

Supplementary Data

Yuan Li and Shaojie Zhang
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL, 32816-2362, USA

1 The Number of Offsprings Produced using Each Mutation Strategy

In the k^{th} generation of the algorithm, each individual in \mathbb{P}_{k-1} (the population of the $k-1^{st}$ generation) reproduces a list of offsprings through a variety of mutation strategies ($\mathbb{M}_1, \dots, \mathbb{M}_Y$). The total number of offsprings that each individual is expected to produce, regardless of the mutation strategy employed, is denoted by a positive integer \mathcal{L} . The number of offsprings, that each individual is expected to produce using mutation strategy \mathbb{M}_y ($1 \leq y \leq Y$) is denoted by $\ell_{\mathbb{M}_y}^k$. \mathcal{L} is constant for each iteration, and $\ell_{\mathbb{M}_y}^k$ is determined adaptively according to the quality of the offsprings produced using \mathbb{M}_y in the previous $(k-1)^{st}$ iteration. In the initial iteration, for $\forall 1 \leq y \leq Y$, we have $\ell_{\mathbb{M}_y}^0 = \mathcal{L}/Y$. From generation k ($k \geq 2$), $b_{\mathbb{M}_y}^{k-1}$ is the number of offsprings in \mathbb{P}_{k-1} produced through mutation strategy \mathbb{M}_y , and

$$\ell_{\mathbb{M}_y}^k = \max \begin{cases} \mathcal{L}_{min} \\ \frac{(b_y^{k-1}/\ell_{\mathbb{M}_y}^{k-1})}{\sum_{y'=1}^Y (b_{y'}^{k-1}/\ell_{\mathbb{M}_{y'}}^{k-1})} \mathcal{L} \end{cases}$$

Mutation strategies having more high quality offsprings in the $(k-1)^{st}$ iteration are allowed to generate more offsprings in the k^{th} generation. In contrast, mutation strategies that perform poorly in the $k-1^{st}$ generation, are also allowed to generate a small number (\mathcal{L}_{min}) of offsprings instead of becoming extinct (the default value of \mathcal{L}_{min} is 3). Therefore, the sum of $\ell_{\mathbb{M}_y}^k$ for $1 \leq y \leq Y$ may be greater than \mathcal{L} .

2 The Discrete Probability Distribution

We define a discrete probability distribution (similar to the discrete Gaussian distribution) over a sample space. Let X be a random variable over \mathbb{R} following a normal

distribution with mean μ ($\mu = 0$) and variance σ^2 . Consider a sample space of n distinguishable objects $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$. The \mathcal{V} -distribution selects a sample v with probability $Pr(v)$ ($Pr(v = v_i) = Pr((i - 1)/n \leq |X| \leq i/n)$ for $1 \leq i \leq n - 1$ and $Pr(v = v_n) = Pr(|X| \geq (n - 1)/n)$). The default value of σ^2 is $1/12$, so that $Pr(|X| \geq 1) = 0.0005$. Consider the set $\{p_l, \dots, p_{(l+n-1)}\}$ and construct \mathcal{V} as follows. If a_t is an addition operation, then $\mathcal{V} = \{v_1 = p_l, \dots, v_{(l+n-1)} = p_u\}$. If a_t is a deletion operation, then $\mathcal{V} = \{v_1 = p_{(l+n-1)}, \dots, v_n = p_l\}$. The actions chain q is chosen from \mathcal{V} with the \mathcal{V} -distribution.

3 Performance of RNAEAPath using different control parameters.

The performance of RNAEAPath using different control parameters is evaluated. Table 1 shows the energy barriers of the best folding pathways found by RNAEAPath over 5 runs with the number of the preserved top offsprings ℓ_1 (ℓ_1 varies from 1 to 16). Table 2 shows the performance of RNAEAPath over 5 runs with the expected size of population ℓ_3 (ℓ_3 varies from 80 to 120). Table 3 shows the performance of RNAEAPath over 5 runs with the number of offsprings \mathfrak{L} (for each individual \mathfrak{L} varies from 80 to 120).

Table 1: Energy barriers of the best folding pathways found by RNAEAPath over 5 runs with different control parameters: ℓ_1 , the number of top offsprings preserved in the next generation, varies from 1 to 16. The energies are measured in kcal/mol.

