SUPPLEMENTARY INFORMATION FOR "RNAIFOLD2T: CONSTRAINT PROGRAMMING DESIGN OF THERMO-IRES SWITCHES"

J.A. GARCIA-MARTIN [1], I. DOTU [2], J. FERNANDEZ-CHAMORRO [3], G. LOZANO [3], J. RAMAJO [3], E. MARTINEZ-SALAS [3], P. CLOTE [1]

[1]: Biology Department, Boston College, Chestnut Hill, MA 02467 (USA).
[2]: Research Programme on Biomedical Informatics (GRIB), Department of Experimental and Health Sciences, Universitat Pompeu Fabra. Dr. Aiguader 88. Barcelona, (Spain).
[3]: Centro de Biologia Molecular Severo Ochoa, Consejo Superior de Investigaciones Cientificas – Universidad Autonoma de Madrid, Nicolas Cabrera 1, 28049 Madrid (Spain).

1. Extended Hairpin and Extended Hairpin with Dangles

In this section, we give detailed definitions of concepts used in the `RNAiFold2T` algorithm to solve the multi-temperature inverse folding problem. We start by identifying a given secondary structure S by its dot-bracket notation $s_1, \ldots, s_n$. `RNAiFold2T` instantiates the RNA sequence $a_1, \ldots, a_n$, whose minimum free energy structure at temperature T_1 [resp. T_2] is S_1 [resp. S_2] by assigning nucleotides to base-paired positions and unpaired positions of S_1 and S_2 in a particular order, defined by the helix and value heuristics applied to a structure decomposition tree for S_1 and S_2. We define two types of decomposition trees: (1) Extended Helix (EH) decomposition tree, and (2) Extended Helix with Dangles (EHwD) decomposition tree. See Figure 1 in the main text for an illustration of the EHwD decomposition tree for a fourU thermometer.

The convex subword $S' = s_i, \ldots, s_j$ of the dot-bracket representation of a secondary structure S is defined to be a *substructure* of S if the dot-bracket expression $s_i, \ldots, s_j$ is a valid secondary structure – i.e. S' is a well-balanced parenthesis expression. An extended helix (EH) [resp. extended helix with dangles (EHwD)] is a maximal substructure S' of S, with closing base pair (i, j) [resp. closing base pair (i, j), including flanking left and right dangle positions $i-1$ and $j+1$, provided the dangles exist in S], defined by the following inductive process, which is motivated by the definition of *order* of a secondary structure [7].

In the base case 0 of the induction, any maximal stem-loop substructure of S [resp. maximal stem-loop substructure of S with flanking left and right dangle, provided the dangle exists in S] is an EH [resp. EHwD], provided no bulge or internal loop has more than 2 adjacent unpaired positions; i.e. bulges have size at most 2, and internal loops are of the form 1×1, 2×1, 1×2, or 2×2. To define EH [rep. EHwD] in the $(k+1)$st inductive

Corresponding author: `clote@bc.edu`.

step, temporarily modify the structure S by replacing all left and right parentheses by a dot for those positions that belong to an EH [resp. EHwD] defined at a previous inductive step $\leq k$. Then an EH [resp. EHwD] in the $(k+1)$st inductive step is any maximal stem-loop substructure of the temporarily modified version of S [resp. maximal stem-loop substructure of the temporarily modified version of S with flanking left and right dangle, if the dangle exists in S], provided no bulge or internal loop has more than 2 adjacent unpaired positions.

The previous definition can be formalized by the following inductive definition. In the base case, define an *extended helix* of S to be a subsequence of the form $S' = s_i, \ldots, s_j$ of S of maximal length, such that: (1) S' is a substructure of S; (2) if $s_x, \ldots, s_y$ is a maximal length subsequence of S' that consists only of dots, then either *(i)* $x, \ldots, y$ are the unpaired positions of a hairpin loop with closing base pair at $(x-1, y+1)$, or *(ii)* $y - x < 2$, which occurs in a bulge, or one side of an interior loop, of size 1 or 2.

In the inductive $(k+1)$st step, define an *extended helix* of S to be a subsequence of the form $S' = s_i, \ldots, s_j$ of S of maximal length, such that: (1) S' is a substructure of S; (2) if $s_x, \ldots, s_y$ is a maximal length subsequence of S' that consists only of dots, then either *(i)* $x, \ldots, y$ are the unpaired positions of a hairpin loop with closing base pair at $(x-1, y+1)$, or *(ii)* $y - x < 2$, which occurs in a bulge, or one side of an interior loop, of size 1 or 2, or *(iii)* position $y+1$ belongs to an EH defined at some step $\leq k$, and $x, \ldots, y$ correspond to the unpaired positions in a left bulge, or left portion of interior loop, of size greater than 2, or *(iv)* position $x-1$ belongs to an EH defined at some step $\leq k$, and $x, \ldots, y$ correspond to the unpaired positions in a right bulge, or right portion of interior loop, of size greater than 2, or *(v)* $x, \ldots, y$ correspond to the unpaired positions in a multiloop or external loop, each of whose components constitutes an EH defined at some step $\leq k$.

An *extended helix with dangles* is analogously defined, except that the leftmost and rightmost positions may constitute a dangle in the original structure S. Leaves of the EH [resp. EHwD] decomposition tree $\mathcal{T}$ consist of all extended hairpins [resp. extended hairpins with dangles] of S that are defined in the base case of the induction, arranged in left-to-right order. Inductively, an EH [resp. EHwD] is defined to be the parent of all proper maximal EHs [resp. EHwDs] that it contains, and these are ordered as daughter nodes in left-to-right order. Finally, the *root* of the decomposition tree $\mathcal{T}$ is the initially given structure S.

Recursively define the *depth* of nodes in decomposition tree $\mathcal{T}$ as follows: the root has depth 0, while a non-root node has depth one greater than its parent. Let $D(k)$ denote the number of nodes in T at depth k. Define the node labels by applying breadth first search; i.e. the root has label 0; nodes at depth 1 have labels $1, \ldots, D(1)$; nodes at depth 2 have labels $D(1)+1, \ldots, D(1)+D(2)$, etc. Define the *height* $ht(x)$ of each node x of decomposition tree $\mathcal{T}$ by induction: if x is a leaf, then $ht(x) = 0$; if x is the parent of nodes $x_1, \ldots, x_m$ and the height of $x_1, \ldots, x_m$ has been defined, then $ht(x) = 1 + \max\{ht(x_1), \ldots, ht(x_m)\}$. In the case of thermosensors, there are two target structures S_1, S_2, and so two decomposition trees $\mathcal{T}_1, \mathcal{T}_2$. In this case, the labels of $\mathcal{T}_1$ are defined as above. If $|\mathcal{T}_1|$ denotes the number of nodes in $\mathcal{T}_1$, then labels of $\mathcal{T}_2$ are defined to be the label, as previously defined, plus $|\mathcal{T}_1|$.

