

# DNA 形態変化におけるエネルギー障壁値の高速近似計算

武田 勉 (Tsutomu Takeda)\*  
 定兼 邦彦 (Kunihiko Sadakane)†

小野 廣隆 (Hiroataka Ono)†  
 山下 雅史 (Masafumi Yamashita)†

## 1 はじめに

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]。

しかし、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変化パスの評価尺度とする手法がある [3, 4, 6, 7, 8, 9, 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。

\*九州大学 大学院システム情報科学府情報工学専攻 (Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University), [takeda@tcslab.csce.kyushu-u.ac.jp](mailto:takeda@tcslab.csce.kyushu-u.ac.jp)

†九州大学 大学院システム情報科学研究院情報工学部門 (Department of Computer Science and Communication Engineering, Graduate School of Information Science and Electrical Engineering, Kyushu University), {ono, mak, sada}@csce.kyushu-u.ac.jp

る。しかし、DNA 分子のとりうる形態は分子長の指数オーダー分存在し [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。結合された塩基の対を塩基対と呼び、相補性に基づき、塩基配列  $X$  は複数の形態を取りうる。

### 2.1 分子の形態変化とエネルギー障壁

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。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<sup>1</sup>。

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 1 step とする。また、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。

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。

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

<sup>1</sup><http://www.tbi.univie.ac.at/~ivo/RNA>

$$P(s_0, s_n)\},$$

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

一般にエネルギー計算は複雑であり、かつ、 $s_0$  から  $s_n$  までの変化パスは無数に存在するため  $E_{bar}(P(s_0, s_n))$  の計算は困難である。図1はエネルギー60の初期形態

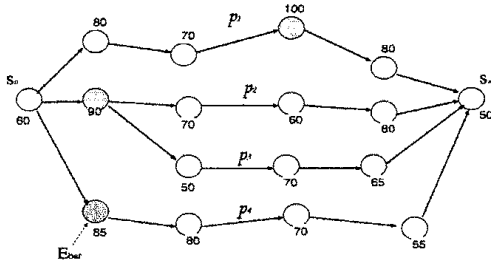


図1: 形態変化パスとエネルギー障壁

$s_0$  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。

### 2.2 エネルギー障壁見積もり問題

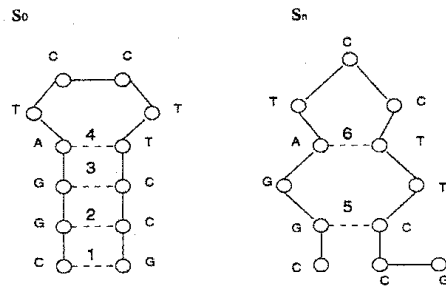
エネルギー障壁見積もり問題は、文字配列  $X$ 、初期形態  $s_0$ 、最終形態  $s_n$ 、を入力として与えたとき、 $E_{top}(p)$  を最小にするパス  $p$  を求める問題である。これを求めるアプローチとしてMH法 [8] や最小流域法 [6] などが提案されている。

本研究においても使用する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を  $s_0$  から除いて、その塩基対(互換な塩基対)を形成するという手順を繰り返す。また、互換でも非互換でもない塩基対を冗長な塩基対と呼ぶ。

図2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  $s_0$  の非互換な塩基対を表している。

### MH法のアルゴリズム

1. 非互換な  $s_0$  の塩基対の数が最も少ない塩基対(複数あるときは解離したときのエネルギー減少が最大またはエネルギー増加が最小となる塩基対を選択)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(形成したときのエネルギー減少が最大またはエネルギー増加が最小となる塩基対から順に選択)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	非互換な $s_0$ の塩基対
5	2,3
6	4

図2: 非互換な塩基対

### 3 局所探索法によるエネルギー障壁近似計算

MH法は比較的良い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が、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高性能な近似解法の構築を目指す。

