

DNA 計算における局所探索法による反応障壁近似計算

武田 勉[†] 小野 廣隆^{††} 定兼 邦彦^{††} 山下 雅史^{††}

[†]九州大学大学院システム情報科学府

^{††}九州大学大学院システム情報科学研究院

E-mail: [†]takeda@tclab.csce.kyushu-u.ac.jp, ^{††}{ono,sada,mak}@csce.kyushu-u.ac.jp

あらまし DNA 計算は DNA 分子の形態変化を計算規則とみなす計算パラダイムであり、望ましい形態変化を行う分子配列の設計（配列設計問題）が様々な角度から研究されている。本研究は配列設計において大きな役割を果たすといわれる、反応エネルギー障壁値を求める問題について考える。エネルギー障壁は、分子反応の（通常は分子長の指数オーダー分存在する）中間形態により定まるものであり、全探索による求解は現時的ではない。本研究では、局所探索法に基づく高速な近似解法を提案する。計算機実験による既存の厳密解法との比較では、多くの配列例で本近似解法により高速に比較的良好な解を得られることが確認された。

キーワード DNA 計算, エネルギー障壁, MH 法, 局所探索法

A Local Search Algorithm for Barrier Heights Estimation in DNA computing

Tsutomu TAKEDA[†], Hiroataka ONO^{††}, Kunihiko SADAKANE^{††}, and Masafumi YAMASHITA^{††}

[†] Dept. of Computer Science and Communication Engineering, Kyushu University

^{††} Dept. of Computer Science and Communication Engineering, Kyushu University

E-mail: [†]takeda@tclab.csce.kyushu-u.ac.jp, ^{††}{ono,sada,mak}@csce.kyushu-u.ac.jp

Abstract DNA computing is a computing paradigm utilizing transitions of DNA molecules as computation. In the context, sequence design problem has been studied to design DNA molecules whose transitions are suitable for computation. In this paper, we consider the problem of estimating the height of the energy barrier that plays an important role in the sequence design. The barrier height depends on intermediate structures in all transition paths. Unfortunately, the number of intermediate structures is exponential of the sequence length, and the exhaustive search is not practical. We propose a fast approximation algorithm based on the local search. Computational experiment shows that our algorithm finds relatively good solutions as the existing exact algorithm in short time.

Key words DNA computing, Barrier Height, Morgan and higgs' Heuristics, Local Search

1. 背景

現在の計算機の計算能力向上にはマイクロ化が不可欠であるが、マイクロ化にも限界が存在することが指摘されている。そこで、近年、生体分子の組み換え規則を利用することで計算速度、エネルギー効率、情報格納量の点で革新的な向上を可能とする、「分子計算」なる新たな計算パラダイムが注目されている。分子計算の中でも、DNA 分子を扱うものを DNA 計算と呼ぶ。

DNA 計算の原理は、DNA 分子がワトソン・クリック相補性に基づいて選択的に水素結合する形態変化を計算と見立てることにある。1994 年、Adleman はこの特徴を利用して有向ハミルトンパス問題を解くことに成功した [1]。

一般に DNA 分子は単に一つの形態をとるだけでなく、複数の形態をとりつつ、それらの形態間を遷移する。このため、この形態変化を「計算」として利用する DNA 計算の設計において、DNA 分子の形態予測（フォールディング問題）、逆フォールディング問題などが重要性を増している。

DNA の塩基配列の形態を予測する方法は古くから研究されている。特に、動的計画法を用いて最小の自由エネルギーを持つ形態を求めるアルゴリズムは生物学実験において幅広く用いられている [12]。一般に、与えられた配列を持つ分子が取り得る最小の自由エネルギーを持つ形態を予測する問題をフォールディング問題 (folding problem) という。

一方、フォールディング問題とは逆に、DNA 反応を計算と

して利用するためには、あらかじめ与えられた形態にフォールドするような塩基配列を求める必要がある。この問題は逆フォールディング問題 (inverse folding problem) と呼ばれ、フォールディングに対する最適化問題として定式化することができる [5]。