2. Cost functions

In this section, we define the cost function first described in equation (7) of [2], as well as a new variant defined from the notion of ensemble defect. To define the cost function of [2], we require some notation. For RNA sequence $\mathbf{a} = a_1, \ldots, a_n$, secondary structure S and temperature T, let $G_T(\mathbf{a})$ denote the ensemble free energy $-RT \ln Z(\mathbf{a})$, and let $E_T(\mathbf{a}, S)$ denote the free energy of $\mathbf{a}$ with respect to structure S at temperature T – both of these values can be computed by `Vienna RNA Package`. Given sequence $\mathbf{a}$ and target structures S_1 resp. S_2 for temperatures T_1 resp. T_2, the cost function of [2] is defined by

$$[E_{T_1}(\mathbf{a}, S_1) - G_{T_1}(\mathbf{a})] + [E_{T_2}(\mathbf{a}, S_2) - G_{T_2}(\mathbf{a})] + c \cdot \{(E_{T_1}(\mathbf{a}, S_1) - E_{T_2}(\mathbf{a}, S_1)) + (E_{T_2}(\mathbf{a}, S_2) - E_{T_1}(\mathbf{a}, S_2))\} \tag{1}$$

where $c > 0$ is a constant to weight the relative importance that the solution has low free energy with respect to target structure versus having high free energy with respect to the non-target structure.

To define a new ensemble defect based cost function, we require some additional notation. For a given RNA sequence $\mathbf{a} = a_1, \ldots, a_n$ and indices $1 \leq i < j \leq n$, recall that the base-pairing probability $p_{i,j}$ is defined by

$$p_{i,j} = \sum_{\substack{S \text{ such that} \\ (i,j) \in S}} \frac{\exp(-E(S)/RT)}{Z} \tag{2}$$

$$= \frac{\sum_{\substack{S \text{ such that} \\ (i,j) \in S}} \exp(-E(S)/RT)}{\sum_S \exp(-E(S)/RT)} \tag{3}$$

where $E(S)$ is the Turner energy of secondary structure S [6], and Z is the *partition function*, defined by $Z = \sum_s \exp(-E(s)/RT)$, where the sum is taken over all secondary structures s of $\mathbf{a}$. The base pairing probabilities $p_{i,j}$ are computed in `RNAfold` [4], which implements McCaskill's algorithm [5]. Now for each fixed position $1 \leq i \leq n$, define the probability distribution $p^*_{i,j}$, for $j \in [1, n]$,

$$p^*_{i,j} = \begin{cases} p_{i,j} & \text{if } i < j \\ p_{j,i} & \text{if } j < i \\ 1 - \sum_{k>i} p_{i,k} - \sum_{k<i} p_{k,i} & \text{if } i = j \end{cases} \tag{4}$$

A secondary structure S of an RNA sequence $\mathbf{a} = a_1, \ldots, a_n$ is defined to be a set of base pairs (i, j) satisfying the following: (1) If $(i, j) \in S$ then a_i, a_j constitute a Watson-Crick or GU wobble base pair. (2) If $(i, j) \in S$ then $j > i + 3$, a condition that requires at least three unpaired bases in each hairpin loop. (3) If $(i, j) \in S$ and $(x, y) \in S$, and if $\{i, j\} \cap \{x, y\} \neq \emptyset$, then $i = x$ and $j = y$, a condition that disallows base triple formation. (4) If $(i, j) \in S$ and $(x, y) \in S$ are distinct base pairs, then either $i < x < y < j$ or $x < i < j < y$ or $i < j < x < y$ or $x < y < i < j$, a condition that disallows pseudoknot formation. Another possible data structure to represent a secondary structure S is an array $s[1], \ldots, s[n]$ of integers, such that $s[i] = i$ when i is unpaired in S, while $s[i] = j \neq i$

when $(i,j) \in S$ or $(j,i) \in S$. Define the Hamming distance between structures s, t as $d_{\mathrm{H}}(s,t) = |\{i : s[i] \neq t[i]\}|$, i.e. the number of positions i in $[1,n]$ where $s[i] \neq t[i]$.

Given a secondary structure S_0 with array representation s_0, the *ensemble defect* $ED(S_0)$ is the expected Hamming distance to s_0 [1] defined by

$$ED(S_0) = \sum_S \frac{\exp(-E(S)/RT)}{Z} \cdot |\{i : s[i] \neq s_0[i]\}| \tag{5}$$

$$= n - \sum_{i \neq j} p^*_{i,j} I[(i,j) \in s_0] - \sum_i p^*_{i,i} I[s_0[i] = i] \tag{6}$$

where I denotes the indicator function. When the sequence $\mathbf{a} = a_1, \ldots, a_n$ and temperature T need to be indicated, we use the notation $ED(\mathbf{a}, S_0, T)$ to denote ensemble defect of $\mathbf{a}$ for target structure S_0 at temperature T. We now define *ensemble defect based cost* as follows:

$$ED(\mathbf{a}, S_1, T_1) + ED(\mathbf{a}, S_2, T_2) - \xi\Big[(E_{T_1}(\mathbf{a}, S_1) - E_{T_1}(\mathbf{a}, S_2) + (E_{T_2}(\mathbf{a}, S_2) - E_{T_2}(\mathbf{a}, S_1)\Big] \tag{7}$$

where $\xi > 0$ is a constant to weight the free energy of folding into the intended structure S_1 [resp. S_2] at temperature T_1 [resp. T_2].

3. Value ordering heuristic

`RNAiFold` determines the order in which values of base-paired positions in the target structure S are assigned, described as follows. If base pairs $(i,j) \in S$ and $(i+1, j-1) \in S$ and positions $i+1, j-1$ are currently instantiated, then base stacking free energies of are determined for each of the base pair choices G-C, C-G, A-U, U-A, G-U, U-G for positions i, j. A random number between 0 and 2 kcal/mol is added to each of the base stacking free energies; subsequently, the base pair (i,j) is instantiated in order of increasing free energy of the resulting list. This value ordering heuristic, denoted by v_0, is the default heuristic used in `RNAiFold 2.0`; see [3] for a more detailed explanation.

For multiple target structures `RNAiFold2T` defines specific ordering heuristics for base-paired positions depending on the pairing status in each target structure and the respective target temperatures. These heuristics also incorporate a random component to ensure that parallel runs explore the search space in a different order. `RNAiFold2T` employs value ordering heuristics v_0 and v_1, where v_1 is summarized in Table 1 and the pseudocode below. Let S denote the target structure at temperature T, S' denote the target structure at temperature T'. Value ordering heuristic v_1 employs base pair instantiation orderings that are specific to the environment in which the base pair is found, i.e. type 0-7 described as follows. Type 0: base pair $(i,j) \in S \cap S'$; Type 1: $(i,j) \in S$, i, j both unpaired in S', $T < T'$; Type 2: $(i,j) \in S$, i, j both unpaired in S', $T' < T$; Type 3: $(i,j) \in S$, either i or j paired differently in S'; Type 4: $(i,j) \in S$, i unpaired in S'; Type 5: $(i,j) \in S$, j unpaired in S'; Type 6: $(i,j) \in S$, $(i-1,j) \in S'$ or $(i+1,j) \in S'$; Type 6′: $(i,j) \in S$, $(i,j-1) \in S'$ or $(i,j+1) \in S'$; Type 7: $(i,j) \in S$, $i-1, i+1$ unpaired in S or $j-1, j+1$ unpaired in S. In the case of types 0,2,7, the same procedure is employed as in v_0; i.e. the