局所探索法は現在の解に少しの変形を加え、より良い解を探索する手法である。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。すなわち、局所探索法は適当な解  $x$  から始め、近傍  $N(x)$  内を探索し、改善解  $x' \in N(x)$  を発見すると  $N(x')$  から同様の探索を行い、近傍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  $p$  を解とみなし、局所探索法を以下の手順で行う。

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

subject to  $p \in P(s_0, s_n)$

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。

### 3.1 初期解生成

先に述べたように、解は  $s_0$  から  $s_n$  までの形態変化パスとする。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。

- 1: MH 法.
- 2: Greedy 法 (Greedy 塩基対形成アルゴリズム+MH 法)

一般に、DNA 分子は塩基対の数が多い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。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.

### 3.2 近傍と近傍探索

解の定義から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して定義する。ただし、この定義によると近傍解の数自身も非常に多くなるため、探索には不向きである。そこで、MH 法の生成するパスを用いて近傍解を定義することを考える。

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  $p'(p, s_i) = p[0, s_{i-1}], MH(s'_i, s_n) >$  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。

#### k-back 近傍

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

本研究ではこれを用いて3種類の近傍探索を行う。

#### • 1-back 近傍探索

常に 1-back 近傍を探索空間とする。

#### • Full-back 近傍探索

常に Full-back 近傍 ( $N(p)_{Full} = \{p'(p, s_i) \mid 1 \leq i \leq h\}$ )

を探索空間とする。

#### • 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 (variable depth neighborhood[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.

### 3.3 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, 154
- 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

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
154(1~5)	158	ランダム	エネルギー最小
154(6~7)	158	実形態	エネルギー最小
154(8)	154	最小 (77+77)*	エネルギー最小
194(1)	194	実形態	エネルギー最小
194(2~3)	194	ランダム	ランダム

表 1: 本研究で用いたインスタンス

ここでの実形態とは、人工的な配列設計による塩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), 154(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
コンパイラ	: gcc 2.95.3

### 3.3.1 実験結果

表2は配列長46, 100, 154, 194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な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法もしくはGreedy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。

### 3.3.2 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

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
154(1)	1.27	69.39
154(2)	-	-
154(3)	-	-
154(4)	-	-
154(5)	11.85	75.82
154(6)	1.94	-48.41
154(7)	262.85	-48.92
154(8)	133.84	-38.34
194(1)	-	-
194(2)	-	-
194(3)	-	-

表2: 最小流域法による実験結果

が最適値と一致するものも多く、解精度は比較的良いと考えられる。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、初期解をGreedy法により生成したのに対しては、インスタンス100(6), 154(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と改善回数が計算時間に影響を与えていると考えられる。しかし、2種類の初期解からの局所探索法は最小流域法では解が求まらない問題例に対しても高速に解を得ることができる。