実際の DNA の形態変化の仕組みの解析は盛んに研究されているものの、未解明な要素が多く、DNA 形態変化パスの予測・設計は困難である。このため、多くの場合、形態変化に相当する化学変化の反応速度に大きな影響を及ぼす反応エネルギー障壁を形態変化の評価尺度とする手法がある [3], [4], [6]~[10]。これは、形態間のエネルギー障壁が高いほど、形態間の変化は容易でなく、逆に、エネルギー障壁が低いほど、変化は容易に起きると見なすことにより、形態変化の起こり易さを測定するものである。

本研究では分子の初期形態・最終形態からエネルギー障壁を現実的な時間で求めることを考える。二つの形態間のエネルギー障壁を求めるには、その間のすべての形態変化パスを網羅的に探索する必要がある。しかし、DNA 分子のとりのう形態は分子長の指数オーダー存在し [2]、計算時間は現実的ではない。そこで、本研究では局所探索法に基づく高速な近似解法を提案する。

2. 準備

DNA 分子は、糖・リン酸基・4種の塩基からなるヌクレオチドが1本鎖状に結合してできる生体高分子である。塩基のアデニン (A) とチミン (T)、シトシン (C) とグアニン (G) は選択的に水素結合する性質を持ち、これ以外の組み合わせでは結合しない。この組み合わせの原理は、ワトソン・クリック相補性と呼ばれる。この規則に基づき、塩基配列 X は複数の形態を取りうる。DNA の形態は塩基配列中の各塩基に1から順に番号を付けたとき、 i, j という形の塩基対の集合と定義される。 i, j は i 番目の塩基と j 番目の塩基からなる対を表す。

DNA 分子は自由エネルギーを持っており、その値は DNA の塩基配列、形態により異なる。各形態のエネルギーは、その形態を構成する各ループ構造のエネルギー (経験的に得られた自由エネルギーのパラメータ) の和によって近似される。ここで、ループ構造とは図1にあるような構造のことである。実線は塩基と塩基を結ぶバックボーン (主鎖)、点線は塩基対を作る水素結合を表しており、塩基対で囲まれた領域をループと呼ぶ。本研究では、ウィーン大学により開発され配布されている Vienna Package によりエネルギー計算を行っている^(注1)。

塩基配列 X に対して取りうる形態の集合を $S(X)$ とする。 $s_0, s_n \in S(X)$ なる2つの形態 s_0, s_n が与えられた時に、 s_0 から s_n までの形態変化パス p は以下で表される。

$$p = \langle s_0, s_1, \dots, s_n \rangle$$

ただし、各 s_i から s_{i+1} への形態変化は、ある1つの塩基対の形成、または解離である。この形態変化を1stepとする。パス長 (形態変化 step 数) が最短のパスをダイレクトパスと呼ぶ。

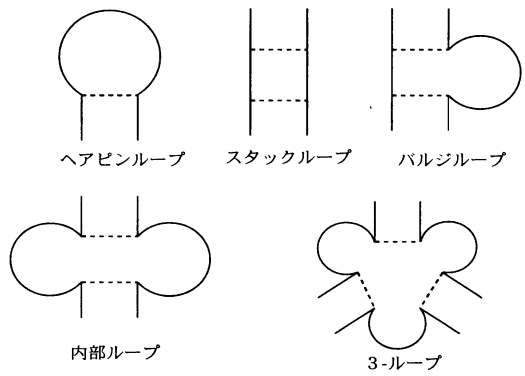


図1 形態を形成するループ構造

ある形態 s のエネルギーを $E(s)$ 、 s_0 から s_n までのパスの集合を $P(s_0, s_n)$ とする。 s_0 から s_n の形態変化におけるエネルギー障壁は以下で定義される。

$$E_{bar}(P(s_0, s_n)) = \min\{E_{top}(p) \mid p \in P(s_0, s_n)\},$$

$$E_{top}(p) = \max\{E(s_i) \mid s_i \in p(i = 0, \dots, n)\}.$$

一般にエネルギー計算の仕組みは複雑で、 $E_{bar}(P(s_0, s_n))$ の計算は困難である。また、 s_0 から s_n までのダイレクトパスの集合を $P_{direct}(s_0, s_n)$ とすると以下の式が成立する。

