

1B16

Elongation 法による DNA の高効率・高精度電子状態計算とその物性解析

(九大院・総理工¹, SCNU², JST-CREST³) 折本 裕一¹, Xie Peng¹, Liu Kai¹,
Gu Feng Long^{2,3}, 青木 百合子^{1,3}

High-efficient DNA electronic structure analysis by *ab initio* elongation method

(Kyushu Univ.¹, SCNU², JST-CREST³) Yuuichi Orimoto¹, Peng Xie¹, Kai Liu¹,
Feng Long Gu^{2,3}, Yuriko Aoki^{1,3}

【序】 DNA をはじめとする生体高分子の物性を理解するには、分子構造と電子状態の関係を正確に把握する必要がある。経験的パラメータを含まない量子化学計算をこの問題に適用できれば、定量的議論が可能となり、生命科学や生体材料分野へ大きな貢献が期待できる。しかし、従来の量子化学計算では系のサイズを N としたときに N^4 に比例する計算コストのため(サイズとともに爆発的に増大)、計算機技術の発達した現在においても生体高分子のような複雑・巨大系を気軽に計算できる段階にはなく、現実的には系の一部をモデルとして扱い、それ以外の部位の影響については目をつぶらざるを得ない。

当研究グループで開発している大規模系演算手法である 3D-Elongation(3D-ELG)法は、生体分子のような三次元ランダム高分子の電子状態を化学的精度を保持しつつ、オーダー N (N^1 に比例した計算コスト)の高効率で得ることができる。本研究では、Elongation 法の巨大生体高分子への有効性を調べるため、様々なタイプの DNA について電子状態の詳細を調べ、二重らせんのピッチ、塩基配列、カウンターイオン等の構造因子と物性の関係を調べる。

【方法】 Elongation 法^{1),2)} は高分子の重合反応を計算機上で実現したものであり、高分子の反応末端にモノマーを付加させながら、逐次的に電子状態を得る手法である。図1に、DNA鎖を例とした Elongation 法による電子状態計算の手順を示した。まず、適当なサイズの出発クラスタについて電子状態計算を行い、正準分子軌道(CMO)を得る。次に、伸長方向の反応末端側に新しいモノマー M (ここでは塩基対ユニット)を付加させるが、従来法であればモノマーを含めた全系を解き直さなければならない。Elongation 法では、先に得た系全体に広がった出発クラスタの CMO をまず、ユニタリー変換によって反応末端側の Active 領域(B 領域)と、逆側の Frozen 領域(A 領域)に局在化させた領域局在化軌道(RLMO)を得る。この時、Frozen(A)領域の RLMO とモノマーは相互作用しないほど離れているので、計算に含めなくても良い。その結果、Active(B)領域の RLMO とモノマーの相互作用だけを解くことで(部分空間に対する Fock 行列の対角化)、モノマーが加わった系全体の電子状態が得られる。この局在化と部分 Fock の対角化を繰り返し、その都度必要のない Frozen 領域の RLMO を次々と計算から外す。解くべき部分空間のサイズは一定となり、計算精度を落とすことなく効率的に巨大系の電子状態が得られる。

しかし上記だけでは、局在化させる前の原子軌道(AO)基底 Fock 行列は系全体のものを常に扱うため、AO Fock 行列の作成と RLMO Fock 行列への変換が系のサイズとともに増大してしまう。そこで、必要のない(軌道の tailing の及ばなくなった)基底関数(AO)そのものを計算から外してしまう AO Cutoff 法によって、この問題を解決している。これにより、Fock 行列の作成・変換に係る時間も系の伸長過程で一定となる。

Fock 作成・変換の次元が一定、及び、モノマー付加後に対角化する部分 Fock 行列の次元が一定となり、Elongation 法では計算時間が N^1 に比例したオーダー N を達成できる。

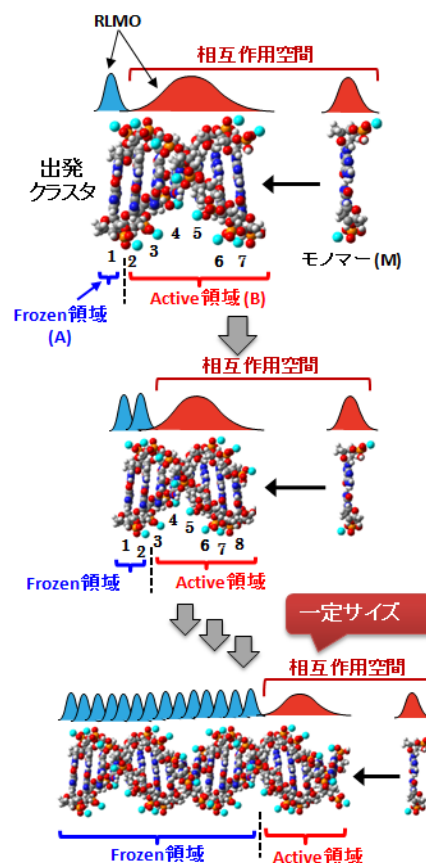


図1 Elongation 法による DNA の高精度・高効率電子状態計算

【結果と考察】 図2に、B-poly(dA)・poly(dT) DNA に対するベンチマークテストの結果を示す (HF/STO-3G)²⁾。まずは計算精度であるが、従来法(全系をダイレクトに解く場合)との全エネルギーの誤差(E(ELG)-E(CONV))は、塩基対ユニット数がn=20(原子数1,324)の時、わずかに 2.59×10^{-9} Hartree/atom(系全体で、0.002 kcal/molの誤差)であり、超高精度に電子状態が得られる。次に計算時間は、図に示すように各サイズごとのCPU時間が、従来法ではサイズとともに急激に増大するのに対し、Elongation法ではAO cutoffの開始とともに、各ステップにかかる時間がほぼ一定となる。

