AMB ALGORITHMS FOR
MOLECULAR BIOLOGY

RESEARCH   **Open Access**

RNA Accessibility in cubic time

Stephan H Bernhart[*], Ullrike Mückstein, Ivo L Hofacker

Abstract

Background: The accessibility of RNA binding motifs controls the efficacy of many biological processes. Examples are the binding of miRNA, siRNA or bacterial sRNA to their respective targets. Similarly, the accessibility of the Shine-Dalgarno sequence is essential for translation to start in prokaryotes. Furthermore, many classes of RNA binding proteins require the binding site to be single-stranded.

Results: We introduce a way to compute the accessibility of all intervals within an RNA sequence in $\mathcal{O}(n^3)$ time. This improves on previous implementations where only intervals of one defined length were computed in the same time. While the algorithm is in the same efficiency class as sampling approaches, the results, especially if the probabilities get small, are much more exact.

Conclusions: Our algorithm significantly speeds up methods for the prediction of RNA-RNA interactions and other applications that require the accessibility of RNA molecules. The algorithm is already available in the program RNAplfold of the ViennaRNA package.

Background

The importance of RNA within living cells has been realized in the last two decades. RNA provides a layer of regulation in eukaryotes, e.g. via miRNA, but also in prokaryotes via small RNAs (sRNAs) and riboswitches. Many of these regulatory functions are mediated by RNA interactions. These interactions are mainly realized through Watson-Crick or wobble base pairing between two RNA molecules. For the initialization of these interactions, a part of the interacting molecules has to be single-stranded. The tendency to be single-stranded is thus also important for the quality of putative target sites of miRNAs [1], siRNAs [2] and most probably sRNAs. Furthermore, the accessibilities of the Shine-Dalgarno sequence and the start codons are indicators of translational efficacy [3]. In addition, RNA accessibility will also influence the efficacy of single-strand binding proteins like *HuR* [4]. As it is not known how big exactly a putative target site is, and where it is located, it is best to know the accessibilities of all possible intervals within a RNA molecule. In particular, programs like RNAup [5,6] or IntaRNA [7] predict RNA-RNA interactions by computing a total binding energy $\delta G_{tot} = \delta G_{int} + \delta G_{open}$, composed of a stabilizing energy for the

intermolecular duplex δG_{int} and the cost of opening the binding sites δG_{open}. The opening energy can be computed from the accessibility, defined as the probability p^u that the binding site is unpaired in equilibrium, via $\delta G_{open} = -RT \ln(p^u)$.

The most naive approach to compute the accessibility of a certain stretch of bases is to use a constrained folding where no base pairs are allowed within a certain stretch of bases and dividing the respective restricted partition function by the unrestricted one. This is done for example in the miRNA target predictor PITA [8]. However, doing this for all n^2 possible intervals requires $\mathcal{O}(n^5)$ time. Ding and Lawrence [9] proposed to compute accessibilities by stochastically sampling structures from the Boltzmann ensemble. Sampling structures can be done in $\mathcal{O}(n^3)$, but necessarily introduces sampling errors which become large if the accessibilities get small, as is necessarily the case for longer regions. In [10], we introduced an algorithm that computes the accessibilities of all intervals of a given length l in cubic time. This leads to a $\mathcal{O}(n^4)$ algorithm when applied to intervals of all possible lengths. In addition, the algorithm could be used as a scanning algorithm that considers only local structures of a maximum length L and runs in $\mathcal{O}(nL^2l)$.

Here we introduce an algorithm to compute the accessibilities or single-strandedness of all intervals of

* Correspondence: berni@tbi.univie.ac.at
Theoretical Biochemistry group, Institute for theoretical chemistry, University of Vienna, Währingerstrasse 17, Vienna, Austria

an RNA molecule in $\mathcal{O}(n^3)$ time and $\mathcal{O}(n^2)$ memory. This is the same complexity as the algorithm to compute the partition function, which is underlying the sampling approaches, like sfold [11], commonly used for this task. However, the probabilities to be accessible for different intervals are not independent. Therefore, if one is interested in more complex questions, e.g. the joint probability that two intervals are accessible, this is no way around sampling the structures of the molecules.

The first predecessor of our algorithm, not yet fully $\mathcal{O}(n^3)$, has been used in RNAup [5]. The RNAplfold program of the ViennaRNA package [12] originally used the algorithm introduced in [10], but has been rewritten to use the efficient version of the algorithm presented below.