$$E_{bar}(P(s_0, s_n)) \leq E_{bar}(P_{direct}(s_0, s_n)).$$

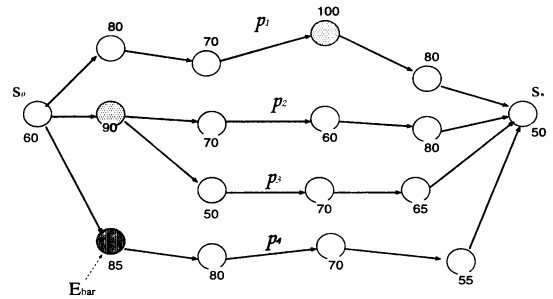


図2 形態変化パスとエネルギー障壁

図2はエネルギー60の初期形態 s_0 とエネルギー50の最終形態の形態変化パスである。図中の円は形態、円に付された数値はその形態の持つエネルギーを表している。形態変化パス p_1, p_2, p_3, p_4 のそれぞれの E_{top} は、 $E_{top}(p_1) = 100, E_{top}(p_2) = E_{top}(p_3) = 90, E_{top}(p_4) = 85$ であり、エネルギー障壁 E_{bar} は $E_{bar} = \min\{E_{top}(p) \mid p \in P(s_0, s_n)\}$ より $E_{bar} = 85$ となる。

3. エネルギー障壁見積もり問題

エネルギー障壁見積もり問題は、文字配列 X 、初期形態 s_0 、最終形態 s_n 、を入力として与えたとき、 $E_{top}(p)$ を最小にするパス p を求める問題である。これを求めるアプローチとして MH 法 [8] や最小流域法 [6] などが提案されている。

3.1 Morgan-Higgs 法 (MH 法) ([8])

MH 法は、塩基対の個数に着目し、塩基対の数をできるだけ多く維持してダイレクトパスを形成するヒューリスティクスで

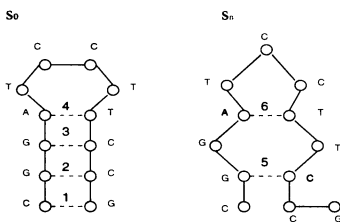
(注1) : <http://www.tbi.univie.ac.at/~ivo/RNA>

ある s_0 と s_n に存在している塩基対のみを塩基対として認め、 s_n の塩基対を形成するために解離しなければならない非互換の塩基対 (incompatible な塩基対) を s_0 から除いて、その塩基対を形成するという手順を繰り返す。

MH 法は計算効率が非常に優れているという利点があるが、MH 法を用いて得られたエネルギー障壁値は必ずしも正しいとは限らず、その精度にもばらつきがある。

MH 法のアルゴリズム

1. incompatible な s_0 の塩基対の数が最も少ない塩基対 (複数あるときはランダムに選択) を s_n から 1 つ選択。
2. 1. で選択した塩基対を s_0 から除き、可能な限り s_n の塩基対 (複数あるときは順序はランダム) を加え、これを形態 s'_0 とする。
3. s'_0 が s_n に一致しない限り、 s'_0 に対して、同様の手順を繰り返す。



s_n の塩基対	incompatible な s_0 の塩基対
5	2, 3
6	4

図 3 incompatible な塩基対

図 3 は塩基配列 $X = \text{CGGATCCTTCCG}$ 、形態 s_0, s_n が与えられたとき、それぞれの塩基対に番号を付し、 s_n に対する s_0 の incompatible な塩基対を表している。

3.2 最小流域法 ([6])

一般に DNA 分子は、様々な経路で形態が変化する。このため形態変化経路は、非常に多くの組み合わせが存在する。様々な経路を解析するために、DNA 配列が与えられた時の配列が取り得る全ての形態を考える (これを形態空間と呼ぶ)。形態空間の各形態にエネルギーを与え、形態間の遷移関係を定義したとき、エネルギー地形というグラフが形成される。エネルギー地形を計算する手法として Flamm たちによる Flooding 法 [9] がある。これにより真のエネルギー障壁値を求めることが可能であるが、実際の生物実験で用いられることのある 100 塩基長を越える配列に対しても全形態とそのエネルギーを求めることは困難である。これに対し、久保田たちにより形態空間上で初期形態だけからなる集合を隣接関係を用いて展開し、その集合が最終形態を包括した時点で終了する最小流域法が提案されている。

