) × 100

表 1 に示すのはタンパク質密度汎関数計算の並列化効率である。計算方法は RI 法で、SVWN 交換相関汎関数を用いた。計算対象はインスリン 6 量体で、アミノ酸 306 残基、原子数 4728、電子数 18552、基底関数の数 26790 である(詳細は 4 章参照)。この計算を 1.3 GHz の Itanium 2 並列計算機のそれぞれ 8, 16, 32, 64 CPU を用いて並列計算を実行した。SCF 繰り返し計算の 1 回目で各種計算にかかる実時間を測定して並列化効率を求めた。

計測した計算はそれぞれ、分子積分、数値積分、行列対角化、行列積で、SCF 繰り返し計算中のディスク入出力時間などはすべてその他にまとめている。並列化アルゴリズムは、分子積分には  $pq$  インデックスによる計算分配を、交換相関項の数値計算には原子ごとの計算分配を適用し、行列対角化と行列積は数値演算ライブラリ ScaLAPACK を用いて並列化を行っている。この例では、1 CPU での計算時間が計測できないため、8 CPU での計算時間を 8 倍して 1 CPU の計算時間とし、並列度と並列化効率を求めている。

並列計算では使用する CPU 数が増大すると、Amdahl の法則により、並列度、並列化効率ともに飽和していく傾向が出てくる。この並列計算例でも CPU 数が増えるにつれて並列化効率が若干下がってきているが、SCF 繰り返し計算の合計時間で並列化効率が 80%以上とおおむね良好な結果が得られている。この計算におけるプログラムの並列化率は 99%以上と見積もられる。およそ 500 CPU 規模の並列計算機で効率を出すことができるレベルである。10000 CPU 規模の並列計算機に対応するためには、さらにこれらの律速ルーチンでデータ転送やディスク I/O な

<sup>23</sup> H. Takashima, S. Yamada, S. Obara, K. Kitamura, S. Inabata, N. Miyakawa, K. Tanabe, U. Nagashima, J. Comp. Chem., 23 (2002) p.1337.

どに工夫が必要になるだろう。

図2のように多種多数のルーチンを渡り歩く SCF 計算では、やむなく中間データをディスクに書き出すことが多く、これもデータ転送とディスク I/O が問題となる。タンパク質のような大規模計算では顕著である。量子化学計算において並列化効率を上げるためには、各律速ルーチンに対して計算機に適した計算アルゴリズムを採用しチューニングするだけでなく、 $h, S, X$  行列や SCF 計算ごとに更新される  $P^{(\lambda)}, U^{(\lambda)}, F^{(\lambda)}, F'^{(\lambda)}, C^{(\lambda)}, C'^{(\lambda)}$  といった行列は、データ転送を極力抑えながら各律速ルーチンで利用しやすい形式でメモリに保持する必要がある。1000 残基タンパク質ではこれらの行列は1個で  $10^{11}$  Byte にのぼるため、保存するには数  $10^{13} \sim 10^{15}$  Byte のメモリが必要となる。10000 CPU を超える計算機なら十分可能な値である。

#### 4. タンパク質全電子計算に特有の技術

この章ではタンパク質のための Kohn-Sham-Roothaan 方程式の解法について解説する。タンパク質は巨大分子であるため低分子と同じ方法では解くことができない。SCF 計算が収束するための第1条件は良好な初期値を準備することであり、第2は滑らかに収束させることである。タンパク質は細胞の中の複雑な環境にあるので、その効果も取り込まなければならない。

##### 4. 1. 計算シナリオと収束方法

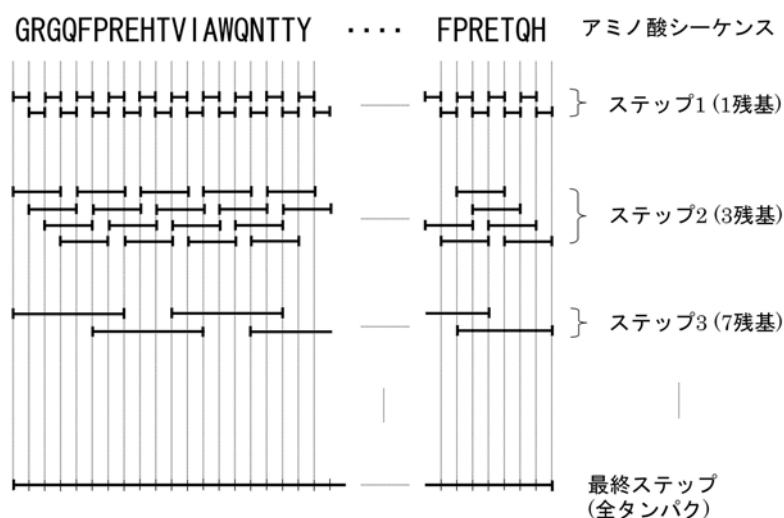
SCF 計算には初期値が必要である。この初期値が悪いと、収束するまでに何百回もの繰り返し計算を行わなければならないか、発散して解が得られなかったりする。実は小さな分子の場合でも難しい問題で、タンパク質の場合、大部分は多数のアミノ酸が結合した巨大分子であるためさらに収束が難しい。アミノ酸は多かれ少なかれ互いに似ているので、わずかな環境の変動で軌道成分の混じり方に影響を受ける。タンパク質は半導体的な性質を持っているので、場合によっては電子が遠くまで走ることもある。その上、密度汎関数法では Hartree-Fock 法よりも被占軌道と空軌道のエネルギーが接近しやすいため、これらの成分の入替えが起きる傾向があり、いっそう収束が難しくなる。このように難しい性格の計算を成功させるためには、タンパク質の構造とそれにとまなう性質を充分理解して注意深く計算をすすめるなければならない。

タンパク質の Kohn-Sham-Roothaan 方程式を解く難しさを克服するために、計算シナリオと呼ぶ新しい手続きを導入した<sup>1)</sup>。初めからタンパク質全体を解こうとはしないで、アミノ酸を1つ1つ解くことから始め、計算する分子を徐々に大きくしていく。図4がその計算シナリオを示したものである。まず、タンパク質のペプチド鎖を形成している個々のアミノ酸残基をそのままの構造で取り出し、窒素には水素を補ってアミノ基としカルボニル炭素には水酸基を付けてカルボン酸にする。原子ごとの電子密度からこのアミノ酸の電子密度を推定して Kohn-Sham-Roothaan 方程式の解を求める。これがステップ1である。

ステップ2では、連続する3個のアミノ酸の解から電子密度を取り出し結合して3残基ペプチドの初期値を作る。この際、付加したアミノ基の水素とカルボン酸の水酸基など不要部分の電子密度は取り除く。この作業はちょうどペプチド結合の生成に当たる。電子密度の推定値が得られたら SCF 計算を行って解を求める。

ステップ3では3残基ペプチドの中央のアミノ酸残基の電子密度のみを取り出し、数残基ペプチドの初期値を作る。このために隣り合う3残基ペプチドは2個のアミノ酸残基が重なるようにしておく。中央のアミノ酸残基の電子密度を両側のアミノ酸残基の影響を取り込んだもの

にするための手続きである。この取り込みは後続のステップでも行う。ペプチドの両端については端のアミノ酸残基の解も取り入れる。この手順を、例えば7残基ペプチド、20残基ペプチドというように次第にペプチドを長くしていき、最後にタンパク質全体のSCF計算を行う。