#### 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

MH 法								Greedy 法							
インスタンス	初期解		1-back		Full-back		VD-back		初期解	1-back		Full-back		VD-back	
	値	時間	値	時間	値	時間	値	時間		値	時間	値	時間	値	時間
46(1)	33.02	0.07	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>
46(2)	21.27	0.09	<b>18.35</b>	0.08	<b>18.35</b>	0.09	<b>18.35</b>	0.09	<b>18.48</b>	0.09	<b>18.48</b>	0.09	<b>18.48</b>	0.09	<b>18.48</b>
46(3)	28.81	0.08	<b>27.24</b>	0.08	<b>27.24</b>	0.08	<b>27.24</b>	0.08	<b>27.24</b>	0.08	<b>27.24</b>	0.08	<b>27.24</b>	0.08	<b>27.24</b>
46(4)	34.41	0.08	<b>31.33</b>	0.08	<b>31.33</b>	0.08	<b>31.33</b>	0.08	<b>31.33</b>	0.08	<b>31.33</b>	0.08	<b>31.33</b>	0.08	<b>31.33</b>
46(5)	18.36	0.09	<b>14.30</b>	0.11	<b>14.30</b>	0.09	<b>14.30</b>	0.10	<b>14.30</b>	0.10	<b>14.30</b>	0.10	<b>14.30</b>	0.10	<b>14.30</b>
46(6)	33.91	0.08	<b>28.60</b>	0.09	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>	0.07	<b>28.60</b>	0.08	<b>28.60</b>	0.08	<b>28.60</b>
46(7)	30.92	0.09	<b>21.71</b>	0.09	<b>19.56</b>	0.09	<b>21.71</b>	0.10	<b>19.56</b>	0.09	<b>21.71</b>	0.10	<b>19.56</b>	0.09	<b>21.71</b>
100(1)	70.89	0.13	<b>63.32</b>	0.16	<b>63.32</b>	0.12	<b>63.32</b>	0.15	<b>63.32</b>	0.15	<b>63.32</b>	0.13	<b>63.32</b>	0.13	<b>63.32</b>
100(2)	59.24	0.13	<b>52.09</b>	0.15	<b>52.09</b>	0.13	<b>52.09</b>	0.13	<b>52.09</b>	0.15	<b>52.09</b>	0.13	<b>52.09</b>	0.13	<b>52.09</b>
100(3)	50.20	0.11	<b>44.50</b>	0.11	<b>44.50</b>	0.11	<b>44.50</b>	0.11	<b>44.50</b>	0.11	<b>44.50</b>	0.12	<b>44.50</b>	0.12	<b>44.50</b>
100(4)	58.62	0.11	<b>52.55</b>	0.12	<b>52.55</b>	0.11	<b>52.55</b>	0.13	<b>52.55</b>	0.13	<b>52.55</b>	0.13	<b>52.55</b>	0.13	<b>52.55</b>
100(5)	68.23	0.11	<b>64.36</b>	0.11	<b>64.36</b>	0.12	<b>64.36</b>	0.12	<b>64.36</b>	0.12	<b>64.36</b>	0.11	<b>64.36</b>	0.11	<b>64.36</b>
100(6)	3.32	0.21	<b>-1.34</b>	0.21	<b>-1.34</b>	0.23	<b>-1.34</b>	0.23	<b>-1.34</b>	0.33	<b>-1.34</b>	0.23	<b>-1.34</b>	0.23	<b>-1.34</b>
154(1)	73.80	0.38	<b>69.39</b>	0.38	<b>69.39</b>	0.37	<b>69.39</b>	0.37	<b>69.39</b>	0.99	<b>69.39</b>	0.99	<b>69.39</b>	0.99	<b>69.39</b>
154(2)	80.23	1.04	<b>73.85</b>	1.04	<b>73.85</b>	1.04	<b>73.85</b>	1.04	<b>73.85</b>	2.74	<b>73.85</b>	1.41	<b>73.85</b>	1.41	<b>73.85</b>
154(3)	87.06	1.05	<b>80.89</b>	1.35	<b>80.89</b>	1.06	<b>80.89</b>	1.06	<b>80.89</b>	2.68	<b>80.89</b>	1.67	<b>80.89</b>	1.67	<b>80.89</b>
154(4)	68.92	0.60	<b>65.18</b>	0.61	<b>65.18</b>	0.61	<b>65.18</b>	0.61	<b>65.18</b>	0.64	<b>65.18</b>	0.65	<b>65.18</b>	0.65	<b>65.18</b>
154(5)	87.55	0.56	<b>75.82</b>	0.56	<b>75.82</b>	0.56	<b>75.82</b>	0.59	<b>75.82</b>	0.87	<b>75.82</b>	0.59	<b>75.82</b>	0.59	<b>75.82</b>
154(6)	-43.49	1.58	<b>-46.27</b>	1.58	<b>-46.27</b>	1.58	<b>-46.27</b>	1.58	<b>-46.27</b>	7.23	<b>-46.27</b>	11.03	<b>-46.27</b>	11.03	<b>-46.27</b>
154(7)	-43.43	3.06	<b>-46.71</b>	5.24	<b>-46.71</b>	3.45	<b>-46.71</b>	3.45	<b>-46.71</b>	8.82	<b>-46.71</b>	10.22	<b>-46.71</b>	10.22	<b>-46.71</b>
154(8)	-35.13	2.16	<b>-36.01</b>	4.38	<b>-36.01</b>	3.98	<b>-36.01</b>	3.98	<b>-36.01</b>	11.15	<b>-37.19</b>	18.22	<b>-37.19</b>	18.22	<b>-37.19</b>
194(1)	-47.48	10.11	<b>-52.73</b>	10.11	<b>-52.73</b>	10.10	<b>-52.73</b>	10.10	<b>-52.73</b>	27.01	<b>-53.70</b>	20.79	<b>-53.70</b>	20.79	<b>-53.70</b>
194(2)	109.01	0.32	<b>102.43</b>	6.17	<b>102.23</b>	5.63	<b>102.23</b>	5.63	<b>102.23</b>	7.93	<b>101.47</b>	1.65	<b>101.47</b>	1.65	<b>101.47</b>
194(3)	102.53	1.12	<b>95.79</b>	12.72	<b>95.79</b>	2.60	<b>95.79</b>	2.60	<b>95.79</b>	12.61	<b>95.79</b>	2.94	<b>95.79</b>	2.94	<b>95.79</b>