最小流域法

1. 始点 s_0 、終点 s_n は与えられている。
2. $\beta, N = \emptyset$. β : 流域, 到達しうる低エネルギー形態の集合
 N : β に隣接する形態の集合
3. 始点 s_0 を集合 β に加える。
4. 3 で集合 β に加えられた集合を v とすると、 v に隣接する形態 $u \in N(v) \setminus (\beta \cup N)$ を N に加える。
5. 集合 N の最も低いエネルギー形態を集合 β に移す。
(複数ある場合は、それらの形態を全て β に加える.)
6. 5 で β に加える点が終点 s_n であつたら終了。
そうでなければ 4 へ戻る。

このアルゴリズムによって求められた集合 β は、始点 s_0 と終点 s_n を結ぶ最小の流域を表現している。 β に含まれる最大エネルギーの形態は真のエネルギー障壁となる。Flooding 法とは異なり必要最低限の形態しか展開しないことから、Flooding 法よりはるかに高速に解を求めることができるものの、エネルギー地形によっては計算量が爆発する可能性がある。

4. 局所探索法によるエネルギー障壁近似計算

MH 法は比較的良好な解を求めることが可能であるが、解精度の保証はない。また、最小流域法は正確な障壁値を見付け出す、問題によっては計算量が爆発する。そこで、本研究では局所探索法に基づくロバスタかつ高性能な近似解法の構築を目指す。

局所探索法において、実行可能解 x に少しの変形を加えることで得られる解集合 $N(x)$ を x の近傍、その変形操作を近傍操作と呼ぶ。局所探索法は適当な解 x から始め、近傍 $N(x)$ 内に改善解が存在しなくなるまで近傍内移動を反復する手法である。最終的に得られた解を局所最適解と呼ぶ。本研究では形態変化パス自体を解とみなし、局所探索法を以下の手順で行う。

$$\begin{aligned} \text{minimize} \quad & E_{top}(p) = \max\{E(s_i) \mid s_i \in p(i = 0, \dots, n)\} \\ \text{subject to} \quad & p \in P(s_0, s_n) \end{aligned}$$

1. 初期解 (形態変化パス) 生成。
2. 近傍 (形態変化パスの集合) 探索。
3. 改善解が無くなれば解を出力し終了。

以下の節では、1, 2 について議論する。

4.1 初期解生成

本研究では以下の 2 種類の方法により生成した初期解を利用する。ここで、解とは s_0 から s_n までの形態変化パスとする。

- 1: 初期解を MH 法により生成。
- 2: 初期解を Greedy 塩基対形成アルゴリズム + MH 法により生成。

一般に、DNA 分子は塩基対の数が多ほど安定であることが多い。Greedy 塩基対形成アルゴリズムはこの性質に基づき、エネルギー減少量が最大 (増加量が最小) の塩基対を順に形成する手法である。形成する塩基対は s_n に対し、incompatible でないもののみである。このアルゴリズムにより得られる形態は一般

には最終形態とは異なるので、MH 法を適用し、解を得る。以下に Greedy 塩基対形成アルゴリズムを記す。

Greedy 塩基対形成アルゴリズム

s : ある形態

$N(s, s_n)$: s に、 s_n に対して incompatible でない塩基対を 1 つ形成した形態の集合

1. $s := s_0$.
2. $N(s, s_n) = \phi$ なら終了。それ以外なら 3. へ。
3. $\min\{E(s') - E(s) \mid s' \in N(s, s_n)\}$ なる s' を s として 2. へ。