第4図: タンパク質の計算シナリオ[1]。

タンパク質のSCF計算を滑らかに収束させるための計算手順。図の最上部はアミノ酸の略号である。途中のH型の記号はそれぞれのステップで計算する分子を示す。ステップ1では個々のアミノ酸を、ステップ2では3残基ペプチドをというように徐々に大きくしていき、最後に全タンパク質の計算を行う。

この種の計算にはいくつか難しい問題点がある。前のステップの電子密度はSCF計算の過程でそのステップの分子全体として成り立つように最適化されたものである。このため電子密度の単純な切り貼りは誤差を生み出し、次のステップの良好な初期値を与える保証はない。そこで計算した電子密度の精度を落とさないようにペプチドを長くしていく方法として開発したのがQCLO法<sup>24</sup>である。

QCLOとは擬カノニカル局在化軌道 Quasi-Canonical Localized Orbital の略称で、1種の局在化分子軌道である。Kohn-Sham 演算子はユニタリ変換によって方程式は不変である。言い換えると、カノニカル分子軌道もそれをユニタリ変換した軌道も表現が違っただけで実質的には同じ解である。適切な基準を選ぶと、カノニカル分子軌道をユニタリ変換して、化学結合や孤立電子対などの局在化軌道を作ることができる。ここで、図4のシナリオの各ステップで計算する分子のことをフレーム分子、分割した軌道空間のことをフラグメントと呼ぶことにする。ステップ2以降はQCLOの計算を行い、これを使って次のステップのフレーム分子の初期値を求める。

フラグメントの分割は結合や孤立電子対などの境で行えば任意であるが、一般にN端-側鎖-ペプチド結合-側鎖-ペプチド結合-側鎖-...-C端のフラグメントに分割し、側鎖のフラグメントに関しては $\alpha$ 炭素を境界に $C_\alpha$ -H結合、 $C_\alpha$ の1sコアおよび側鎖のすべての結合や孤立電子対はこの側鎖のフラグメントに所属させ、N- $C_\alpha$ 結合は左側のペプチドのフラグメントに、 $C_\alpha$ -C結合は右側のペプチド結合にそれぞれ帰属させると一意に分割できる。

<sup>24</sup> H. Kashiwagi, H. Iwai, K. Tokieda, M. Era, T. Sumita, T. Yoshihiro, F. Sato, Mol. Phys., 101 (2003) p. 81.

フレーム分子のカノニカル分子軌道が得られたら、まず通常の局在化軌道を求める。1つのフラグメントのQCLOを求めるには、そのフラグメントに所属する局在化軌道を選び出し係数行列 $\mathbf{C}_{lo}$ を作る。この行列の縦の次元はフレーム分子の基底関数の数であり、横の次元はそのフラグメントに所属する局在化軌道の数である。 $\mathbf{C}_{lo}$ を用いて次式のようにフレーム分子全体のKohn-Sham行列 $\mathbf{F}$ を線形変換すると、そのフラグメントのKohn-Sham行列 $\mathbf{F}'$ が得られる。

$$\mathbf{C}_{lo}^\dagger \mathbf{F} \mathbf{C}_{lo} = \mathbf{F}' \quad (4.1)$$

この行列の次元はフラグメントに所属する局在化軌道の数になる。これを対角化するとQCLOが得られる。

$$\mathbf{F}' \mathbf{C}' = \mathbf{C}' \mathbf{E} \quad (4.2)$$

1つのフラグメントのQCLOは(4.1)式と(4.2)式を解いて得られる軌道として定義される。フラグメントの局在化軌道のみで展開しているから局在化軌道であり、(4.2)式のカノニカルな解であるから、そのフラグメントのカノニカル軌道としての性格を兼ね備えている。この性質から擬カノニカル局在化軌道と名づけた。

次のステップ3以降では2つの計算方式がある。1つは、フレーム分子の局在化軌道を求めなおして(4.1)式の基底として用いる方法がある。もう1つは基底として前のステップで求めたQCLOを使う方法である。前のステップの小さなフラグメントのQCLOを用いて $\mathbf{C}_{lo}$ を作り、大きなフラグメントのQCLOを計算する。前者の方が精度はよいが計算労力が大きい<sup>25</sup>。

計算シナリオの途中で得られるQCLOは分割したフレーム分子の解であり、またフラグメントごとに基底関数を限定したものである。したがって、全分子のカノニカル軌道をユニタリ変換して得られる局在化分子軌道とは異なる。しかし、各ステップのQCLOは全分子の局在化軌道のかなりよい近似になっていると考えられる。

QCLOをSCF計算の初期値作成に用いるためには次のような手順をふむ。アミノ酸1個ずつの計算から3残基ペプチドのSCF計算までは前に述べた手順と同じである。ステップ3の数残基ペプチドの電子密度の初期値をステップ2のQCLOから求める。切断と接続はフラグメント単位で行う。以下のステップも同様である。この効果は5.1節で明らかになる。

## 4. 2. 計算構造

計算に用いる分子構造の決定は大変重要である。化学分子の場合、基準となる分子構造は実験で得られた構造以外にも、典型的な結合長や結合角、二面角を与えて組み上げた模型構造を採用したり、シミュレーションで用いるKohn-Sham方程式のもとでエネルギー最小構造を求めたりする。これらの構造で分子の持つ基本的な性質を理解することに成功してきた。

タンパク質も分子であるから同じことが期待できるが、タンパク質には基準となるあるいはシミュレーションに適切な構造を決定する際に、特に考慮に入れなければならない問題がある。以下に簡単に整理してみよう[2]。

### 1. 実験構造

タンパク質立体構造は比較的容易にX線構造解析や多次元NMR法などの実験で求めることができるようになった。多くの構造はPDB<sup>26</sup>に登録されている。これをシミュレーション構造に

<sup>25</sup> N. Nishino-Uemura, T. Hirano, F. Sato, J. Chem. Phys. 127 (2007) p.184106.

<sup>26</sup> <http://www.rcsb.org/pdb/home/home.do>

採用することが最もよく行われている。しかし、X 線構造解析では一般に水素の位置が見えないし、窮屈で位相の揃った結晶場中の構造である。NMR 法にしても、原子間距離から構造推定するためモデルがたくさんできてしまう。両者に共通で、原子核位置を決定するリファインメントには経験的な力場を用いたモデル作りの過程が入り込む。

## 2. 系の取り扱い

化学分子では孤立系のシミュレーションが標準的に行われるが、タンパク質の性質や機能を調べるのに適するかどうかは疑わしい。少なくとも古典法では溶媒を考慮に入れる。ただし、その古典法にしても完璧ではない。例えば、0.05 M 緩衝液の中に 5  $\mu\text{M}$  のタンパク質が入っているとすると。水は 55 M だから、タンパク質とイオンと水の分子数比は 1:20000:10000000 である。このような環境をシミュレーションで作出すことは難しい。溶媒分子の取り扱いもモデル化が必要となる。タンパク質の表面をどう扱うかも悩ましい問題である。