Since our algorithm is based on McCaskill's dynamic programming (DP) algorithm for the partition function [13], we briefly recapitulate the algorithm as implemented in the ViennaRNA package. As usual, we consider only (pseudo)knot-free secondary structures and use the Turner nearest-neighbor energy model [14]. In the inside (forward) recursion, the partition function $Q(i, j)$ of a sequence interval i, j is split into a part where i is unpaired and a part where i is paired:

$$Q(i, j) = \overbrace{Q(i + 1, j)}^{i\text{ is unpaird}} + \overbrace{\sum_{i < k \leq j} Q^B(i, k) Q(k + 1, j)}^{i\text{ is paired}} \quad (1)$$

Here, $Q^B(i, j)$ is the partition function for a stretch between i, j given that (i, j) form a base pair and $Q(a, b) = 1$ if $a \geq b$. A base pair can either close a hairpin loop, an interior loop (including bulge loops and stacks) or a multi (branch) loop.

$$Q^B(i, j) = \overbrace{\mathcal{H}(i, j)}^{\text{Hairpin}} + \overbrace{\sum_{i < k < l < j} \mathcal{I}(ij, kl) Q^B(k, l)}^{\text{Interior Loop}}$$
$$+ \overbrace{\sum_{i < k < j} Q^M(i + 1, u) Q^{M1}(u + 1, j - 1) a(i, j)}^{\text{Multi Loop}} \quad (2)$$

Here $\mathcal{H}(i, j)$ is short for the Boltzmann factor of a hairpin loop closed by the base pair (i, j): ($\exp^{-\beta E_{\text{hairpin}}}$, with $\beta = \dfrac{1}{RT}$), $\mathcal{I}(ij, kl)$ stands for the Boltzmann factor of the interior loop enclosed by the base pairs (i, j) and (k, l). For the multi loop contributions, $Q^M(i, j)$ holds the partition function for a part of a multi loop, and $Q^{M1}(i, j)$ are multi loop contributions that contain exactly one stem, where i belongs to the outermost base pair of this stem.

$$Q^M(i, j) = Q^M(i, j - 1)c + \sum_{i < u < j} (Q^M(i + 1, u) + c^{(u-i)}) Q^B(u + 1, j - 1)b$$
$$Q^{M1}(i, j) = Q^{M1}(i, j - 1)c + Q^B(i, j)b$$

$Q^{M1}(i, j)$ is needed to keep the multi loop decomposition unambiguous. The factors a, b and c are Boltzmann weighted contributions for closing a multi loop, adding a base pair or an unpaired base to the multi loop, respectively. Note that here, the size of the interior loops is limited to keep the algorithm cubic in time. See Figure 1 for a graphical representation of the recursion.

In the outside (backward) recursion (see Figure 2), the pair probabilities of a RNA molecule of length n, that is the quotient between the partition function of all states containing the base pair, $Q^{bp}(i, j)$ and the total partition function is computed.

$$p(i, j) = \frac{Q^{bp}(i, j)}{Q(1, n)} = \frac{Q^B(i, j) \hat{Q}^B(i, j)}{Q(1, n)} \quad (3)$$

Here, $Q^{bp}(i, j)$ is decomposed into an outer ($\hat{Q}^B(i, j)$) and an inner ($Q^B(i, j)$) partition function. The computation of the outer partition function of a base pair is split into two cases: The trivial case where no base pair is enclosing the base pair (i, j), and the case where there exists at least one base pair (k, l) with $k < i < j < l$.

$$\hat{Q}^B(i, j) = \overbrace{Q(1, i - 1) Q(j + 1, n)}^{\text{no enclosing base pair}} + \overbrace{\sum_{k < i < j < l} \mathcal{I}(kl, ij) \hat{Q}^B(k, l)}^{\text{Interior Loop}}$$
$$\left. \begin{array}{l} + \sum_{k < i < j < l} \hat{Q}^B(k, l) Q^M(k + 1, i - 1) Q^M(j + 1, l - 1) ab \\ + \sum_{k < i < j < l} \hat{Q}^B(k, l) Q^M(j + 1, l - 1) abc^{(i-k-1)} \\ + \sum_{k < i < j < l} \hat{Q}^B(k, l) Q^M(k + 1, i - 1) abc^{(l-j-1)} \end{array} \right\} \text{Multi Loop} \quad (4)$$

The interior loop contribution is again kept cubic by the size restriction of the interior loops. However, to keep the multi loop part of the algorithm $\mathcal{O}(n^3)$, we need to split the double sums over k and l into two sequential ($\mathcal{O}(n)$) sums with the help of two auxiliary arrays:

$$Q_B^M(i, l) = \sum_{1 \leq k < i} \hat{Q}^B(k, l) a Q^M(k + 1, i - 1)$$
$$Q_C^M(i, l) = \sum_{1 \leq k < i} \hat{Q}^B(k, l) a c^{(i-k-1)}$$