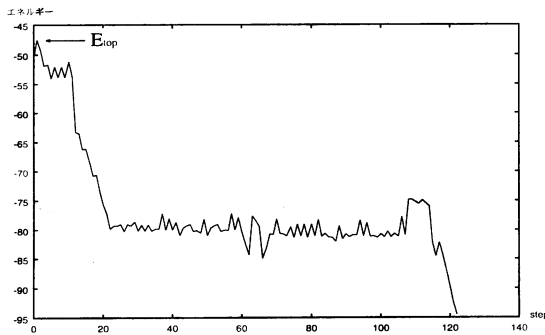


図 4 MH 法による初期解

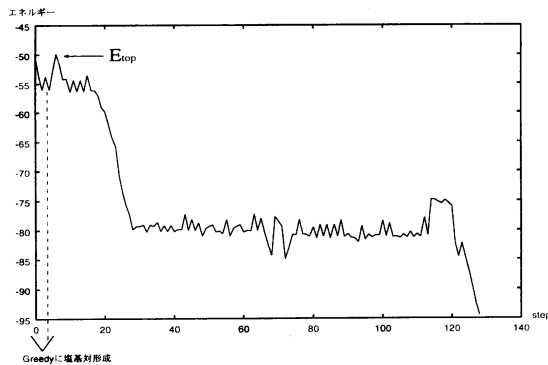


図 5 Greedy+MH 法による初期解

図 4, 5 は横軸が形態変化 step 数、縦軸がエネルギーを表す 2 種類の初期解生成の図である。図 5 のように Greedy に塩基対を形成後、MH 法を適用することで $E_{top}(p)$ の値が、図 4 の $E_{top}(p)$ の値より低くなる初期解を生成できる例である。

4.2 近傍と近傍探索

変化パス上の形態の一部を別の形態に変えたパスを近傍解とする。MH 法による s_i から s_j への変化パスを $MH(s_i, s_j)$ 、パス p 上の s_i から s_j への部分パスを $p[i, j]$ 、パス p 上の s_i を s'_i に変更して構成したパス p' を $p'(p, s_i) = \langle p[0, s_i-1], MH(s'_i, s_n) \rangle$ とする。パス p 上で最大エネルギー値をとる形態を s_h とし、以下のように探索近傍を定義する。

• 1-back 近傍

$$N(p)_1 = \{p'(p, s_h)\}.$$

• k-back 近傍

$$N(p)_k = \{p'(p, s_i) \mid h - k + 1 \leq i \leq h\}.$$

• Full-back 近傍

$$N(p)_{Full} = \{p'(p, s_i) \mid 1 \leq i \leq h\}.$$

本研究ではこれらを用いて 3 種類の近傍探索を行う。

• 1-back 近傍探索

常に 1-back 近傍を探索空間とし、改善解が無くなるまで、解の改善を繰り返す。計算時間は小さい。

• Full-back 近傍探索

常に Full-back 近傍を探索空間とし、改善解が無くなるまで、解の改善を繰り返す。計算時間はエネルギーが最大になる step h の値に比例し大きくなる。

• VD-back 近傍探索

はじめに 1-back 近傍を探索空間とし、改善解が無ければ近傍を徐々に拡大する。解の改善が行われれば、再び近傍探索は 1-back 近傍から行う。このような可変近傍 (variable depth neighborhood) 内での探索を繰り返す [11]。

VD-back 近傍探索

1. $k = 1$.
2. k -back 近傍内に改善解が存在すれば改善し、1 へ。無ければ 3 へ。
3. 近傍を $(k+1)$ -back 近傍に拡大し、2 へ。拡大できなければ終了。

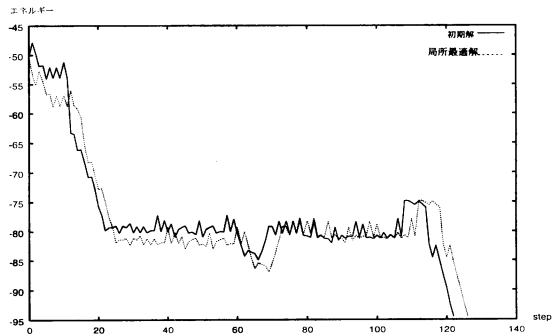


図 6 MH 法による初期解と Full-back 近傍探索による局所最適解

図 6 はインスタンス (塩基配列長: 202, 初期形態: 実験に使用されている形態, 最終形態: エネルギー最小形態) に対して MH 法による初期解と Full-back 近傍探索による局所最適解の例である。横軸が形態変化 step 数、縦軸がエネルギーを表している。初期形態のエネルギー -50.42 、最終形態のエネルギー -94.45 に対して MH 法が見つけた障壁値 (初期解) は -47.87 となっている。これに対し、局所探索法を適用した結果、 -52.73 なる値が発見された。