## 3. 最適化

タンパク質は最安定構造を求めることも困難である。取り扱う原子数を  $N_{atom}$  とすると、 $(3N_{atom} - 6)$ 次元空間の最適化問題となるからである。タンパク質は数千~数十万原子からなる分子であるし、もしこれに溶媒分子を含むならば  $N_{atom}$  に水分子とイオンを加えなければならない。当面は  $N_{atom}$  を小さく抑える局所構造最適化が現実的であるが、その手法にしても注意が必要である。タンパク質に水素を付けてこれを最適化することを考えてみよう。よく活性中心となるセリン、トレオニン、チロシン、システインの OH 基や SH 基の水素の位置は、他の原子が邪魔をして 2 面角が  $360^\circ$  回れず、最初に付加した位置に依存してしまうことが多々ある。ヒスチジンのイミダゾール環は  $\epsilon$ 位と  $\delta$ 位に窒素がある。他に情報が無ければ、どちらに水素が付くか分からないし、両方に結合してカチオンとなることもある。適切なモデルでないとタンパク質の反応機構を間違えて推定してしまうかもしれない。

## 4. 量子計算固有の問題

PDB に登録されている立体構造は、化学的な見地からは異常な構造歪みや原子間衝突などを残したままリファインメントが行われているケースが多々見られる。水素付加を行うと、しばしば原子間衝突が起こってしまう。量子化学計算結果はこのような構造に敏感で、反応性の高い軌道を与える。タンパク質は N 末端、C 末端以外にも多くの解離基を持っている。孤立系でタンパク質の密度汎関数計算を収束させることはほぼ絶望的である。

結局、タンパク質の計算構造は一意には決まらない。上記の問題点を考慮しつつ、シミュレーションによって明らかにしたい内容に基づき、適切な構造を決定すべきである。これこそがタンパク質シミュレーションの最も重要なプロセスの 1 つなのである。

これまでの経験で著者らが開発した、現在の計算機環境でできるモデリング法を「おまけ」として紹介しておこう [2]。タンパク質密度汎関数計算を比較的安全に達成できる計算構造を構築するテクニカルな話であり、決定版でもない。

本来、分子の立体構造は、シミュレーションを行う支配方程式、すなわち本ケースならば Kohn-Sham 方程式に基づいて最適化されることが望ましい。将来は実現されるであろうが、今のところ構造最適化は古典分子力場を用いて行うのが現実的である。精度は劣るが、立体構造の歪みや原子間距離の異常の修正に大いに役に立つ。化学分子でも、第 1 原理でなんなく構造最適化できるようになったのはごく最近の話で、古典モデリングは標準的に行われていた。

構造最適化には、分子動力学 (MD) 法やモンテカルロ (MC) 法といった系の状態をシミュレーシ

オンする手法と、最急降下法や共役勾配法といったエネルギーの極小値を解く手法がある。前者は大域的な構造探索がなされるため、立体構造の歪みを修正するばかりでなく初期構造を大きく変化させてしまうことがある。後者は初期構造近傍のエネルギー極小点へ収束するため、初期構造はそれほど大きくは変化しない。実験データがあり、これになるべく近い構造で立体構造の歪みや局所構造異常を修正したい場合は、後者を選ぶ方が無難かもしれない。またこのとき、溶媒に連続近似を選択することも一考である。

それでも、タンパク質の構造最適化はなかなか収束しないことが多い。膨大な自由度が原因である。ただし、古典分子力場での構造最適化であるから、不必要なまでに精度を高め、厳しい収束判定を課す必要は無い。適切な結果に至らない場合は、その都度原因を解析し、立体構造を修正する。その妥当性は最終的には計算者の判断に委ねられる。

表面の取り扱いには様々なモデルが考えられるが、その1つに解離基の周りにこれと反対の電荷を持ったイオンを近づける緩衝液モデルがある。実際には、これらのイオンは水分子同様、動き回っていると考えられるが、溶媒が近づける表面にある解離基の近くにカウンターイオンを付加するモデルで比較的安定に密度汎関数計算を達成できる。

タンパク質を取り巻く水は極性を持った液体の分子としては特殊なものである。当然ながら、タンパク質に影響を与えている。溶媒としての水分子のモデルは様々ある。もしあらわに扱うのであれば、その構造最適化にもタンパク質と同等のモデルを採用し、溶媒水分子の構造を再現させることが望ましい。また、タンパク質の特定の部位に取り込まれて動きが鈍い水分子がある。これらは少なくともタンパク質の一部としてあらわに扱うべきである。

これらの方法で決定された構造によるタンパク質の全電子計算では、この系全てを量子系として扱うこともあるし、量子系はタンパク質のみでイオンや水分子は点電荷として取り扱うこともある。このような選択もシミュレーションの目的次第である。とはいえ、溶媒分子や細胞膜などの環境まで量子化学計算を行うのは大きな負担である。複数のサブユニットから構成される大型のタンパク質の場合には、量子化学計算を活性中心を含むサブユニットに限定せざるを得ないこともある。このような場合に用いるのがQM/MM法である。

一般的なQM/MM法はその名称が示すとおり、全系を量子化学的に取り扱う部分と古典的な分子力場による部分に分割して計算する方法で、量子化学計算の計算労力が大きいため考え出されたものである。基本的なアイデアは、計算機の能力が低分子を計算するのにも不十分であった時代に提唱されたものである。1976年にはWarshelとLevittによるタンパク質の先駆的な計算<sup>27</sup>が行われ、今でもよく利用されている。

QM/MM領域の切り分け方は原理的には任意であるが、量子系領域にタンパク質全体、古典系領域に溶媒を指定すれば、比較的うまく行くだろう。ここでは、タンパク質の全電子計算を行う際の周囲の水の効果を取り込む便宜的な取り扱いとして、タンパク質のまわりの水分子を点電荷として扱う方法を紹介する。水の水素原子と酸素原子はそれぞれいくらかの正、負の電荷を帯びている。そこで、水分子の水素、酸素に点電荷を割り当てるのである。この水のモデルはタンパク質に静電場を与えるのみである。一般に点電荷と点双極子により任意の場のポテンシャルは表現できる<sup>27</sup>ため、溶媒の水が外場としてタンパク質に与える効果を記述できる。

この静電場は、タンパク質の全電子計算にはKohn-Sham行列の1電子項(2.13)式の第2項である原子核からのポテンシャル・エネルギーの形で、外場として点電荷を追加すればよい。

<sup>27</sup> H. Nakamura, J. Mol. Graphics, 11 (1993) p. 30.

$$h_{pq} = \int g_p^*(\mathbf{r}) \left[ -\frac{1}{2} \Delta - \sum_A^{N_{atom}} \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} + \sum_B^{N_{ext}} \frac{q_B}{|\mathbf{r} - \mathbf{R}_B|} \right] g_q(\mathbf{r}) d\mathbf{r} \quad (4.3)$$

第 3 項が外場で、 $B$  は外場のすべての点電荷を走り、 $q_B$  はその電荷である。この式には分子軌道の係数も電子密度も含まれていないので、SCF の繰り返しで再計算の必要がない。追加された 1 回分の積分計算のコストが増えるのみで、経済的に溶媒効果を取り込むことができる。

#### 4. 3. 収束法

タンパク質など、巨大分子の SCF 計算で使用される収束法について簡単にコメントしておこう。紙面の都合から文献を挙げる程度にとどめる。これまで述べてきたように、SCF 計算が成功する鍵は適切な計算シナリオと良い初期値の作成である。それに加えて、収束法は SCF 計算の過程を制御し計算の達成を補助する技術である。収束法には大きく分けて 2 つの役目がある。1 つは SCF 計算を安全に収束へと導くことであり、もう 1 つは収束を加速する役目である。プロジェクション法<sup>28</sup>、軌道の重なり法、レベルシフト法<sup>29</sup>やシンプルミキシング法はどちらかといえば前者の方法で、Anderson 法<sup>30</sup>や DIIS 法<sup>31</sup>は後者の性格が強い。それぞれ特徴があるので適切な場面で用いると効果を示す。