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

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 46(1, 3~6), 100(1~3), 154(5), 194(3) に関しては初期解中の最大エネルギーの値は、Greedy 法を用いた方が MH 方だけの初期解以上の結果を得ている。しかし、実験結果からほとんどのインスタンスに対して、計算時間では単純に MH 法による初期解からの探索の方が比較的良好な結果が得られた。Greedy 法による初期解からの探索の方が改善回数が多いという理由から、計算時間には改善の回数が大きく影響していると考えられるが、インスタンス 154(8), 194(1,2) のように、MH 法による初期解からの局所探索法より良い精度の結果を得る問題例も存在する。

4 局所探索法の高速化

本章では解生成に用いる MH 法の性質を利用し、局所探索法をより高速に行うことを試みる。MH 法は塩基対数を保ったダイレクトパスを発見するため、以下の性質が成り立つ。

異なる形態  $s_a, s_{a'}$  と最終形態  $s_n$  に対して  $s \in MH(s_a, s_n)$  かつ  $s \in MH(s_{a'}, s_n)$

を満たす  $s$  が存在するなら、 $s_b, s_{b'}, s_c$  が存在。ただし、 $MH(s_a, s_n) = \langle MH(s_a, s_b), MH(s_c, s_n) \rangle$ ,  $MH(s_{a'}, s_n) = \langle MH(s_{a'}, s_{b'}), MH(s_c, s_n) \rangle$

図 3 は、解と近傍解の一致する部分パスの省略を模式的に表したものである。10,30,40 という数値は形態変化 step を表している。

この例においては近傍解を生成する際に元の解と step 30 から step 40 までの変化パスが一致するため、これを利用することにより近傍解生成の高速化を図る。