free energy for each base pair choice G-C, C-G, A-U, U-A, G-U, U-G for base pair (i, j) is tabulated, a random value between 0 and 2 kcal/mol is added to the free energy, and the list is sorted in increasing order. Base pairs are then tried in that order. In case 1, the procedure is opposite that of 0,2,7; i.e. the list is sorted in decreasing order and base pairs are then tried in that order. In case 3, the following fictive *pseudo-energies* -2.90,-2.23,-1.90,-1.37,-1.03,-0.10 are assigned respectively to base pairs G-C, G-U, U-G, U-A, C-G, A-U. A random value between 0 and 2 kcal/mol is added to each pseudo-energy, and then base pairs are tried according to increasing pseudo-energies. In the remaining cases 4,5,6,6′, the same fictive *pseudo-energies* are taken, but are instead assigned to the base pairs indicated in Table 1. For instance, in case 4, pseudo-energies -2.90,-2.23,-1.90,-1.37,-1.03,-0.10 are assigned respectively to G-C, G-U, U-G, U-A, C-G, A-U. A random value between 0 and 2 kcal/mol is added to each pseudo-energy, and then base pairs are tried according to increasing pseudo-energies.

The following pseudocode and Table 1 together summarize the different types defined in `RNAiFold2T` and its preferred value order. In the pseudocode, `StrList` is a 0-indexed array $[S_1, S_2]$ (resp. $[S_1, \ldots, S_m]$) of structures in the 2-temperature (resp. m-temperature) inverse folding problem. Let `index[S]` denote the index in `StrList` for structure S.

Pseudocode for value ordering for m-temperature inverse folding.

```
for (S in StrList)
    T = folding temperature of S
    for each bp(i,j) in S
        if ((i-1 and i+1 is unpaired in S) or (j-1 and j+1 is unpaired in S))
            Type7
        else
            if (S == StrList[m-1])
                S'= StrList[0]
            else
                S'=StrList[index[S]+1]
            T' = folding temperature of S'
            if (i,j) unpaired in S'
                if (T<T')
                    Type1
                else
                    Type 2
            else if (bp[i]=j in S')
                Type0
            else if (i and j paired in S')
                Type 3
            else if (i is paired in S')
                if (bp[i] in S' adjacent to j)
                    Type 6
                else
                    Type 4
            else if (j is paired in S')
                if (bp[j] in S' adjacent to i)
                    Type 6
                else
                    Type 5
```

4. Additional helix ordering heuristics

Apart from the helix heuristics $overlap_1, overlap_2$ described in the Methods section of the main text, `RNAiFold2T` implements two additional heuristics, $overlap_3, overlap_4$, for variable ordering at the helix level. The final order, determined by a CP search, is the

permutation σ that minimizes $\sum_{i=0}^{N}\sum_{j=0}^{N} diff_{\alpha}(\sigma, H_i, H_j)$, as described in the main text. We summarize the four helix heuristics below.

- $overlap_1(H, H')$ is 1 if $[i, j] \cap [i', j'] \neq \emptyset$, or equivalently $\max(i, i') \leq \min(j, j')$; otherwise $overlap_1(H, H')$ is 0.
- $overlap_2(H, H')$ is the number of positions k in H and H', for which k is base-paired in both H and H'.
- $overlap_3(H, H')$ is the total number of nucleotide positions k in H and H' (including possible bulges of size 1 or 2, as well as internal loops of sizes 1×1, 1×2, 2×1, and 2×2).
- $overlap_4(H, H')$ is a normalized version of $overlap_2(H, H')$ where the number of overlapping base paired positions in H and H' is divided by the number of base paired positions in H. (Note that $overlap_4$ is not necessarily symmetric.)

5. Structural difference between target structures

Following a referee's suggestion, we investigated the relation between the base pair distance $d_{\mathrm{BP}}(S_1, S_2)$ between target secondary structures S_1, S_2 and the time required by `RNAiFold2T` to find a solution.

In the following, S_1 [resp. S_2] denotes a given target RNA secondary structure at temperature T_1 [resp. T_2], where $T_1 < T_2$. Since some natural thermosensors involve *unzipping* (see main text), we considered both $d_{\mathrm{BP}}(S_1, S_2)$ as well as the number $|S_2 - S_1|$ of base pairs belonging to S_2 but not S_1. Indeed if S_2 is obtained from S_1 but *unzipping*, then we expect that $S_2 - S_1 = \emptyset$.

Using Rfam data from the paper, the Pearson correlation was computed between run time to find a solution using LNS from RNAiFold2T and the base pair distance between the two target structures S_1 at temperature T_1 and S_2 at temperature T_2. We additionally generated artificial RNAs of length 40, computed the MFE structure S_1 at 20°C and S_2 at 40°C. By analyzing the data, two problem instances were obtained for each base pair distance $k = 1, \ldots 15$, where each problem instance is stipulated by target structure S_1 at 20°C and S_2 at 40°C. Subsequently, LNS from RNAiFold2T was executed on each problem instance 20 times, each run bound by 30 minutes. Statistics for each problem instance were gathered for the number of runs (maximum of 20 runs) for which LNS returned a solution, the total time required for the 20 attempted runs (unsuccessful runs contributed 30 minutes), the total time required for the successful runs (run times of 30 minutes for unsuccessful runs were removed).

Results from both experiments support the following conclusions:

(1) There is moderate negative correlation (≈ 0.5) between the number of solved structures and the base pair distance $d_{\mathrm{BP}}(S_1, S_2)$ between S_1 and S_2.
(2) There is moderate negative correlation (≈ 0.5) between the number of solved structures and the number $|S_2 - S_1|$ of base pairs belonging to S_2 but not to S_1.
(3) There is moderate positive correlation (≈ 0.5) between average runtime to obtain a solution (total runtime for solved problems divided by the number of solved problem instances) and $d_{\mathrm{BP}}(S_1, S_2)$.

(4) There is moderate positive correlation (≈ 0.5) between average runtime to obtain a solution (total runtime for solved problems divided by the number of solved problem instances) and $|S_2 - S_1|$.
(5) There is no correlation between the average cost (sum of cost function values for all solved problem instanced divided by number of solved problem instances) and $d_{\mathrm{BP}}(S_1, S_2)$.
(6) There is no correlation between the average cost and $|S_2 - S_1|$.
(7) There is a weak to moderate correlation ($0.35 - 0.57$) between the average cost and the number $|S_2|$ of base pairs belonging to structure S_2.
(8) There is a high correlation (0.96) between $d_{\mathrm{BP}}(S_1, S_2)$ and $|S_2 - S_1|$.

The data is available in a supplementary spreadsheet.