Instance	Control Parameter: ℓ_1					
	1	4	7	10	13	16
rb1	22	22	22	22	22	22
rb2	7.4	7.5	10	6.5	6.5	6.5
rb3	16.7	16.7	17.1	16.7	16.7	16.7
rb4	16.9	16.9	16.9	16.9	16.9	16.9
rb5	21.44	21.44	21.44	21.44	21.44	21.44
hok	20.2	20.1	20.2	20.1	20.1	20.1
SL	13	13	12.9	12.9	13	13
attenuator	8.6	8.5	8.5	8.5	8.5	8.5
s15	6.6	7.1	7.1	7.1	6.6	7.1
sbox leader	5.2	5.2	5.2	5.2	5.2	5.2
thiM leader	12.3	12.3	12.3	12.3	12.3	12.3
ms2	6.6	6.6	6.6	6.6	6.6	6.6
HDV	16.7	16.8	16.7	16.8	16.8	16.8
dsrA	8	8	8	8	8	8
ribD leader	9.5	9.5	9.5	9.5	9.5	9.5
amv	5.74	5.74	5.74	5.74	5.74	5.74
alpha operon	6.5	6.1	6.1	6.1	6.1	6.1
HIV-1 leader	8.9	8.9	8.9	8.9	8.9	8.9

Table 2: Energy barriers of the best folding pathways found by RNAEAPath over 5 runs with different control parameter: ℓ_3 , the size of population in each generation varies from 80 to 120. The energies are measured in kcal/mol.

Instance	Control Parameter: ℓ_3				
	80	90	100	110	120
rb1	22	22	22	22	22.4
rb2	6.5	7.4	6.5	6.5	6.5
rb3	16.7	17.1	16.7	16.7	16.7
rb4	16.9	16.9	16.9	16.9	16.9
rb5	21.44	21.44	21.44	21.44	21.44
hok	20.1	20.9	20.1	20.7	20.1
SL	13	13	12.9	13	13
attenuator	8.5	8.6	8.5	8.6	8.5
s15	7.1	6.6	7.1	6.6	6.6
sbox leader	5.2	5.2	5.2	5.2	5.2
thiM leader	12.3	12.3	12.3	12	12
ms2	6.6	6.6	6.6	6.6	6.6
HDV	16.8	16.8	16.8	16.7	16.8
dsrA	8	8	8	8	8
ribD leader	9.5	9.5	9.5	9.5	9.5
amv	5.74	5.74	5.74	5.74	5.74
alpha operon	6.1	6.1	6.1	6.1	6.1
HIV-1 leader	8.9	8.9	8.9	8.9	8.9

Table 3: Energy barriers of the best folding pathways found by RNAEAPath over 5 runs with different control parameters: \mathcal{L} , the number of offsprings that each individual should generate varies from 80 to 120. The energies are measured in kcal/mol.

Instance	Control Parameter: \mathcal{L}				
	80	90	100	110	120
rb1	22	22	22	22	22
rb2	7.4	6.5	6.5	6.5	6.5
rb3	17.5	16.7	16.7	16.7	16.7
rb4	16.9	16.9	16.9	16.9	16.9
rb5	21.44	21.44	21.44	21.44	21.44
hok	20.5	20.1	20.1	20.1	20.1
SL	12.9	13	12.9	13	13
attenuator	8.5	8.5	8.5	8.6	8.5
s15	7.1	6.6	7.1	7.1	7.1
sbox leader	5.2	5.2	5.2	5.2	5.2
thiM leader	12.3	12.3	12.3	12	12.1
ms2	6.6	6.6	6.6	6.6	6.6
HDV	16.7	16.8	16.8	16.7	16.8
dsrA	8	8	8	8	8
ribD leader	9.5	9.5	9.5	9.5	9.5
amv	5.74	5.74	5.74	5.74	5.74
alpha operon	6.1	6.1	6.1	6.1	6.1
HIV-1 leader	8.9	8.9	8.9	8.9	8.9

4 Running Time

Table 4 shows the running time of RNAEAPath on 18 conformational switches using the default parameters. The tests are performed on a 32 bit, 2.4 GHz Quad-processor, 3.2GB memory computer, the operating system is Fedora 11. The minimum execution time is 1 minute (in the case of rb4) and 421 minutes (in the case of hok), and on average RNAEAPath terminates in 43 minutes.

Table 4: Running time of RNAEAPath on 18 conformational switches using the default parameters.

Instance	Running Time (minutes)
rb1	34
rb2	16
rb3	22
rb4	1
rb5	17
hok	421
SL	13
attenuator	13
s15	10
sbox leader	20
thim leader	45
ms2	10
HDV	20
dsrA	13
ribD leader	52
amv	14
alpha operon	15
HIV-1 leader	34

5 The Best Folding Pathways Found by RNAEAPath

The best folding pathways between two conformational structures of 18 riboswitches found by RNAEAPath over 5 runs using the default parameters are shown as follows.