SCF 計算を安全に収束へと導く技術は特にタンパク質のような巨大分子で欠かすことができない。やさしく注意深く収束解に導かないかぎり SCF 計算は振動したり発散したりしてしまう。

分子軌道計算では、あらかじめ計算する分子の電子数と同時に電子の軌道への詰め方を決めておかなければならない。閉殻の基底状態では、Pauli の排他原理に従って、普通は、エネルギーの低い軌道から順に 2 つずつ電子を詰めていく。それから密度行列を計算し、SCF 計算が繰り返される。ところが、タンパク質の軌道エネルギーは固体のバンドのように密に分布し、被占軌道と空軌道とのギャップが狭い。その実測値は実験方法や測定条件によって値が大きく変わるが、おおよそ 3~7 eV と見積もられている<sup>32,33</sup>。これを Hartree-Fock 法で計算すると、しばしば 10 eV 以上のギャップが開いて<sup>34</sup>絶縁体となってしまう。Kohn-Sham 法、特に局所密度法ではかえって 1~2 eV と狭くなる<sup>1,35</sup>。SCF 計算の始めでは、被占軌道と空軌道とのギャップがより狭くなり、密度汎関数法ではしばしば軌道の順番も入れ替わる。このような症状は SCF 計算の初期値が悪いほど顕著となる。開殻系の計算では、閉殻と空軌道の狭いギャップの間に開殻軌道がくるので、ますます収束が難しくなる。このように軌道エネルギーが密に分布している場合に正しく電子を詰める方法が、SCF  $n$  回転目の閉殻・開殻・空軌道がどの程度  $n-1$  回目の軌道の空間に含まれているかを示す指標を計算するプロジェクション法、今回の分子軌道と前回の軌道の重なり大きさで対応関係を判断する軌道の重なり法、特定の軌道について望みの値だけ軌道エネルギーをずらすレベルシフト法や、電子状態の急激な乱れを防ぐため今回

<sup>28</sup> 岡崎功, 九州工業大学情報工学部博士論文(2000).

<sup>29</sup> R. Carbo, J. A. Hernandez, F. Sanz, Chem. Phys. Lett., 47 (1980) p. 581.

<sup>30</sup> D. G. Anderson, J. Assoc. Comp. Machinery, 12 (1965) p. 547.

<sup>31</sup> P. Pulay, Chem. Phys. Lett., 73 (1980) p. 393.

<sup>32</sup> D. D. Eley, D. I. Spivey, Trans. Faraday Soc., 56 (1960) p. 1432.

<sup>33</sup> J. Ladik, F. Bogar, B. Penke, Int. J. Quant. Chem., 98 (2004) p. 522.

<sup>34</sup> S. Suhai, J. Kaspar, J. Ladik, Int. J. Quant. Chem., 17 (1980) p. 995.

<sup>35</sup> T. Inaba, N. Tsunekawa, T. Hirano, T. Yoshihiro, H. Kashiwagi, F. Sato, Chem. Phys. Lett., 434 (2007) p. 331.



の SCF 計算で求めた物理量を前回の物理量で薄めるシンプルミキシング法である。

一方、SCF 計算の繰り返し回数が 100 回、200 回かかっても正しく収束するならば、計算結果の議論には問題はない。しかし、巨大分子の量子計算はコストが高く時間がかかる。混ぜ合わせ率を状況に合わせて変化させる動的なミキシング法である Anderson 法や、過去の物理量を使って残差を最小にする外挿を行う DIIS 法などの収束加速はそのまま計算時間の短縮、しいては計算の実現に貢献する。したがって、収束加速法は実際の計算達成には欠かせない技術である。

適用の仕方は様々で、シンプルミキシング法、Anderson 法、DIIS 法は同時に適用できないが、軌道の重なり法やプロジェクション法、レベルシフト法はこれらと組み合わせることができる。

## 5. 将来展望：タンパク質反応シミュレーション

### 5. 1. タンパク質全電子計算の現状

1997 年の Challacombe と Schwegler の先駆的な P53 の Hartree-Fock 法による計算 (3836 軌道)<sup>36</sup>に端を発し、2000 年には著者らのヘムタンパク質シトクロム *c* の密度汎関数計算 (9600 軌道)<sup>1</sup>と高田らの PKC の Hartree-Fock 計算 (8672 軌道)<sup>5</sup>が達成された。本格的な生体分子量子シミュレーションの歴史はまだほんの 10 年である。

この節では、これまで解説してきたいろいろな技法を使ったインスリン 6 量体の密度汎関数法計算<sup>20,35</sup>を紹介する。その計算規模は 3 章で示したが、執筆時点で世界最大のカノニカル分子軌道計算であり、本稿が展望する将来のタンパク質計算への一里塚である。

インスリンは血液中の糖분을制御するタンパク質ホルモンである。インスリン単量体は 21 残基の A 鎖と 30 残基の B 鎖がジスフィルド結合でつながった小型のタンパク質である。単量体の表面には解離基が多く、凝集して 6 量体になる傾向が強い。糖尿病のための製剤は高濃度のため 6 量体が形成された状態で投与されるが、ホルモンとして働くのはインスリン単量体である。体内で単量体に分解する時間が長くかかり効き目がすぐに現れないことが問題になっていて、アミノ酸残基の置換などにより分解しやすい製剤も開発されている。

実は、インスリン 6 量体の計算は稲葉らによって 2005 年に成功していた。その後、この巨大系で安定した解法を探るべく、様々な技法が考案されテストされた。その結果、最終的に採用されたのは以下で説明する方法である。まず、分子構造についての予備計算を行った。NMR 法で得られた分子構造を用いているが、データベースの構造には局所的な歪みがあって量子化学計算に耐えられない。また水溶液中にあるタンパク質の計算をするために、水和イオンや溶媒の配置も重要である。この過程で 4.2 節のような手順が開発された。

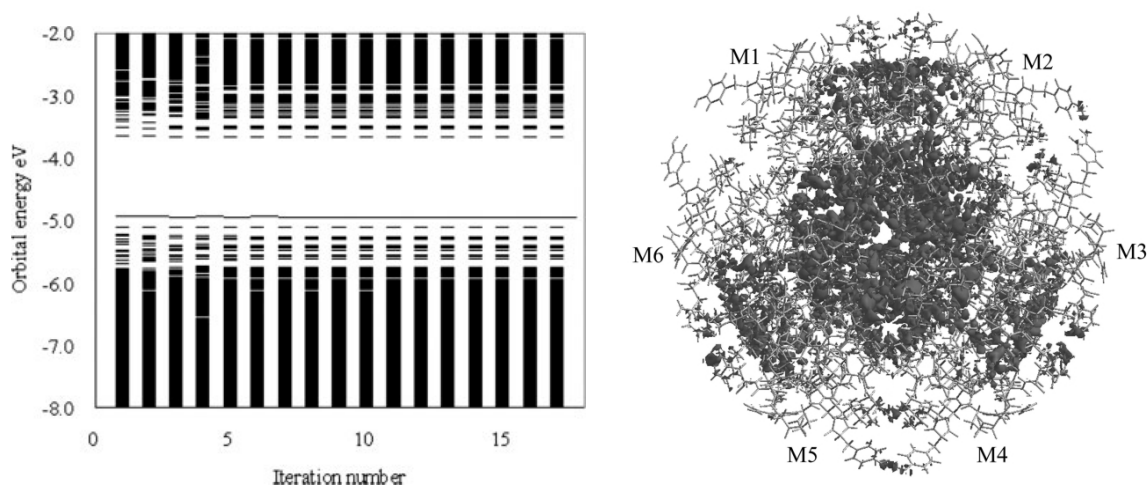
全電子計算は、4.1 節の QCL0 法を採用するとともに、各単量体の全体計算が終わった段階で、他の単量体の原子および周辺の水分子を点電荷として計算に取り込むことが安全な収束へのもう 1 つのポイントとなった。最終ステップの全電子計算の収束加速には Anderson 法が有効であることを発見し、これを用いて SCF 繰り返し計算 17 回で収束した。1.3 GHz の Itanium 2 プロセッサ 64 台 (理論ピーク 333 GFLOPS) で 65 時間であった。