References

[1] R. M. Dirks, M. Lin, E. Winfree, and N. A. Pierce. Paradigms for computational nucleic acid design. *Nucleic. Acids. Res.*, 32(4):1392–1403, 2004.
[2] C. Flamm, I.L. Hofacker, S. Mauer-Stroh, P.F. Stadler, and M. Zehl. Design of multi-stable RNA molecules. *RNA*, 7:254–265, 2001.
[3] J.A. Garcia-Martin, P. Clote, and I. Dotu. RNAiFold: A constraint programming algorithm for RNA inverse folding and molecular design. *J Bioinform Comput Biol*, 11(2):1350001, 2013. in press.
[4] R. Lorenz, S. H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P. F. Stadler, and I. L. Hofacker. Viennarna Package 2.0. *Algorithms. Mol. Biol.*, 6:26, 2011.
[5] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.
[6] D. H. Turner and D. H. Mathews. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic. Acids. Res.*, 38(Database):D280–D282, January 2010.
[7] M. S. Waterman. Secondary structure of single-stranded nucleic acids. *Studies in Foundations and Combinatorics, Advances in Mathematics Supplementary Studies*, 1:167–212, 1978.

Type	Condition	Value order
$0^{\dagger}$	$(i,j) \in S \cap S'$	GC-CG-AU-UA-GU-UG
$1^{\ddagger}$	$(i,j) \in S$, i,j both unpaired in S', $T < T'$	UG-GU-UA-AU-CG-GC
$2^{\dagger}$	$(i,j) \in S$, i,j both unpaired in S', $T' < T$	GC-CG-AU-UA-GU-UG
3	$(i,j) \in S$, either i or j paired differently in S'	GC-CG-GU-UG-AU-UA
4	$(i,j) \in S$, i unpaired in S'	GC-GU-UG-UA-CG-AU
5	$(i,j) \in S$, j unpaired in S'	CG-UG-GU-AU-GC-UA
6	$(i,j) \in S$, $(i-1,j) \in S'$ or $(i+1,j) \in S'$	UG-GU-AU-UA-CG-GC
$6'$	$(i,j) \in S$, $(i,j-1) \in S'$ or $(i,j+1) \in S'$	UG-GU-AU-UA-CG-GC
$7^{\dagger}$	$(i,j) \in S$, $i-1,i+1$ unpaired in S or $j-1,j+1$ unpaired in S	GC-CG-AU-UA-GU-UG

TABLE 1. **Value ordering for base pairs used in `RNAiFold2T`.** Assume that S [resp. S'] is the target structure at temperature T [resp. T']. We consider cases where $(i,j) \in S \cap S'$, $(i,j) \in S - S'$, etc. Despite the order indicated in the table, the implementation in `RNAiFold2T` includes a random component, so that different parallel runs will explore the search space in a different fashion. This effects only the order of base pair value assignments, but not the completeness of CP – regardless of value order, CP involves a complete search of the search space using a branch-and-prune strategy. Types marked with a dagger † [resp. ‡] correspond to increasing [resp. decreasing] base stacking free energies, as described in Section 5 of this Supplementary Information.

EMBL acc.	n	RF family	o-1-v0	o-1^{t}-v0	o0-v0	o1-v0	o2-v0	o-1-v1	o-1^{t}-v1	o0-v1	o1-v1	o2-v1
M13767.1/3-60	58	λthermo	0	**58**	17	17	23	0	4	10	8	6
CP000243.1/1246604-1246546	59	λthermo	0	**59**	16	17	14	1	0	13	4	8
CP000026.1/2520723-2520781	59	λthermo	0	0	1	0	3	0	14	4	**10**	7
CP001144.1/624595-624537	59	λthermo	0	9	22	19	11	0	9	**31**	24	30
AY736146.1/34404-34346	59	λthermo	0	**16**	5	0	0	0	0	1	4	2
CP000647.1/1773227-1773291	65	FourU	12	87	13	86	**94**	11	84	14	86	82
CP001127.1/1302123-1302187	65	FourU	0	71	19	73	**74**	7	62	8	62	60
CP001144.1/2031534-2031470	65	FourU	0	0	0	0	0	0	0	0	0	0
ACDJ01000026.1/381061-380989	73	ROSE_2	0	0	0	**32**	7	0	0	1	2	11
ABWL02000023.1/393416-393344	73	ROSE_2	0	1	1	10	11	4	0	2	3	**22**
CP000653.1/14627-14699	73	ROSE_2	0	97	91	92	**100**	0	98	**100**	99	**100**
CP000036.1/3699544-3699616	73	ROSE_2	0	0	0	0	0	0	0	1	0	**2**
CP000026.1/3798554-3798481	74	ROSE_2	0	16	0	0	0	0	11	2	1	**20**
AE017220.1/3951363-3951290	74	ROSE_2	2	0	11	86	**88**	5	0	14	86	83
BAAW01000185.1/6674-6747	74	ROSE_2	0	77	97	99	88	0	75	**100**	62	83
CP000647.1/4480191-4480116	76	ROSE_2	0	2	2	0	**24**	0	1	1	0	9
CP000009.1/1450710-1450627	84	ROSE	0	89	88	3	0	10	91	**94**	1	3
AP003017.1/94542-94451	92	ROSE	**73**	17	59	16	50	17	39	21	42	24
AE007872.2/441983-442075	93	ROSE	76	72	85	84	78	96	96	**98**	94	93
AE007872.2/51225-51317	93	ROSE	9	11	2	1	3	2	**76**	4	10	8
AL591985.1/872145-872052	94	ROSE	**92**	0	0	2	1	13	0	58	21	27
BA000012.4/1943819-1943723	97	ROSE	93	67	62	59	64	70	97	95	96	**100**
RU55047.1/3106-3215	110	ROSE	0	**2**	0	0	0	0	0	1	0	0
AJ003064.1/2697-2806	110	ROSE	2	0	0	5	1	1	3	0	**10**	2
U55047.1/5180-5291	112	ROSE	4	2	0	0	0	40	22	42	**57**	53
AJ010144.1/622-738	117	ROSE	0	90	84	75	91	58	96	**98**	95	97
AJ003064.1/2430-2312	119	ROSE	0	94	0	0	0	66	**98**	74	66	60
Total			363	937	675	776	825	401	976	887	943	**992**
Str. Solved			9	20	18	18	19	15	18	**25**	23	**25**

TABLE 2. `RNAiFold2T` **heuristic combination test.** Test of combinations of helix ordering heuristics ($o-1$, $o-1^{t}$, $o0$, $o1$, $o2$) and value ordering heuristics ($v0$, $v1$), using `RNAiFold2T CP`. Benchmarking statistics were obtained from 100 runs, with time limit set of 10 minutes per run, performed on a Core2Duo PC (2.8 GHz; 2 Gbyte memory; CentOS 5.5). $v0$ denotes the value ordering used in `RNAiFold 2.0`; $v1$ stands for the new value ordering from Supplementary Table 1; $o-1$ denotes the helix ordering for a single structure, where the helix search order is from leaves to root in the EHwD decomposition tree of the first target structure alone (with intermediate checks whether any fully instantiated sequence for any fully instantiated sequence for an EHwD of the second structure folds into S_2 at T_2). $o-1^{t}$ denotes the helix ordering for a single structure, as in $o-1$, except that the EHwD decomposition tree is for the target structure at the *higher temperature*. $o0$ denotes the helix ordering from leaves to root in the combined decomposition tree for both target structures, as depicted in Figure 1 of the main text, while $o1$ [resp. $o2$] denotes the helix ordering heuristic 1 ($overlap_1$) [resp. heuristic 2 ($overlap_2$)], described in Section 4 of this Supplementary Information.