5. 実装と実験結果

本研究では以下の条件を組み合わせたインスタンスを用いる。

- ・配列長：46, 100, 158
- ・初期形態：実形態, ランダム形態
- ・最終形態：ランダム形態, エネルギー最小形態

ここでの実形態とは、逆フォールディングの配列設計による塩

インスタンス	配列長	初期形態	最終形態
46(1~5)	46	ランダム	エネルギー最小
46(6~7)	46	ランダム	ランダム
100(1~5)	100	ランダム	エネルギー最小
100(6)	100	最小(50+50)*	エネルギー最小
158(1~5)	158	ランダム	エネルギー最小
158(6~7)	158	実形態	エネルギー最小
158(8)	158	最小(79+79)*	エネルギー最小

表1 本研究で用いたインスタンス

基配列を利用した分子機械の動作実験(生物実験)に使用されたものである[7]。また、表1のインスタンス100(6), 158(8)の初期形態は、各塩基配列を半分に分割し、それらのエネルギー最小形態を組み合わせで作成した形態である。これらのインスタンスに対して、初期解生成(MH法, Greedy塩基対形成アルゴリズム + MH法), 近傍探索(1-back近傍探索, Full-back近傍探索, VD-back近傍探索)をそれぞれ組み合わせた局所探索法を実装し、厳密解法である最小流域法の実験結果との比較を行う。ただし、近傍探索は近傍内を全て調べて複数の改善解の中から最も改善が大きいものへ改善する最善移動戦略による。(これに対し、近傍内を調べて最初に発見した改善解に移動する戦略を即時移動戦略と呼ぶ。)本研究での実験環境は次の通りである。

CPU : Pentium4 2.4GHz
MEM : 512MB
OS : VineLinux2.6

5.1 実験結果

表2は配列長46, 100, 158の各インスタンスに対して、最小流域法を用いた実験結果である。時間はCPU時間(秒)、値は求められたエネルギー障壁値(kcal/mol)を表し、横線が引かれている部分はメモリ溢れのため結果が求まらなかったものである。インスタンス46(6,7)を含め、それ以外のインスタンスに対しても、最終形態がランダムのインスタンスに対しては、最小流域法で結果を得ることは困難である。また、配列長が長いインスタンスに対しては、インスタンスによって計算時間が大きく異なる。

表3は各インスタンスに対して、初期解をMH法により生成した後、1-back近傍探索, Full-back近傍探索, VD-back近傍探索による局所探索法を行った結果である。時間はCPU時間(秒)、値はそれぞれで得られた解の E_{top} の値を表し、太字で書かれた数値は真のエネルギー障壁値と一致した値である。ほとんどのインスタンスに対して比較的短い時間で最小流域法と同等の精度の解を発見している。

インスタンス	時間(秒)	値: 最適値(kcal/mol)
46(1)	0.11	28.60
46(2)	0.14	18.35
46(3)	0.14	27.24
46(4)	0.13	31.33
46(5)	0.14	14.30
46(6)	-	-
46(7)	-	-
100(1)	0.50	63.32
100(2)	3.22	52.09
100(3)	4.74	44.50
100(4)	0.74	52.55
100(5)	0.43	64.36
100(6)	0.17	-1.34
158(1)	1.27	69.39
158(2)	-	-
158(3)	-	-
158(4)	-	-
158(5)	11.85	75.82
158(6)	1.94	-48.41
158(7)	262.85	-48.92
158(8)	133.84	-38.34