図 5 にインスリン 6 量体の最終ステップの軌道エネルギー変化を示した。SCF 計算の最初の 3 ~ 4 回にわずかな変化が見られるが、それ以降の軌道エネルギーのパターンはほとんど変化していない。良好な初期値と収束制御によりスムーズに収束に向かっている様子が見られる。バ

<sup>36</sup> M. Challacombe, E. Schwegler, J. Chem. Phys., 106 (1997) p. 5526.

バンドギャップも約 1.3 eV で安定していて、被占軌道が空軌道で置き換わるような気配は一切ない。全エネルギーの初期値と収束値の差は 0.15 au であった。

単量体で計算を止めても電子密度は十分よいように思えるが、実際は広く電子分布が変化する。図 6 は単量体の計算と 6 量体の計算結果における電子密度分布の差を示したものである。このような結果から、1 章で述べたように、タンパク質は凝集などの摂動に対し、広範囲にわたって少しずつ電子分布を変えることも立証された。



第 5 図: インスリン 6 量体の最終ステップの軌道エネルギー変化<sup>20</sup>(左)。

縦軸は軌道エネルギー，横軸は SCF 計算の繰り返し回数。下部が占有軌道のエネルギーで，一番上のつながっている横線は最高被占軌道のエネルギー，その上は空軌道のエネルギーである。

第 6 図: インスリンの 6 つの単量体と 6 量体の電子密度分布差(右)。

M1, M2 などは単量体である。6 つの単量体と 6 量体との電子密度分布の差を図示した。単量体ごとの計算では他の単量体の原子にはマリケン電荷を置いている。濃淡の差はプラス，マイナスを表す。

本稿で解説したタンパク質の Kohn-Sham-Roothaan 方程式を解くための技法を用いて、すでに 100 種類のタンパク質の計算が行われている。さらに改良が進めば、タンパク質波動関数計算の完全自動化も夢ではないだろう。

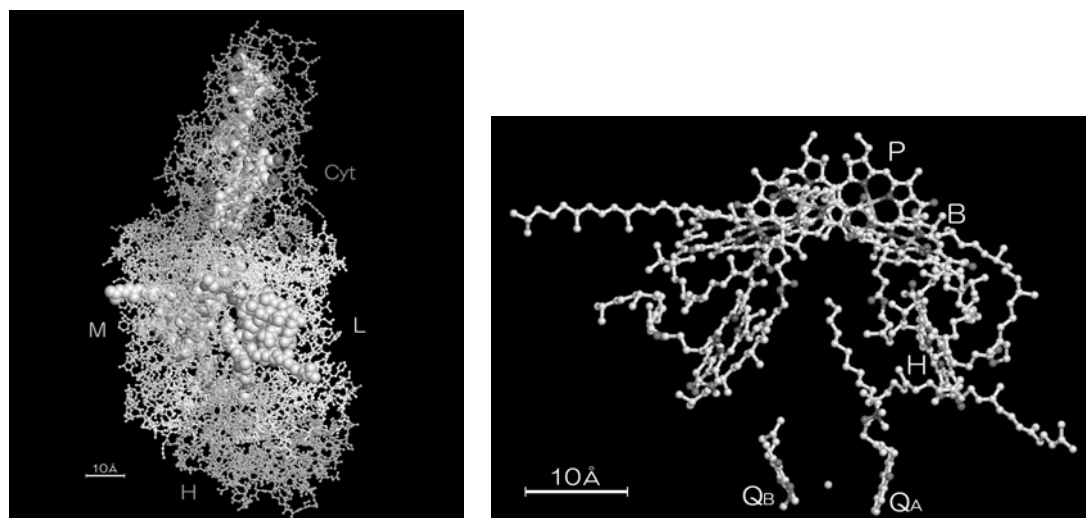
## 6. 2. タンパク質反応シミュレーション

最後に、今後飛躍的な発展が期待されるタンパク質密度汎関数法による反応シミュレーションの将来を展望する。

タンパク質で特徴的な反応に分子内電子移動や酵素による触媒反応がある。ミトコンドリア内膜の呼吸鎖では、タンパク質複合体が整然と並配置されている。これらのタンパク質はアミノ酸残基以外の補欠分子族を豊富に持っていて、電子はタンパク質内でも 10 Å を超える長距離を移動し、さらに複合体から複合体へと電子を伝達し、最終的には酸素を水に変換してエネルギー物質の生産を行う。高度に組織化された、無駄のない生体のエネルギー生産工場である。

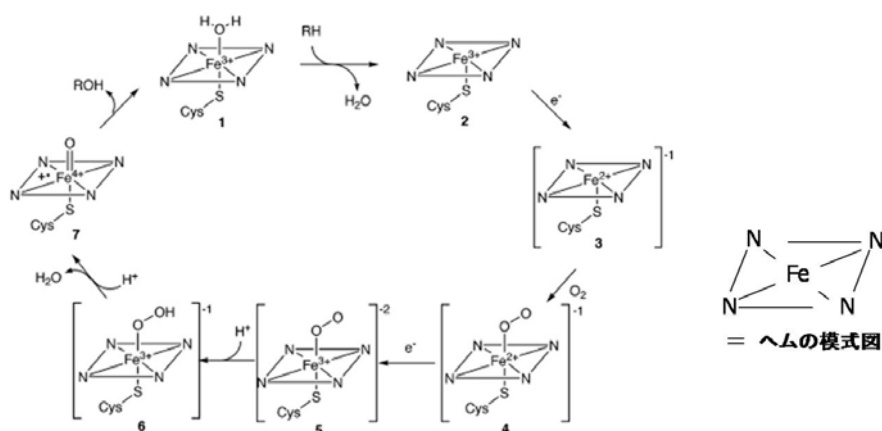
光合成システムは生物界の大部分のエネルギー生産を行っている重要なエネルギー工場である。図 7 は、Deisenhofer らによって初めて明らかにされた紅色光合成細菌の光合成反応中心タンパク質の構造である。様々な色素分子を持ち、高度に組織化されている点で呼吸鎖と同じ

である。2つのクロロフィルがスペシャルペアと呼ばれる2量体を構成している。これが集光を受け持つアンテナタンパク質からの光エネルギーを受けて励起し、電荷分離反応の基点となる。色素群はほぼ2回対称的に配置されているが、電子は図の向かって右側の経路のみを移動する。電子はスペシャルペアからサテライト・クロロフィルを経てフェオフィチンに移り、さらにその下のキノンに移動する。光合成反応中心は光のエネルギーを高い効率で化学エネルギーに変換する初期過程を担う精密分子である。