Parameters			RNAiFold2T		Frnakenstein		SwitchDesign	
EMBL acc.	n	Rfam family	solved	Avg. cost	solved	Avg. cost	solved	Avg. cost
M13767.1/3-60	58	λ thermo	**30**	-1.12	**30**	-1.97	**30**	-3.66
CP000243.1/1246604-1246546	59	λ thermo	**30**	-1.22	**30**	-1.77	**30**	-3.70
CP000026.1/2520723-2520781	59	λ thermo	**29**	1.66	27	-1.09	27	-2.14
CP001144.1/624595-624537	59	λ thermo	**30**	0.54	27	-1.16	28	-2.37
AY736146.1/34404-34346	59	λ thermo	**30**	0.20	26	-1.30	24	-2.64
CP000647.1/1773227-1773291	65	FourU	**30**	2.50	**30**	2.82	16	0.95
CP001127.1/1302123-1302187	65	FourU	**30**	2.43	**30**	1.91	26	0.37
CP001144.1/2031534-2031470	65	FourU	3	2.31	**28**	1.48	25	0.93
ACDJ01000026.1/381061-380989	73	ROSE_2	11	3.75	**19**	2.19	8	1.10
ABWL02000023.1/393416-393344	73	ROSE_2	**2**	4.08	0	-	**2**	1.56
CP000653.1/14627-14699	73	ROSE_2	**26**	3.87	6	2.60	12	0.84
CP000036.1/3699544-3699616	73	ROSE_2	**30**	1.61	**30**	1.48	28	0.40
CP000026.1/3798554-3798481	74	ROSE_2	**30**	1.60	**30**	2.16	23	0.21
AE017220.1/3951363-3951290	74	ROSE_2	**30**	3.83	**30**	3.90	11	0.84
BAAW01000185.1/6674-6747	74	ROSE_2	**30**	3.09	**30**	4.29	28	0.99
CP000647.1/4480191-4480116	76	ROSE_2	**30**	2.82	**30**	2.35	25	0.22
CP000009.1/1450710-1450627	84	ROSE	**30**	2.63	**30**	2.07	26	0.31
AP003017.1/94542-94451	92	ROSE	**30**	2.25	**30**	2.05	**30**	0.14
AE007872.2/441983-442075	93	ROSE	**30**	4.67	**30**	4.09	16	1.10
AE007872.2/51225-51317	93	ROSE	**30**	6.88	27	5.62	12	2.08
AL591985.1/872145-872052	94	ROSE	**29**	6.13	20	3.70	19	1.44
BA000012.4/1943819-1943723	97	ROSE	**30**	5.08	**30**	5.16	23	1.61
RU55047.1/3106-3215	110	ROSE	**30**	6.43	7	3.47	14	1.03
AJ003064.1/2697-2806	110	ROSE	**30**	7.46	29	7.43	7	1.99
U55047.1/5180-5291	112	ROSE	6	8.08	1	7.66	**17**	1.58
AJ010144.1/622-738	117	ROSE	**30**	7.58	**30**	5.82	18	1.70
AJ003064.1/2430-2312	119	ROSE	4	8.91	0	-	**9**	3.04
Total/Avg. Cost			**680**	3.63	637	2.60	534	**0.37**
Str. Solved			**27**		25		**27**	

TABLE 3. **Benchmark for sequences shorter than 130 nt.** Summary of the computational results for Rfam structures shorter than 130 nucleotides comparing `RNAiFold2T` (LNS), `SwitchDesign` and `Frnakenstein`. Benchmarking was performed over 30 runs with time limit set to 30 minutes for each run, measured on a Core2Duo PC (2.8 GHz; 2 Gbyte memory; CentOS 5.5).

			RNAiFold2T		Frnakenstein		SwitchDesign	
EMBL acc.	n	Rfam family	solved	Avg. cost	solved	Avg. cost	solved	Avg. cost
M55160.1/297-426	130	PrfA	2	5.59	0	-	**4**	1.80
AJ002742.1/161-290	130	PrfA	**10**	2.69	**10**	2.13	9	0.30
X72685.1/1303-1435	133	PrfA	**10**	6.05	0	-	9	0.70
AAEU020001321/310682-310830	149	Hsp90_CRE	0	-	0	-	**3**	1.52
AAPQ010065501/448359-448507	149	Hsp90_CRE	**10**	6.17	6	5.50	1	1.45
AY1220801/193-344	152	Hsp90_CRE	**10**	4.18	5	3.16	5	0.84
X038111/879-1030	152	Hsp90_CRE	**10**	4.31	9	3.05	4	1.41
L23115.1/459-834	376	CspA	0	-	0	-	0	-
AF017276.1/479-855	377	CspA	**10**	9.47	0	-	9	2.33
ABJD02000101.1/494629-495045	417	CspA	0	-	0	-	0	-
CP000647.1/4296745-4297166	422	CspA	0	-	0	-	0	-
CP000653.1/190938-191361	424	CspA	0	-	0	-	0	-
ABWM02000027.1/244578-245001	424	CspA	0	-	0	-	0	-
ABWL02000023.1/204545-204970	426	CspA	0	-	0	-	0	-
ABEH02000004.1/260631-260204	428	CspA	0	-	0	-	0	-
CP000946.1/174765-174338	428	CspA	0	-	0	-	0	-
CP000822.1/4602942-4603373	432	CspA	0	-	0	-	0	-
ACCI02000028.1/45583-45145	439	CspA	0	-	0	-	0	-
AAOS02000014.1/93711-93269	443	CspA	0	-	0	-	0	-
AALD02000025.1/10362-9916	447	CspA	0	-	0	-	0	-
ABXW01000053.1/281917-281471	447	CspA	0	-	0	-	0	-
Total/Avg. Cost			**62**	5.50	30	3.46	44	**1.29**
Str. Solved			7		4		**8**	

TABLE 4. **Benchmark for sequences of length greater than 130 nt.** Summary of the computational results for Rfam structures of length at least 130 nucleotides comparing `RNAiFold2T` (LNS), `SwitchDesign` and `Frnakenstein`. Benchmarking was performed over 10 runs with time limit set to 60 minutes for each run, measured on a Core2Duo PC (2.8 GHz; 2 Gbyte memory; CentOS 5.5).