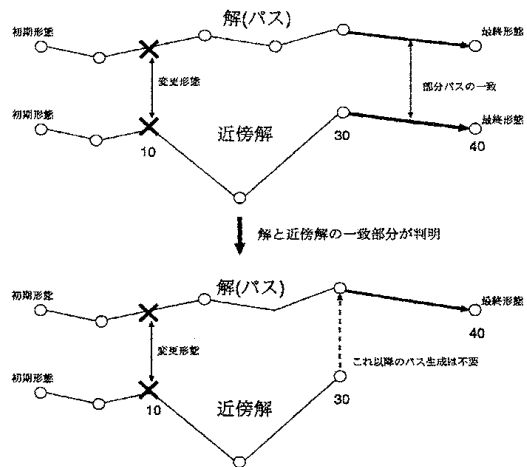


図 3: 変化パスにおける一致部分の省略

### 4.1 パス省略法

本節では上記の考察に基づき、パス探索の省略をいかに行うかについて述べる。

近傍探索においてパス省略を行うことを考えた際、できるだけ似通った解との比較を行うことが望ましい。これは、かけはなれた解との比較では無駄な比較によるオーバーヘッドの存在によりパス省略の利点が少ないからである。そこで、この似通った解を保持する仕組として、近傍探索による解の改善を段とみなした改善段(improve-stairs)を考える(図4)。 $i$ 回目の改善において、近傍解の変更元の変化パス(base-path) $p_{base}$ に対し、形態  $s_a \in p_{base}$  が  $s_{a'}$  に変更され生成された段を以下で定義する。

$$\text{改善段 } i = p[s_{a'}, s_n]$$

図4は、改善段を模式的に表現したものであり、横軸が形態変化 step、縦軸が改善回数(改善度)を表している。この例では初期解(パス長:60)に対し、step 20の形態を別の形態に変更して得られた部分パスを改善段1(改善段1を含む改善解の長さは62)、その後、step 25の形態を別の形態に変更して得られた部分パスを改善段2として表している。2回の改善により得られた解は太線で表される(パス長:64)。

本手法では、この改善段を構成する形で過去の解を保持し、パス省略のための比較対象パス(解)を効率的にチェックすることを図る。

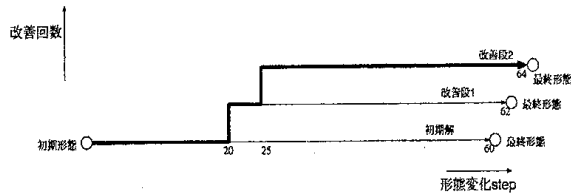


図4: improve-stairs による改善解の表現

### 4.2 形態の照合と incompatible-table

上記のパス省略法においては、各形態の比較(照合)を行う必要があるが、素朴な方法ではこれに配列長のオーダーの計算の時間がかかる。そこで本節ではこの照合の高速化について論ずる。

MH法における形態変化パス生成では、各中間形態において、最終形態内の塩基対に対する非互換な塩基対の個数により塩基対の解離・形成の順序を定める。そ

こで、近傍生成の際に、図5のように変化パスにおける各中間形態に非互換な塩基対の個数を incompatible-table として記憶させておく。最終形態  $s_n$  を形成している塩基対の個数が  $m$  のとき、各塩基対に  $1, \dots, m$  と番号を付す。ある中間形態  $s_i (1 \leq i \leq n)$  における  $s_n$  の塩基対  $j (1 \leq j \leq m)$  に対する非互換な塩基対の個数を  $t_j$  とすると incompatible-table は以下で定義される。

$$\text{incompatible-table}(s_i) = \langle t_1, \dots, t_m \rangle$$

近傍解生成において、improve-stairsにより定まる比較解を用い、各 step ごとに比較解と近傍解の中間形態の incompatible-table を比較する。ある条件を満たし、かつ、incompatible-table が一致すれば、それ以降の変化パスは一致する。条件は近傍解の種類により異なるが、例えば、近傍解生成時の形態変更が互換な塩基対を形成するものであれば、比較パスと近傍解の同 step の中間形態どうしを比較し、変化パス中の形態の一部を別の形態に変更した際に形成された互換な塩基対が比較解においても形成されることである。他の場合もほぼ同様に部分パスの一致を判断できる。