第7図：紅色細菌の光合成反応中心タンパク質の構造(左)とM, L, H鎖中の色素(右) [1]<sup>37</sup>

(左)紅色細菌 *Rhodospseudomonas viridis* の光合成反応中心タンパク質。太陽光からのエネルギーによって電荷分離を起こし、太陽電池のような働きをする。シトクロム鎖 Cyt, M鎖, L鎖, H鎖と呼ばれる4つのサブユニットからできていて、スペースフィルモデルで表された多数のヘテロ分子を持っている。(右) M, L, Hサブユニットの色素, P: スペシャルペア (バクテリオクロフィル2量体), B: サテライト・バクテリオクロフィル, H: バクテリオフェオフィチン, QA: メナキノン, QB: ユビキノン。Pが励起状態に上がると電子が飛び出し、0.65 psで右側のBへ移動, そこから3.5 psでHへ, 200 psでQAに, 100 μsでQBに移動する。



第8図：ヘムタンパク質P450の1原子付加反応サイクル。

<sup>37</sup> J. Deisenhofer, O. Epp, K. Miki, R. Huber, H. Michel, J. Mol. Biol., 180 (1984) p. 385.

酵素は化学反応を生体温度で効率よく進行させる触媒である。酵素反応では基質と酵素が鍵と鍵穴の関係にある。形状の整合性は酵素反応の始まりを左右するが、化学反応の第1ステップにすぎない。形が似ていても酵素反応を停めてしまうのが阻害剤である。酵素反応が促進されるか抑制されるのか、またその効率ほどの程度かを解析するためには化学反応シミュレーションが必要である。酵素タンパク質は基質、水分子などを取り込んで特有の反応を進める。図8はヘムタンパク質 P450 が行う 1 原子付加反応のサイクルを示している。全体としてみれば



という反応であるが、実に様々な反応中間体を經由していることが分かる。これはわずかな活性化エネルギーで効率よく反応を進めるための生体の基本戦略である。

タンパク質反応シミュレーションにとっては反応そのものが的確に記述できることが大切である。電子移動や化学反応は電子が関わるため記述には量子化学計算が欠かせない。kcal/mol の活性化エネルギーで進行するタンパク質の反応を定量的にシミュレーションするためには、電子相関効果の取り込みが欠かせない。数千～数万原子からなる巨大分子の量子化学計算と高い計算精度という、相反する要求を両立させる方法として、本稿ではタンパク質の密度汎関数法を解説してきた。その最終的な目的はタンパク質反応の定量的な解析である。

以下に2つのターゲットについて簡単に説明しよう。

## 電子移動シミュレーション

電子移動とはある2つの系の間で電子が移動する現象である。豊富で詳細な実験をもとに、Marcus らの研究<sup>38</sup>に端を発し、電子移動独自の理論体系が編み出されている。時間が関与する反応の記述には、時間に依存する Schrödinger 方程式の解法が必要である。その近似解法として摂動法により、電子移動反応速度  $k_{ET}$  は、電子が単位時間あたりに始状態から終状態へと遷移する確率として表現される。これは Fermi の黄金律として知られており、

$$k_{ET} = \frac{2\pi}{\hbar} |V|^2 (FC) \quad (6.2)$$

と書かれる。ここで、 $\hbar$  は Planck の定数を  $2\pi$  で割った定数、 $V$  は電子カップリング行列要素と呼ばれ、電子の波動関数から計算される項である。 $(FC)$  は Frank-Condon 因子と呼ばれ、原子核の波動関数などから計算される。

反応速度が非常に速い電子移動反応では Born-Oppenheimer 近似よりもさらに近似を進めた原子核の波動関数と電子波動関数を切り離す Condon 近似を採用することによって、(6.2) 式のように電子移動反応速度  $k_{ET}$  は  $|V|^2$  と  $(FC)$  の積で表されるため、速度因子を別々に議論することができる。タンパク質の量子計算による本格的な  $(FC)$  の見積もりはしばらく待たねばならないだろうが、古典的分子動力学計算による知見は集まりつつある。上記の光合成反応中心タンパク質の  $(\text{P}^*-\text{H})$  から  $(\text{P}^+-\text{H})$  への電荷分離反応において、右側周りの経路では最高速に移動できる配置を持っており、左側周り経路ではこの条件から外れているらしいこと。さらに、遅く見積もられすぎる  $(\text{P}^+-\text{H}) \rightarrow (\text{P}-\text{H})$  の電荷再結合反応速度が、構造変化を加味すれば実験値に近づくといった報告<sup>39</sup>がある。シミュレーションには仮定が多いが、タンパク質の構造や構造変

<sup>38</sup> R. A. Marcus, N. Sutin, Biochim. Biophys. Acta, 811 (1985) p.265.

<sup>39</sup> M. Massimo, J. N. Gehlen, D. Chandler, M. D. Newton, J. Am. Chem. Soc., 115 (1993)

化に基づく因子の究明が可能になりつつある。

タンパク質の電子移動反応の電子による因子  $V$  に関しては、これまでのところタンパク質波動関数を直接用いた計算は行われていない。とはいえタンパク質の密度汎関数計算から精密な  $V$  の値が見積もられるようになるのはもうすぐである。様々な計算法の諸考察については、これを先導的に理論構築してきた Newton の総説<sup>40</sup>を参考にするとよい。

電子波動関数による直接計算法の考え方は簡単である。光合成反応中心タンパク質の光励起による電子移動過程の電子カップリング要素  $V$  は  $z$  方向の振動電場  $F\sin(\omega t)$  と電子の双極子モーメント  $(-e\mathbf{r})$  ( $\mathbf{r} = (x, y, z)$ ) との相互作用から以下と表される。

$$V = \int \Psi_f \frac{zF}{2} \Psi_i d\mathbf{r} \quad (6.3)$$

すなわち原理的に始状態波動関数  $\Psi_i(P^* - H)$  と終状態波動関数  $\Psi_f(P^+ - H)$  の電子波動関数が精度良く見積もられればよいことがわかる。これらの波動関数の計算に配置間相互作用法を用い、電子カップリング要素を計算するのが分子軌道計算の正攻法である。

現在のところ、配置間相互作用計算法の拡張である SAC-CI 法を用いた長谷川・中辻らの計算<sup>41</sup>が先駆的な例である。光合成反応中心タンパク質における、クロロフィル 2 量体-フェオフィチン、続くフェオフィチン-キノン間における色素間で電子カップリング行列要素を計算し、光合成色素の励起構造に関する有用な知見を与えた。ところで、タンパク質の軌道は図 1 のシトクロム  $c$  のように、主に色素の成分を持つ軌道でもタンパク質全域に広がる。タンパク質電子移動の解明にはこの広がりを考慮に入れることが重要である。タンパク質の励起波動関数計算では、1 電子励起の記述が有効な密度汎関数法-配置間相互作用 (DFT-CIS) 法や Runge-Gross の定理から発展した時間依存密度汎関数 (TDDFT) 法が有力な候補である。タンパク質の軌道を考慮に入れた計算法は、光合成反応中心タンパク質における  $(P^*-H)$  から  $(P^+-H)$  への電荷分離反応がサテライト・クロロフィルを経由する理由の解明に欠かせないだろう。

しかし、タンパク質のすべての軌道における励起を考えると 1 電子励起ですら、ハミルトニアン行列の次元は天文学的な数になってしまう。タンパク質を、例えばクロロフィル 2 量体を含む空間やフェオフィチンを囲む空間などのいくつかの軌道空間に分割する必要があるだろう。これらの空間の波動関数の記述には 4 章で説明した QClO を用いるとよい。QClO は分子のある部分に局在し、その部分のカノニカル分子軌道の性格を持つ直交局在化軌道であった。このように分割しておくとならば必要な軌道の数を絞ることができ、計算量が大幅に減る。空間内の CIS 計算をすれば励起エネルギーや振動子強度が求まり、QClO の直交性を利用して各空間の間の励起計算をすれば電子移動の確率を求めることができる。このような手法を用いて、近い将来、本格的なタンパク質電子移動シミュレーションが行われるようになるだろう。

## 化学反応シミュレーション

化学反応は、物質がそれ自身か他の物質と作用し合って、元の物質とは異なる物質を作り出す現象である。分子の構造と作用する過程で理論づけられる。そのため、タンパク質の化学反

---

p. 4178.