EMBL	FRNA	SD-upd.	SD-sel	RNAiFold2T
CP000243.1/1246604-1246546	675/24,431 (36.2)	535/16,436 (30.7)	23/775 (33.7)	177,428/2,427,236 (13.7)
CP000026.1/2520723-2520781	296/11,529 (38.9)	598/18,519 (31.0)	16/787 (49.2)	68,800/674,593 (9.8)
CP001144.1/624595-624537	341/12,334 (36.2)	342/11,706 (34.2)	27/1,146 (42.4)	216,809/3,665,692 (16.9)
AY736146.1/34404-34346	321/11,976 (37.3)	290/10,163 (35.0)	20/641 (32.1)	64,853/1,027,058 (15.8)
M13767.1/3-60	811/27,008 (33.3)	520/14,733 (28.3)	18/682 (37.9)	49,598/533,629 (10.8)
Average	489/17,456 (35.7)	457/14,311 (31.3)	21/806 (38.8)	115,498/1,665,642 (14.4)

TABLE 5. **Number of solutions for 2-temperature inverse folding with target structures for λ phage CIII thermoswitches from Rfam family RF01804.** Column headers: EMBL accession code, `Frnakenstein`, `SwitchDesign` (with updates – see main text), `SwitchDesign` (selection – see text), `RNAiFold2T`. For each program, run time was 24 hours, where a restart was forced if no new solution was found within 1 hour. Results are presented as *A/B (C)*, where A is the number of distinct solutions returned, B the number of distinct solutions after additionally testing all single point mutations of sequences from A, and C is the ratio of B over A. Clearly, `RNAiFold2T` computes two orders of magnitude more solutions than the other methods.

Accession number, family	Structures S_1, S_2	Temp T_1, T_2	Size
CP000026.1/2520723-2520781	.((((.(((.(((((.(((((.....(((....)))..)))))))))))))))...)	32	59
Lambda thermo	.....(((..(((.....)))..))).....	55	59
CP001144.1/624595-624537	.((((.(((.((((.(((((.((..((....))..))...)))))))))))))))...)	32	59
Lambda thermo	.....(((..(((.....)))..))).....	55	59
AY736146.1/34404-34346	.((((.(((.(((((.(((((.(((((....)))))...)))))))))))))))...)	32	59
Lambda thermo	.....(((..(((.....)))..))).....	55	59
M13767.1/3-60	.((((.(((..(((.(((((((.((((((....)))))).)))))))).))))))).))))))..	32	58
Lambda thermo	.....(((..(((.....)))..))))...	62	58
CP000243.1/1246604-1246546	.((((.(((..(((.(((((((.(((((....))))))).)))))))).))))))).)))))...	32	59
Lambda thermo	.....(((..(((.....)))..)))))....	62	59
CP001144.1/2031534-2031470	((.((((((...)))))))).....(((((.(((((((((....))))))))).))))).....	37	65
FourU	.....(((.(.((.....((..(((((....)))))..))....)).).)))).	58	65
CP001127.1/1302123-1302187	((.((((((...)))))))).....(((((.(((((((((....))))))))).))))).....	37	65
FourU	.....(((.....(((((.((..(((((....)))))..)).)))))...)))).	58	65
CP000647.1/1773227-1773291	((.((((((...)))))))).....(((((.(((((((((....))))))))).))))).....	37	65
FourU	.....(((.....(((((.(((((((((....))))))))).)))))...)))).	45	65
CP000653.1/14627-14699	(((((((((.((((((.(((.....)))))))))...)))...)))))).....((.....)))...	20	73
ROSE 2	...(((((..((((((.(((.....)))))))))..........(((....)))))))).....	42	73
ABWL02000023.1/393416-393344	....((.....(((((.(((.....)))))))))...((..(((.(((....)))))))....)).....))...	20	73
ROSE 2	..(((((.(((((((.(((.....)))))))).....((.....)).....))))))).....	42	73
ACDJ01000026.1/381061-380989	....((.....((((.(((.....)))))))).((((..(((.(((....)))))))....)))...))...	20	73
ROSE 2	..(((((...((((((.(((.....)))))))).....(((....))).)))))).....	42	73
CP000036.1/3699544-3699616	..(((((.(((((...(((....(((((((.....).))).))).))).))).))))))).....	20	73
ROSE 2	..(((((.(((((...(((....(((((((.....).))).))).))).))).))))))).....	42	73
AE017220.1/3951363-3951290	....(((.(..(((((.(((.....)))))))).)))((((.(((.....)))))))...)))....)))...	20	74
ROSE 2	....(((.....((((.(((.....)))))))).....(((.(((.(((....)))))))...)))....)))...	42	74
CP000026.1/3798554-3798481	....((((((..(((((.(((.....)))))))).....(..(((.(((....)))))))..)..))))).....	20	74
ROSE 2	....((((((..(((((.(((.....)))))))).....(((.(((....))))))).....	42	74
BAAW01000185.1/6674-6747	.((.((((((..(((((.(((.....)))))))).(((..(((.(((....)))))))..)))))))).))..	20	74
ROSE 2	.((.(((((((.(((((.(((.....)))))))).....(((.(((....)))))))....)))))).))..	42	74
CP000647.1/4480191-4480116	.(((....))).(((((.(((.....)))))))).....(((.(((.(((....)))))))..))).....	20	76
ROSE 2	.(((....))).(((((.(((.....)))))))).....(((.(((....)))))).....	42	76
CP000009.1/1450710-1450627	((((((.((...)))))))).....(((((.....))))).(((((((.(((((((.(((....)))))))))).))))))).	20	84
ROSE	((((((.(....).))))))....(((((.....))))).(((..(.(((((((.(((....)))))))))).).)..))).	42	84
AP003017.1/94542-94451	...(((....))))).....((((((((((.....)))))))))).((((((..(((((((.(((....)))))))))).))))))..	20	92
ROSE	...(((....))))).....((((((((((.....)))))))))).((..(((..(((((((.(((....)))))))))).)..)..))..	42	92
AE007872.2/51225-51317	((((((.(((....)))))))))...(((((((((((((.....)))))))))))).((.(((((((.(((....)))))))))).)).)))).	20	93
ROSE	((((((.(((....)))))))))..(((((((((((((.....)))))))))))).....(((((((.(((....)))))))))).....))))	42	93
AE007872.2/441983-442075	((((((.(((....))).)))))).((.....)).((((.(.....).)))).((((((.(((((((.(((....)))))))))).)))))...	20	93
ROSE	((((((.(((....))).)))))).....((((.(.....).)))).((((((.(((((((.(((....)))))))))).)))))...	42	93
AL591985.1/872145-872052	.(((((.((.....)))))))((.(((((((((.......))))))))))..((((((.(.(((((((.(((...)))))))))).))))))..	20	94
ROSE	.(((((.((.....)))))))......(((((((.(((.......))).))))))).((((((((((.(((...)))))))))).))))....	42	94
BA000012.4/1943819-1943723	((((((.((......)))))))....((.((((((((((.......)))))))))).((((.(((((((.(((....)))))))))).)))).)).	20	97
ROSE	((((((.((......)))))))).....(((((((((((.......)))))))))))).(((((.(((((((.(((....)))))))))).))))....	42	97
AJ003064.1/2697-2806	(((((((....)))))))(((((((.((....)))))))))...(..((((((.....)))))))..))(((.((.(((((((.(((....)))))))))).)).)))..	20	110
ROSE	(((((((....)))))))(((((((.((....)))))))))......(((((((.....)))))))...(((.((.(((((((.(((....)))))))))).)).))))..	42	110
U55047.1/3106-3215	.(((((.....)))))((((((.((....)))))))).....((((((.((((....)))).)))))).(((((.(((..((.(((....)))))..))).))))))..	20	110
ROSE	.(((((.....)))))).(((((((...((.(((((.......))))))..)))...))))))).....((((((.(((...(.(((....))).).)..))).))))))..	42	110
U55047.1/5180-5291	.(((((.....)))))(((((.(((....))))))))..(((.((((((.((((....)))).)))))))))...(((((...((.(((........))))).))).)).	20	112
ROSE	.(((((.....)))))(((((.(((....)))))))).....(((((.(((((....)))).)))))).(((((.(((...(.(((....))).).)..))).))))))..	42	112
AJ010144.1/622-738	((((((....))))))(((((((.((....))))))))...)))).....(((((((.....)))))))).(((.(((..((.(((....)))))..))).))))..)).	20	117
ROSE	((((((....))))))...((((((.((....)))))))).....(((((((.....)))))))).(((.(((...(.(((....))).).)..))).))))..)).	42	117
AJ003064.1/2430-2312	.((((((....))))))...((((((.((....)))))))).((.(((.(((.(((.((((((.....))))))))))))....((((........)))).))))).((.....)).	20	119
ROSE	.((((((....))))))((.((((((.((....)))))))).))...........(((((((.....)))))))).(((.(((...(.(((....))).).)..))).))))..)).	42	119