表2 最小流域法による実験結果

インスタンス	初期解		1-back		Full-back		VD-back	
	値	時間	値	時間	値	時間	値	
46(1)	33.02	0.08	28.60	0.09	28.60	0.09	28.60	
46(2)	21.27	0.12	18.35	0.13	18.35	0.11	18.35	
46(3)	28.81	0.09	27.24	0.09	27.24	0.10	27.24	
46(4)	34.41	0.10	31.33	0.10	31.33	0.11	31.33	
46(5)	18.36	0.16	14.30	0.18	14.30	0.14	14.30	
46(6)	33.91	0.10	28.60	0.09	28.60	0.11	28.60	
46(7)	30.92	0.14	21.71	0.21	19.56	0.14	21.71	
100(1)	70.89	0.21	63.32	0.26	63.32	0.21	63.32	
100(2)	59.24	0.25	52.09	0.31	52.09	0.25	52.09	
100(3)	50.20	0.19	44.50	0.22	44.50	0.18	44.50	
100(4)	58.62	0.19	52.55	0.19	52.55	0.18	52.55	
100(5)	68.23	0.18	64.36	0.18	64.36	0.18	64.36	
100(6)	3.32	0.56	-1.34	0.74	-1.34	0.55	-1.34	
158(1)	73.80	0.89	69.39	0.92	63.39	0.90	69.39	
158(2)	80.23	1.59	73.85	2.96	73.85	1.60	73.85	
158(3)	87.06	2.09	80.89	3.02	80.89	2.10	80.89	
158(4)	68.92	1.32	65.18	1.29	65.18	1.30	65.18	
158(5)	87.55	1.12	75.82	1.45	75.82	1.15	75.82	
158(6)	-43.49	2.50	-46.27	3.94	-46.27	3.95	-46.27	
158(7)	-43.43	4.75	-46.71	9.68	-46.71	8.28	-46.71	
158(8)	-35.13	4.19	-36.01	32.61	-36.01	7.64	-36.01	

表3 初期解(MH法)に対する1-back, Full-back, VD-back近傍探索による実験結果

また、表4は各インスタンスに対して、初期解をGreedy塩基対形成アルゴリズム + MH法により生成した後、1-back近傍探索, Full-back近傍探索, VD-back近傍探索による局所探索法を行った結果である。この手法もほとんどのインスタンスに対して、比較的短い時間で良い解を発見している。

インスタンス	初期解		1-back		Full-back		VD-back	
	値	時間	値	時間	値	時間	値	時間
46(1)	33.02	0.10	28.60	0.09	28.60	0.10	28.60	
46(2)	35.23	0.11	18.48	0.13	18.48	0.13	18.48	
46(3)	28.81	0.10	27.24	0.08	27.24	0.11	27.24	
46(4)	33.41	0.09	31.33	0.10	31.33	0.10	31.33	
46(5)	15.10	0.10	14.30	0.18	14.30	0.12	14.30	
46(6)	33.02	0.09	29.30	0.09	29.30	0.10	29.30	
46(7)	36.38	0.13	21.71	0.30	19.56	0.13	21.71	
100(1)	70.89	0.19	63.32	0.20	63.32	0.20	63.32	
100(2)	59.24	0.22	52.09	0.29	52.09	0.24	52.09	
100(3)	50.20	0.19	44.50	0.23	44.50	0.18	44.50	
100(4)	61.78	0.22	52.55	0.26	52.55	0.23	52.55	
100(5)	72.45	0.18	64.36	0.22	64.36	0.21	64.36	
100(6)	26.74	0.66	5.32	1.10	1.65	1.52	1.23	
158(1)	87.30	1.66	69.39	1.66	69.39	1.66	69.39	
158(2)	81.90	2.27	76.41	4.07	76.41	2.28	76.41	
158(3)	89.88	2.80	80.89	4.48	80.89	2.80	80.89	
158(4)	73.59	1.25	65.18	1.23	65.18	1.24	65.18	
158(5)	84.69	1.51	75.82	1.80	75.82	1.47	75.82	
158(6)	-22.77	5.07	-43.99	12.66	-46.27	16.95	-46.27	
158(7)	-29.39	2.99	-43.97	14.86	-46.71	14.50	-46.71	
158(8)	-11.52	6.20	-34.75	20.27	-37.19	29.27	-37.19	

表4 初期解 (Greedy + MH 法) に対する 1-back, Full-back, VD-back 近傍探索による実験結果

5.2 結果に対する考察

最小流域法との比較

実験結果から局所探索法により得られた局所最適解が最適値と一致するものも多く、解精度は比較的良好と考えられる。また、計算時間もほとんどのインスタンスに対し、最小流域法よりも良い結果を得ている。この傾向はインスタンスの配列長が長くなる程、顕著に表れる。しかし、インスタンス 100(6)、158(6) のように最小流域法の方が計算時間が速い問題例もある。これはインスタンスが最小流域法により解き易い形をしているのも一つの原因であるが、局所探索法における近傍サイズの大きさが計算時間に影響を与えていると考えられる。