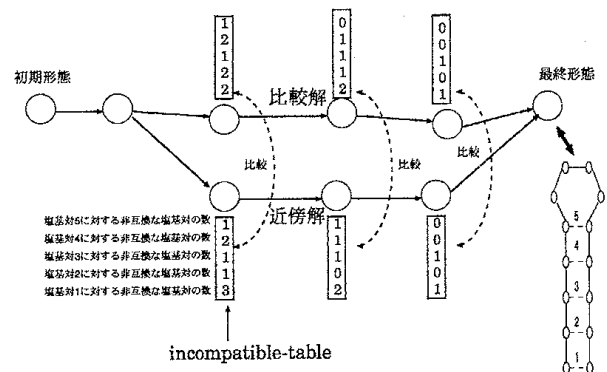


図5: 解と近傍解の incompatible-table の比較

### 4.3 実験結果と考察

表4は初期解をMH法, Greedy法により生成した後、パス省略法を用いた局所探索法を行った結果である。パス省略法により得られる解はパス省略を用いる前と同じなので、計算時間のみを掲載する。この結果は表3に対して、配列長が長い程、パス省略の効果が現れている。

パス省略法を用いた局所探索法は、配列長が小さいインスタンスに対してはあまり効果が得られないが、配列長が大きいインスタンスに対してはパス省略を用

インスタンス	初期解: MH 法			初期解: Greedy 法		
	時間 (1-b)	時間 (F-b)	時間 (VD-b)	時間 (1-b)	時間 (F-b)	時間 (VD-b)
46(1)	0.08	0.08	0.07	0.08	0.08	0.08
46(2)	0.08	0.08	0.09	0.09	0.09	0.09
46(3)	0.07	0.08	0.08	0.08	0.07	0.07
46(4)	0.08	0.08	0.08	0.08	0.08	0.07
46(5)	0.09	0.09	0.09	0.08	0.09	0.09
46(6)	0.07	0.09	0.08	0.09	0.08	0.08
46(7)	0.08	0.09	0.08	0.09	0.10	0.09
100(1)	0.12	0.13	0.11	0.12	0.14	0.12
100(2)	0.12	0.14	0.12	0.12	0.14	0.12
100(3)	0.11	0.11	0.10	0.10	0.10	0.10
100(4)	0.11	0.10	0.10	0.12	0.13	0.12
100(5)	0.11	0.11	0.11	0.10	0.12	0.11
100(6)	0.17	0.18	0.18	0.20	0.26	0.38
154(1)	0.28	0.28	0.28	0.63	0.63	0.63
154(2)	0.87	0.87	0.87	0.99	1.91	1.00
154(3)	0.85	1.03	0.84	1.24	1.75	1.24
154(4)	0.48	0.48	0.48	0.42	0.42	0.41
154(5)	0.44	0.44	0.44	0.54	0.54	0.55
154(6)	0.66	0.65	0.65	1.80	2.90	4.42
154(7)	1.46	2.43	2.03	0.86	3.41	3.85
154(8)	1.08	2.36	2.06	1.61	5.15	7.91
194(1)	3.61	3.62	3.61	7.28	13.11	10.83
194(2)	0.22	3.59	3.28	0.78	4.96	1.22
194(3)	0.71	6.78	1.85	0.70	6.86	1.87

表 4: パス省略法による局所探索法の計算時間

いない場合よりも大幅に計算時間が短縮されている。これは、インスタンスの配列長が大きくなる程、変化パス長が長くなる傾向からパス省略の長さが増加するからである。また、配列長が大きいインスタンスは近傍探索の際の近傍の大きさも大きいものが多く、パス省略の回数が増えるのも原因である。

## 5 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。

## 参考文献

- [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 of an RNA

Molecule", ISMB 1997, 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 RNA Secondary Structures", Monatsh. Chem. 125, pp. 167-188, 1994.
- [6] M. Kubota, M. Hagiya, "Minimum Basin Algorithm: An Effective Analysis Technique for DNA Energy Landscapes", DNA10, pp. 202-213, 2004.
- [7] M. Kubota, K. Ohtake, K. Komiya, K. Sakamoto, M. Hagiya, "Branching DNA Machines Based on Transitions of Hairpin Structures", CEC2003, 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 Trees on Poset-Valued Landscapes", J. Gen. Prog. Evol. Machines 4, pp. 7-20, 2003.
- [10] H. Uejima, M. Hagiya "Analyzing Secondary Structure Transition Paths of DNA/RNA molecules", 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.