図2では塩基対(A...T)をスタックさせた場合を見てきたが、ここでさらに図3に示すように4種類の塩基対を様々な組み合わせで n_{whole} 個スタックさせることを考える。まず従来法では、塩基対 n_{whole} 個からなるDNAの全ての可能性(全組み合わせ数 $4^{n_{whole}}$ 通り)を計算するための時間 T_{CONV} は式(1)のように見込まれる。

$$T_{CONV} \sim (m \cdot n_{whole})^4 \cdot 4^{n_{whole}} \quad (1)$$

ここで、 m は塩基対ユニット当たりの基底関数の数であり、系のサイズに相当する $m \cdot n_{whole}$ の4乗に、組み合わせ数 $4^{n_{whole}}$ が掛け合わさった時間がかかる。

次にElongation法であるが、電子状態を逐次的に求めていくため、塩基対 n_{whole} 個の系を計算するためには $(n_{whole}-1)$ 個の結果が必要となり、さらに $(n_{whole}-1)$ 個のためには $(n_{whole}-2)$ 個の結果が必要というようにそれぞれ前段階の計算結果が必要となる。しかし一方で、途中段階の計算結果を全て保存しておけば、途中までの配列が同じ場合には、最初から計算する必要はなく、保存した結果を再利用して続きから計算を開始することができる。これらを考慮すると、Elongation法で塩基対 n_{whole} 個の全組み合わせ($4^{n_{whole}}$ 通り)を計算するための時間 T_{ELG} は式(2)のようになる。

$$T_{ELG} \sim (m \cdot n_{actv})^1 \cdot 4^1 + (m \cdot n_{actv})^1 \cdot 4^2 + \dots + (m \cdot n_{actv})^1 \cdot 4^{n_{whole}} = (m \cdot n_{actv})^1 \sum_{k=1}^{n_{whole}} 4^k$$

$$\sim \frac{4}{3} (m \cdot n_{actv})^1 \cdot 4^{n_{whole}} \quad (2)$$

ここで、 n_{actv} は図1の相互作用空間に含まれるユニット数であり、伸長過程で一定のサイズである。式展開の途中、等比数列の和の公式、および $4^{n_{whole}} \gg 1$ という条件を用いた。式(2)から分かるように、Elongation法の計算が目的サイズだけでなく、全てのスタックサイズそれぞれの全組み合わせの計算を含んでいるにもかかわらず、最終的には一定かつ小さな相互作用空間のサイズ $m \cdot n_{actv}$ の1乗に、目的サイズでの組み合わせ数 $4^{n_{whole}}$ が掛け合わさっただけの時間で済むことが分かる。具体的に $n_{whole}=100$ として計算時間の比を見積もると、DNAで想定される $m \sim 770$ (STO-3G), $n_{actv}=5$ を用いることで $T_{ELG}/T_{CONV} \sim 1/(6.8 \times 10^{15})$ となり、従来法とは比較にならないほど計算コストを削減できる。しかも、上述のように計算途中の各サイズの結果を既に全て獲得しており、鎖長依存性などの有用な情報をも抽出できる。

今後Elongation法によって、二重らせんのピッチ(A, B型など)、様々な塩基配列、カウンターイオン等が系の安定性や電子状態(状態密度など)とどのような関係を持っているか詳細に調べていく。

- 1) A. Imamura, Y. Aoki and K. Maekawa, *J. Chem. Phys.*, **95**, 5419 (1991).
- 2) Y. Aoki and F. L. Gu, *Phys. Chem. Chem. Phys.*, **14**, 7640 (2012).

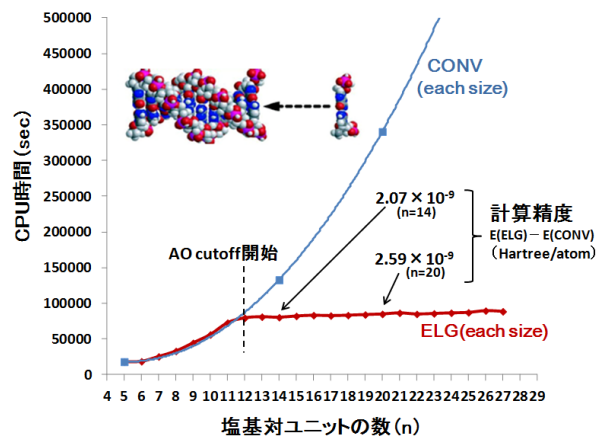


図2 DNAの計算時間の比較²⁾

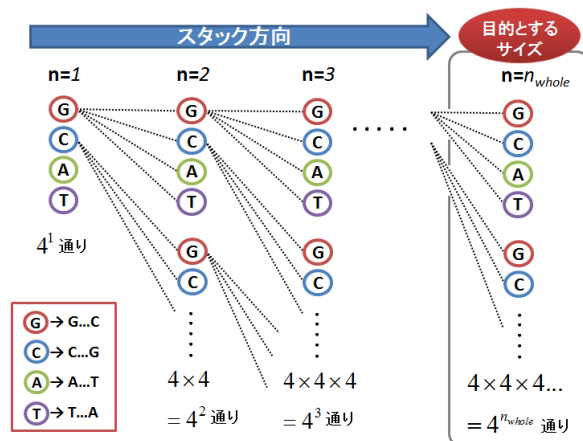


図3 塩基対ユニットを n_{whole} 個スタックさせる場合の全組み合わせ数