近傍探索による違い

配列長が短いインスタンスに対する 1-back, Full-back, VD-back 近傍探索のそれぞれの計算時間の差は小さい。これは、1-back, VD-back 近傍探索は解の改善回数が多く、Full-back 近傍探索では改善回数が少ないため、局所探索法において探索された近傍サイズの合計が、結果的に同程度になっているためである。一方、配列長が長いインスタンスにおいては、それぞれの近傍探索による計算時間は 1-back < VD-back < Full-back の傾向が表れている。インスタンスによっては Full-back 近傍探索の方が VD-back 近傍探索より速く終了しているが、これも解の改善回数が原因である。また、解精度においては Full-back, VD-back 近傍探索の方が 1-back 近傍探索より良い解を得る傾向にある。

初期解生成法による違い

インスタンスによって初期解は Greedy + MH 法を用いた方が MH 方だけの初期解より良い結果を得られるが、実験結果からほとんどのインスタンスに対して解精度、計算時間ともに単純に MH 法による初期解からの探索の方が比較的良好の結果を得ている。解精度は初期解の値、計算時間は局所探索法における改善の回数が影響していると考えられる。しかし、インスタンス 158(8) のように、MH 法の初期解による局所探索法より良い結果を得る問題例も存在する。

6. まとめ

局所探索法を利用することで、厳密解法では解くことが困難な多くのインスタンスにも適応できる、エネルギー障壁近似計算法を提案した。実験結果から解精度は比較的良好く、計算時間も多くのインスタンスに対して良い結果が得られることを確認した。今後の課題としては、最適解の下界値の見積もり、またそれに基づく局所探索法の高性能化、さらにメタ戦略の採用による高性能化などがあげられる。

文 献

- [1] L. Adleman, "Molecular Computing of Solutions to Combinatorial Problems", Science 266, pp 1021-1024, 1994
- [2] J. Cupal, C. Flamm, P.F. Stadler "Density of States, Metastable States, and Saddle Points Exploring the Energy Landscape an RNA Molecule", Proceedings of ISMB-97, pp 88-91, 1997
- [3] C. Flamm, W. Fontana, I.L. Hofacker, P. Schuster "RNA folding at elementary step resolution", RNA6, pp 325-338, 2000
- [4] C. Flamm, I.L. Hofacker, P.F. Stadler, M.H. Wolfinger "Barrier Trees of Degenerate Landscapes", Z. Phys. Chem, pp 155-173, 2002
- [5] I.L. Hofacker, W. Fontana, P.F. Stadler, S. Bonhoeffer, M. Tacker, P. Schuster "Fast Folding and Comparison of DNA Secondary Structures", Monatshefte für Chemie (Chemical Monthly) 125, pp 167-188, 1994
- [6] M. Kubota, M. Hagiya, "Minimum Basin Algorithm: An Effective Analysis Technique for DNA Energy Landscapes", Machines 4, pp 7-20, 2003
- [7] M. Kubota, K. Ohtake, K. Komiya, K. Sakamoto, M. Hagiya, "Branching DNA Machines Based on Transitions of Hairpin Structures", Proc. of the 2003 Congress on Evolutionary Computation (CEC2003) 4, pp 2542-2548, 2003
- [8] S. R. Morgan, P. G. Higgs, "Barrier heights between ground states in a model of RNA secondary structure", J.Phys.A: Math. Gen. 31, pp 3153-3170, 1998
- [9] P. F. Stadler, C. Flamm, "Barrier Tree on Poset-Valued Landscapes", Machines 4, pp 7-20, 2003
- [10] H. Uejima, M. Hagiya "Analyzing Secondary Structure Transition Paths of DNA/RNA molecules", Preliminary Proc. of 9th Int. Meeting of DNA Based Computers (DNA9), pp 92-96, 2003
- [11] M. Yagiura, T. Yamaguchi, T. Ibaraki, "A Variable Depth Search Algorithm for the Generalized Assignment Problem", in: S. Voss, S. Martello, I.H. Osman and C. Roucairol, eds., Meta-Heuristics: Advances and Trends in Local Search Paradigms for Optimization, Kluwer Academic Publishers, pp 459-471, 1999
- [12] M. Zuker, P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamic and auxiliary information", Nucl. Acids. Res. 9, pp 133-148, 1981