TABLE 6. **Thermometers of length at most 130 nt used in benchmarking.** Each RNA corresponds to two successive rows, where the first row contains the EMBL accession code, target structure S_1, temperature T_1 for S_1, and structure length, while the second row contains the type of thermosensor, target structure S_2, temperature T_2 for S_2, and structure length. Data for thermometers of length 130-447 nt is too large to represent in a table, hence is available in an Excel file.

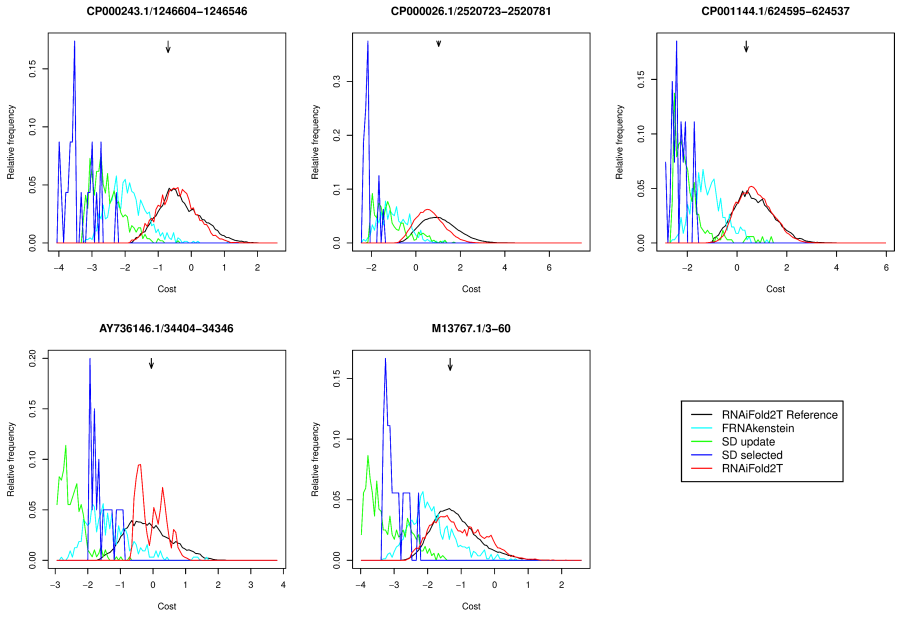

FIGURE 1. Relative histogram for the cost, as defined in equation (1) also appearing in equation (1) of the main text, for the solutions returned by `RNAiFold2T`, `SwitchDesign` and `Frnakenstein`, given target structure S_1 [resp. S_2] at temperature T_1 [resp. T_2] for λ phage CIII thermoregulators from Rfam family RF01804 the number of solutions returned for each method is indicated in column A of Table 1, which also gives EMBL accession codes. To produce a reference distribution, the black curve for `RNAiFold2T`. Reference was produced by running `RNAiFold2T` for several days. Remaining curves are for `Frnakenstein` (light green), `SwitchDesign` (dark green and purple) and `RNAiFold2T` (red). Arrows indicate the cost values for the real λ phage CIII thermoregulators from Rfam RF01804. This figure is similar to Figure 4 from the main text, except that the histograms of `SwitchDesign` and `Frnakenstein` are created from the output of these programs, *without* adding 1-point mutants that also fold into the target structures.

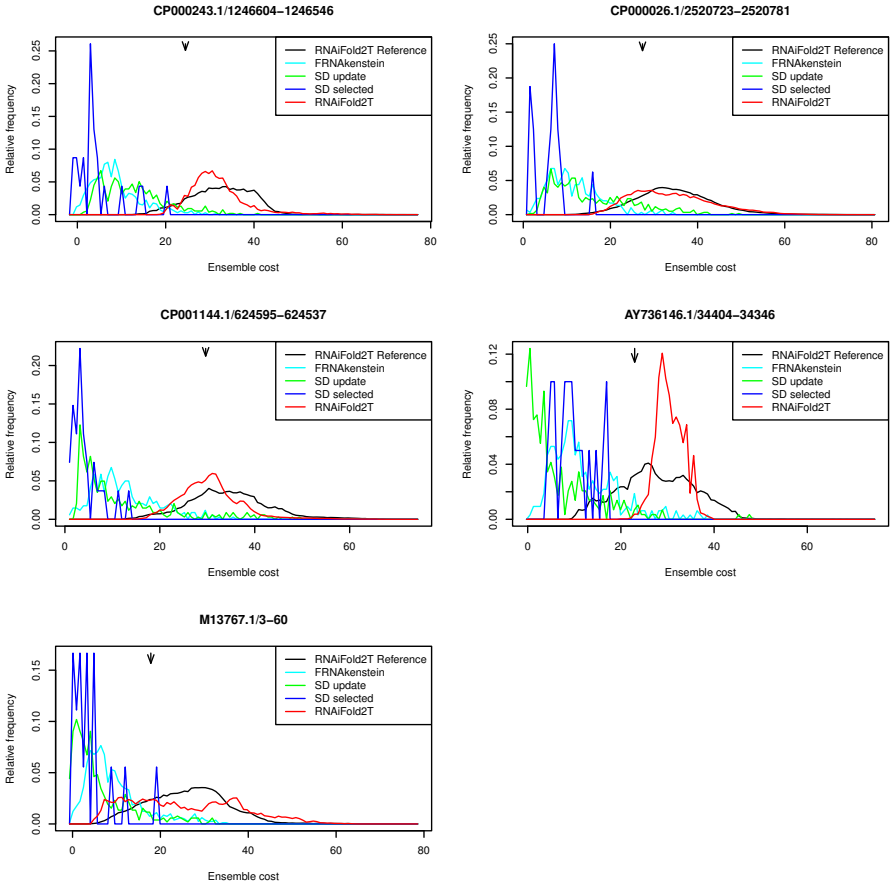

FIGURE 2. Relative histogram for ensemble defect cost, defined in equation (7), for the solutions returned by `RNAiFold2T`, `SwitchDesign` and `Frnakenstein`, given target structure S_1 [resp. S_2] at temperature T_1 [resp. T_2] for λ phage CIII thermoregulators from Rfam family RF01804 the number of solutions returned for each method is indicated in column A of Table 1, which also gives EMBL accession codes. To produce a reference distribution, the black curve for `RNAiFold2T`. Reference was produced by running `RNAiFold2T` for several days. Remaining curves are for `Frnakenstein` (light green), `SwitchDesign` (dark green and purple) and `RNAiFold2T` (red). Arrows indicate the cost values for the real λ phage CIII thermoregulators from Rfam RF01804. The Pearson correlation coefficient between `SwitchDesign` cost and ensemble defect based cost is 0.6545963