<sup>40</sup> M. D. Newton, Chem. Rev., 91 (1991) p. 767.

<sup>41</sup> J. Hasegawa, H. Nakatsuji, J. Phys. Chem., B102 (1998) p. 10420.

応シミュレーションには、溶媒、温度の効果などを考慮に入れながら、化学結合の組み換えなどの分子の電子状態変化を精度よく追跡できる計算方法が必要になる。1964年にHohenbergとKohnが密度汎関数原理を発表した<sup>9</sup>が、これは温度0における理論である。その翌年、Merminが有限温度への拡張を発表した<sup>42</sup>。これにより、密度汎関数法—分子動力学(DFT-MD)法、密度汎関数法—モンテカルロ(DFT-MC)法が有力な計算方法となることが保証された。

その目的は、反応系から遷移状態を通して生成系に至る反応経路(反応座標)の解析である。どのように分子が構造変化し、どのように結合が組み変わるのかを追跡する反応機構モデル解析の出発点は、遷移状態の存在を仮定した遷移状態理論である。化学反応速度の差異は活性化自由エネルギー $\Delta F$ を通して現れるため $\Delta F$ が最も重要な物理量である。 $\Delta F$ の正確な見積もりには、電子相関を取り込んだ密度汎関数法が役立つ。

タンパク質反応の $\Delta F$ および反応経路の求め方の基本的な原理は一般的な化学分子の方法と同じである。電子移動反応に比べて格段に遅い化学反応にはBorn-Oppenheimer近似が適用限界である。この近似では分子構造が変化するとすぐに電子状態が追従する。言い換えれば、電子状態は常に分子構造の明示的な関数として表してよい。そのため、様々な分子構造でエネルギーを計算することによって、最も起こりやすい軌跡と登るべき山の高さを決定すればよい。

化学反応に関わる原子の数は反応系を構成する全原子である。この事情はタンパク質においても変わらない。酵素が基質と結合する際、その前後で結合部位の形状は変化しタンパク質全体に影響をおよぼす。構造の自由度を底面の軸にして、ポテンシャル・エネルギーを縦軸で示した俯瞰図で考えると、地形には複数の谷が存在する。すなわちこのエネルギーの極小点における原子核座標を持つ分子の構造が安定構造である。これらは化学反応の反応系と生成系のエネルギー状態の候補である。反応が起きるためには谷から谷へと分子が渡る必要がある。最も省エネルギーな行き方は尾根の峠を通る道筋である。峠の辺りは馬の鞍の形をしているため鞍部あるいは鞍点と呼ばれる。この鞍点が遷移状態である。したがって、反応系の安定点と鞍点のエネルギー差が活性化自由エネルギーである。これが低いほど反応速度は速い。

反応系、生成系、そして遷移状態はエネルギーの原子核座標による1次微分(グラジエント)が零となる点である。鞍点を極小点、極大点から識別するためには、2次微分の値を計算して判定を行う。エネルギーの2次微分行列はヘッシアンと呼ばれ、これを対角化すれば力の定数を得る。安定構造すなわち極小点は全ての固有値が正のもので、遷移状態すなわち鞍点は負の固有値を1つだけ持つものである。負の固有値の固有ベクトルは反応経路の方向を向いており、この方向のみポテンシャル・エネルギー超曲面は降っていて、他の方向は全て登っている。

以上のように、原理的には反応系および生成系の安定構造と遷移状態構造を求め、活性化自由エネルギー $\Delta F$ を見積もることができれば、化学反応シミュレーションの大半の目的が達成される。さらに、反応系から遷移状態、生成系に至る分子構造と作用の変遷、すなわち反応経路を追跡できれば、その化学反応機構が十分に解析できたことになる。その代表がIRC・DRC法<sup>43,44</sup>で、量子化学計算で化学反応経路を求める標準的な手法である。これは、反応系・生成系の安定構造や遷移状態が求められた場合、それぞれの構造が一連の反応に対応しているかを調べる手段である。タンパク質の化学反応も原理は同じである。

<sup>42</sup> N. D. Mermin, Phys. Rev., A137 (1965) p.1441.

<sup>43</sup> C. Gonzalez, H. B. Schlegel, J. Chem. Phys., 90 (1989) p.2154.

<sup>44</sup> K. M. Dieter, J. P. Stewart, J. Mol. Struct. (Theochem), 163 (1988) p.143.

IRC 法や DRC 法は、反応系—遷移状態—生成系の素反応において、作用最小の法則にしたがう反応経路を探す試みであり、これらは理想的な反応経路なので重要である。しかし、実際の反応は必ずしもこのような作用最小の経路を進むわけではない。他の峠も迂回して反応が進む確率は決して小さくない。しかも、そもそも安定構造が 1 つに決まる保障すらない。

このような性質を持つタンパク質の化学反応解析を行うためには、分子動力学法で温度や初期構造などの計算条件を変えてタンパク質の運動を追いかけること、モンテカルロ法で様々な条件によるカノニカルアンサンブルを計算することが直接的な方法である。これらの方法をより広い反応空間すなわち位相空間に適用できれば、大域的な反応座標探索や熱力学量の統計平均を得ることが期待できる。すなわち、究極の化学反応シミュレーションは様々な熱統計力学サンプリングによる自由エネルギー計算に他ならない。

自由エネルギーの差は、

$$F(\lambda_j) - F(\lambda_i) = \int_{\lambda_i}^{\lambda_j} d\lambda \left\langle \frac{\partial E^{(\lambda)}}{\partial \lambda} \right\rangle_{NVT} \quad (6.4)$$

と計算される。ここで、 $\lambda$  は反応経路を示すパラメタで、被積分部分は Kohn-Sham エネルギーの  $\lambda$  に対する偏微分係数のカノニカル平均である。古典系のエネルギーを用いて精力的に研究が進められてきた自由エネルギー計算法の量子版である。系こそまだ小さいが、密度汎関数法においても様々なシミュレーションが行われ、良好な結果が得られつつある<sup>45</sup>。

大きな系が取り扱える古典系では、反応座標の周辺を効果的にサンプリングする様々な方法が次々と開発されている。そのほとんどが再重み付け (reweighting) 技術を応用したもので、バイアスをかけたり、束縛条件を与えたりして、サンプリングする位相空間の範囲を狭めている。量子計算に基づくタンパク質のシミュレーションにも適用可能なものが多い。