Figure 6: The best folding pathway between two conformational structures of hok found by RNAEAPath over 5 runs using the default parameters.

```

>SplicedLeaderAB
AACUAAAACAAUUUUUGAAGAACAGUUUCUGUACUUCUUAUUGGUUAUGAGACUUUC
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-10.7)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-7.8)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-10.1)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-6.1)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-7.4)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-8.1)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-5.82)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-2.56)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-1.53)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-0.01)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-0.2)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(1.5)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(0.7)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-0.7)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(0.2)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(0.2)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(1.8)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(1.9)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(0.1)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(2.2)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(1.4)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(1.9)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-2.6)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(0.8)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-0.9)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-2.4)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-2.4)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-3)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-3.9)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-7.1)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-7.3)
..(((..(((..(((..(((..(((..(((..)))..)))..)))..)))..)))..(-9)
The optimal folding pathways has barrier energy 12.9.

```

Figure 7: The best folding pathway between two conformational structures of SL found by RNAEAPath over 5 runs using the default parameters.


```

>s15
CUGGGAUCGCUGAAUUAGAGAUCCGGUCCUUIUCAUUCUUAUUAUACUUUUGGAGUUUUAAAUGUCUCUAAGUACU
..(((.(.....))))).....(((.(.....)))))... (-17.9)
..(((.(.....))))).....(((.(.....)))))... (-14.3)
..(((.(.....))))).....(((.(.....)))))... (-17.9)
..(((.(.....))))).....(((.(.....)))))... (-13.9)
..(((.(.....))))).....(((.(.....)))))... (-17.9)
..(((.(.....))))).....(((.(.....)))))... (-17.2)
..(((.(.....))))).....(((.(.....)))))... (-12.4)
..(((.(.....))))).....(((.(.....)))))... (-11.5)
..(((.(.....))))).....(((.(.....)))))... (-12.6)
..(((.(.....))))).....(((.(.....)))))... (-10.8)
..(((.(.....))))).....(((.(.....)))))... (-14.4)
..(((.(.....))))).....(((.(.....)))))... (-15)
The optimal folding pathway has barrier energy 7.1.

```

Figure 9: The best folding pathway between two conformational structures of s15 found by RNAEAPath over 5 runs using the default parameters.


```

>ms2
GGGUGGACCCUUUCGGGGUCCUGCCUACAACUJCCUGJCGAGCUAAUGCCAUUUUUAUGUCUUUAGCGGAGAC
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
(((.....))).....(((.....))).....
The optimal folding pathway has barrier energy 6.6 kcal/mol.
(-32.7)
(-29.3)
(-32.6)
(-28.9)
(-28.8)
(-28.8)
(-28)
(-30)
(-28.2)
(-26.5)
(-26.1)
(-27.3)

```

Figure 12: The best folding pathway between two conformational structures of ms2 found by RNAEAPath over 5 runs using the default parameters.

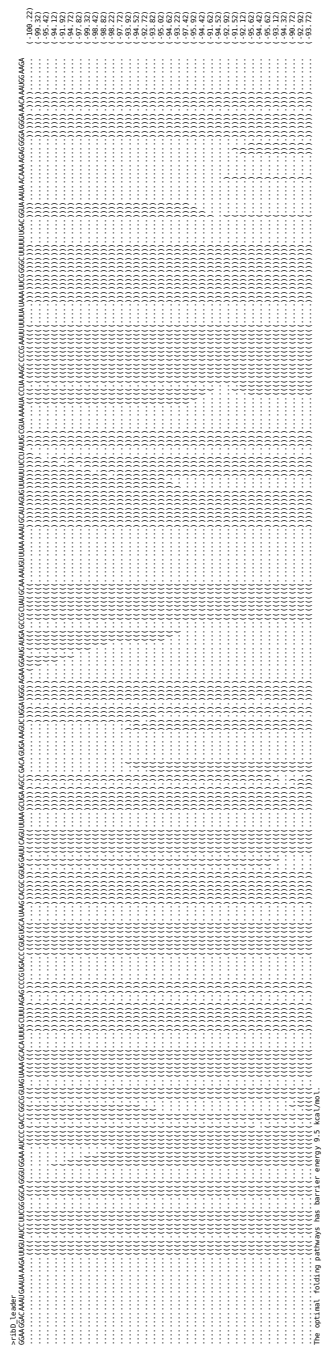


Figure 15: The best folding pathway between two conformational structures of ribD leader found by RNAEAPath over 5 runs using the default parameters.