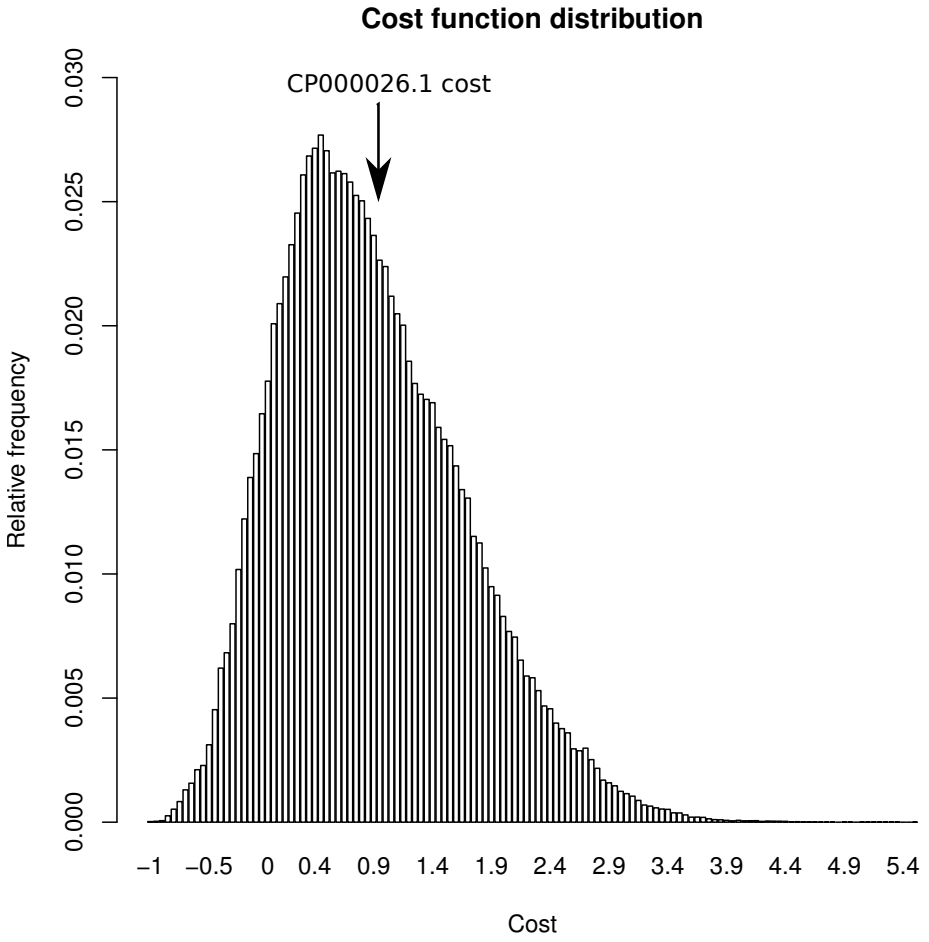

FIGURE 3. Distribution of the cost function for all possible solution sequences of lambda phage RNA [EMBL:CP000026.1/2520723-2520781] MFE structure at 32°C and 55°C. The arrow indicates the cost of the original sequence for the given structures.

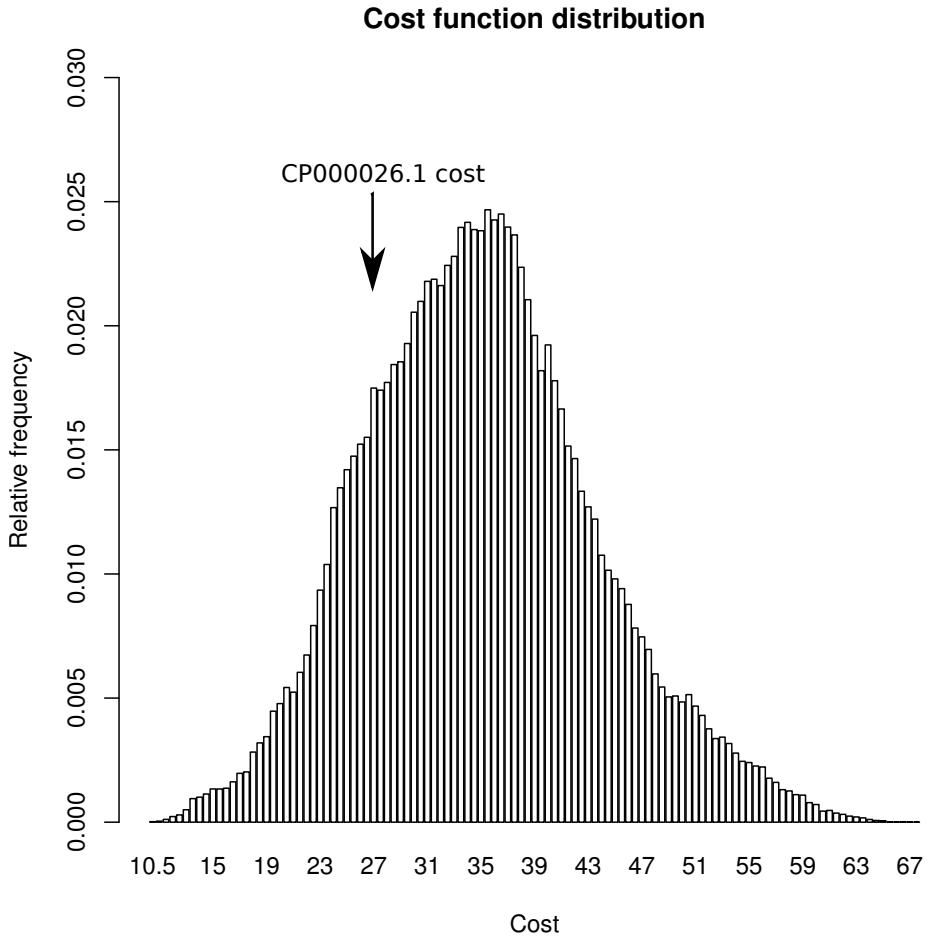

FIGURE 4. Distribution of the cost function based on ensemble defect for all possible solution sequences of lambda phage RNA [EMBL:CP000026.1/2520723-2520781] MFE structure at 32°C and 55°C. The arrow indicates the cost of the original sequence for the given structures.