また、反応時間の短い化学反応は、効果的に位相空間の探索を狭めることができるためシミュレーション達成は近いだろう。代表的な例はロドプシンの励起状態ダイナミクスである。ロドプシンは色素レチナールがオプシントタンパク質 (約 350 残基) とシッフ塩基結合した光受容膜タンパク質で、光によって 11-*cis*-レチナールから all-*trans*-レチナールに異性化する。この反応は、溶液中よりもオプシン中の方が高速で、わずか数百 fs といった時間スケールで反応が進む。その理由や反応機構は、近い将来、タンパク質量子化学計算で精度良く解析することができるだろう。

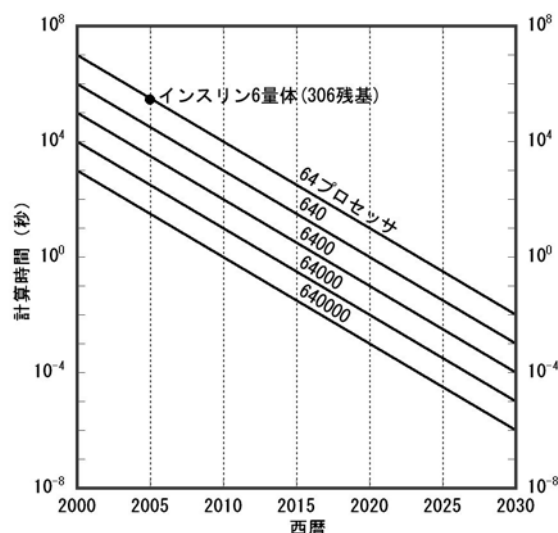
## まとめ

図 9 はタンパク質密度汎関数計算の計算時間予測グラフである。図の横軸は西暦年で、2000 年から 2030 年までふつである。縦軸はタンパク質 1 構造の計算時間を秒単位の対数軸で表したものである。Moore の経験則が維持され、アルゴリズムとソフトウェアの改良によって 10 年で 10 倍高速化するという効果を加えると、計算は 10 年で 1000 倍速くなる。さらに、計算機の超並列化によっても計算時間が短縮される。このグラフを見ると、2030 年には 64 万台 (コア数でもよい) の超並列計算機を使用して、300 残基タンパク質の固定構造の計算が  $10^6$  秒で計算できる。これは 2000 年における古典力場による計算時間とほとんど変わらない。

タンパク質密度汎関数法による本格的な反応シミュレーションはこれからの研究課題である

<sup>45</sup> B. L. Trout, M. Parrinello, J. Phys. Chem., B, 103 (1999) p. 7340.

が、2030年までにはタンパク質の励起や運動が確実に追跡できるようになっているはずである。タンパク質の本格的な励起状態や電子移動の計算は2010年ごろまでにはスーパーコンピュータで達成され、2020年には個人のコンピュータで難なく実行できるだろう。同様に、タンパク質の化学反応のシミュレーションも2020年までには本格的なものが登場し、2030年にはもはや普通に行われるようになっていくに違いない。



第9図:密度汎関数法によるタンパク質の計算時間予測[1]。

プロセッサの台数および計算時間は2005年のインスリン6量体の計算をほぼ基準とし、それぞれの斜め線は $10^{-3}/10$ 年の傾きで引かれている。 $10^{-2}$ はプロセッサ高速化のMooreの経験則によるもの、 $10^{-1}$ はアルゴリズムやソフトウェア技術の向上によるものである。

タンパク質の励起エネルギー、振動子強度の計算や自由エネルギー計算の結果は、実験結果と直接比較することができる。密度汎関数の原理によりさらに高品位な交換相関汎関数が見出され、それをを用いた計算結果が実験値に接近するという好ましいサイクルが生まれるだろう。様々なタンパク質についてのシミュレーションが蓄積されれば、タンパク質の反応を制御している要因を電子レベルで分析することが可能になるだろう。生命の分子機構を解明し理解することは人類最大の科学的テーマの一つである。

光合成反応中心や光異性化反応の機構が完全に理解されれば、これらを基にした太陽電池、光センサー、非線形光学材料の設計へと応用されることになるだろう。天然や人工光合成システムから効率的に電気エネルギーを獲得し、超伝導エネルギー貯蔵装置に大電力を備蓄できるというような時代が来るかもしれない。

タンパク質の化学反応機構の解析は創薬の分野で多くの注目が集まっている。例えば、ヘムタンパク質のP450 (CYP)で代謝分解できない薬剤は体内に残り続けるため、薬剤の設計にはP450との相互作用解析が欠かせない。代謝産物が、DNAや細胞組織に作用して、癌や壊死を引き起こす問題にも関係する。タンパク質の化学反応シミュレーションにより、基質の結合性や活性化エネルギー、反応速度などが予測できる。系統的にシミュレーションを行うことによって知見を蓄積し、基質特異性や反応特性を制御できるようになるだろう。

生体触媒は工業製品の生産にも利用されつつある。これまでの工業触媒では重金属が多く使



われており、毒性のある副産物が問題になる。これに対して、P450には過酸化水素を酸化剤とし副産物は水だけというような特性がある。このように、生体が行う触媒反応はクリーンな反応を推進する。生体触媒の設計にもタンパク質の化学反応シミュレーションが威力を発揮するだろう。

低分子ではすでに実績のある量子化学シミュレーションの信頼性は高く、反応予測、反応設計ができるレベルに達している。密度汎関数法によるタンパク質反応シミュレーションが実用的になれば、基礎科学のみならず様々な応用分野でも大きな貢献をすることだろう。

本稿で使用した一部のデータは、文科省のいくつかのプロジェクト(革新的シミュレーションソフトウェアの研究開発、次世代生命体統合シミュレーションの研究開発、イノベーション基盤シミュレーションソフトウェアの研究開発)で得られた成果を利用した。また、一部の計算には統計数理研究所の統計科学スーパーコンピュータシステムを利用した。

## 参 考 文 献

最後にさらに詳しい内容に興味を持たれた方のために、関連する著書を数遍リストアップした。

### 生体分子量子化学計算関連

[1] 佐藤文俊, 恒川直樹, 吉廣保, 平野敏行, 井原直樹, 柏木浩:”タンパク質密度汎関数法”, 柏木浩監修, 森北出版, (2008). 【注: 本稿は本拙書の抄録的位置づけである。】

[2] 上村(西野)典子, 佐藤文俊, 恒川直樹, 西村康幸, 平野敏行, 吉廣保, 甘利真司, 加藤昭史, 小林将人, 田中成典, 中野達也, 福澤薫, 望月祐志: “プログラムで実践する 生体分子量子化学計算 –ProteinDF/ABINIT-MP の基礎と応用”, 佐藤文俊, 中野達也, 望月祐志編, 森北出版, (2008).

### 密度汎関数法

[3] R. G. Parr, W. Yang 著, 狩野覚, 吉田元二, 関元訳: “原子・分子の密度汎関数法”, シュプリンガー・フェアラーク (1996).

[4] 里子允敏, 大西楯平: “密度汎関数法とその応用—分子・クラスターの電子状態”, 講談社 (1994).

### 量子化学計算・分子軌道法全般

[5] 藤永茂: “分子軌道法”, 岩波書店 (1980).

[6] A. Szabo, N. S. Ostlund 著, 大野公男, 阪井健男, 望月祐志訳: “新しい量子化学—電子構造の理論入門<上・下>”, 東京大学出版会 (1987).

[7] T. Helgaker, P. Jorgensen, J. Olsen: “Molecular Electronic-Structure Theory”, John Wiley & Sons, (2000).