With these arrays, the multi loop part of Eqn 5 becomes:

$$\hat{Q}^B(i, j)[\text{Multi Loop}] = \sum_{j < l \leq n} Q_B^M(i, l) Q^M(j + 1, l - 1)b$$
$$+ \sum_{j < l \leq n} Q_C^M(i, l) Q^M(j + 1, l - 1)b$$
$$+ \sum_{j < l \leq n} Q_B^M(i, l) c^{(l-j-1)}b$$

Each entry of the arrays $Q_B^M(i, l)$ and $Q_C^M(i, l)$ can be computed in linear time. In fact, $Q_C^M(i, l) = Q_C^M(i - 1, l) c + \hat{Q}^B(i - 1, l)$ can be computed

Figure 1 Partition function folding. DP decomposition of RNA partition function folding. See text for detailed explanation.

Figure 2 Pair probability computation. Computation of pair probabilities. A single base pair (i, j) either lies within another base pair or not (First term on top). Base pair (i, j) and the enclosing base pair (k, l) either form an interior loop (rightmost term on top), or a multi loop. Based on the location of the other components within the multi loop, three possibilities for the closing of a multi loop exist. Here, the contributions collected in auxiliary array Q_B^M are colored green, in Q_C^M red.

recursively in constant time. The full probability computation is then:

$$p(i,j) = \frac{Q^B(i,j)}{Q(1,n)} \left(\overbrace{\frac{}{Q(1,i-1)Q(j+1,n)}}^{\text{Hairpin}} + \overbrace{\sum_{k<i<j<l} \mathcal{I}(kl.ij)\hat{Q}^B(k,l)}^{\text{Interior Loop}} + \right.$$
$$\left. + \underbrace{\sum_{j<l\le n} Q^M_B(i,l)c^{(l-j-1)} + \sum_{j<l\le n}(Q^M_B(i,l) + Q^M_C(i,l))Q^M(j+1,l-1)}_{\text{Multi Loop}} \right) \quad (5)$$

Algorithm

We now extend the McCaskill Algorithm to compute the probabilities $p^u(x, x + L)$ that a sequence interval $x \ldots x + L$ contains no paired bases. As can be seen in Figure 3, an unpaired stretch $x \ldots x + L$ is either not enclosed by a base pair, which again is the trivial case, or there is a base pair (i, j) enclosing the unpaired region such that $i < x < x + L < j$ (we call the sum of these contributions $Q^{ub}(x, x + L)$).

$$xp^u(x, x+L) = \frac{\overbrace{Q(1, x-1)Q(x+L+1, n)}^{\text{no enclosing base pair}} + Q^{ub}(x, x, +L)}{Q(1, n)}$$

The enclosing base pair (i, j) can again either close a hairpin, an interior loop or a multi loop.

$$Q^{ub}(x, x+L) = \sum_{i<x<x+L<j} \hat{Q}^B(i,j)\mathcal{H}(i,j)$$
$$+ \sum_{i<x<x+L<k<l<j} \hat{Q}^B(i,j)Q^B(k,l)\mathcal{I}(ij,kl)$$
$$+ \sum_{i<k<l<x<x+L<j} \hat{Q}^B(i,j)Q^B(k,l)\mathcal{I}(ij,kl)$$
$$+ \sum_{i<x<x+L<j} \hat{Q}^B(i,j)Q^M(i+1,x-1)Q^M(x+L+1,j-1)$$
$$+ \sum_{i<u<x<x+L<j} \hat{Q}^B(i,j)Q^M(i+1,u)Q^{M1}(u+1,x-1)$$
$$+ \sum_{i<x<x<L+u<j} \hat{Q}^B(i,j)\underbrace{Q^M(x+L+1,u)Q^{M1}(u+1,j-1)}_{Q^{M2}(x+L+1,j-1)}$$

To keep the computational complexity low, we use decompositions akin to the ones used for the pair probability and partition function computation. Hairpin terms are computed in two steps: First, we consider the case where the 5' base of the enclosing pair is directly adjacent to our region of interest:

$$Q^H(x, x+L) = \sum_{x+L<j} \mathcal{H}(x-1,j)\hat{Q}^B(x-1,j)$$

Suppose the enclosing pair starts at some position $i - 1 \le x - 1$, then the 3' end of the pair, j, has to be downstream of the unpaired region, i.e. at least at position
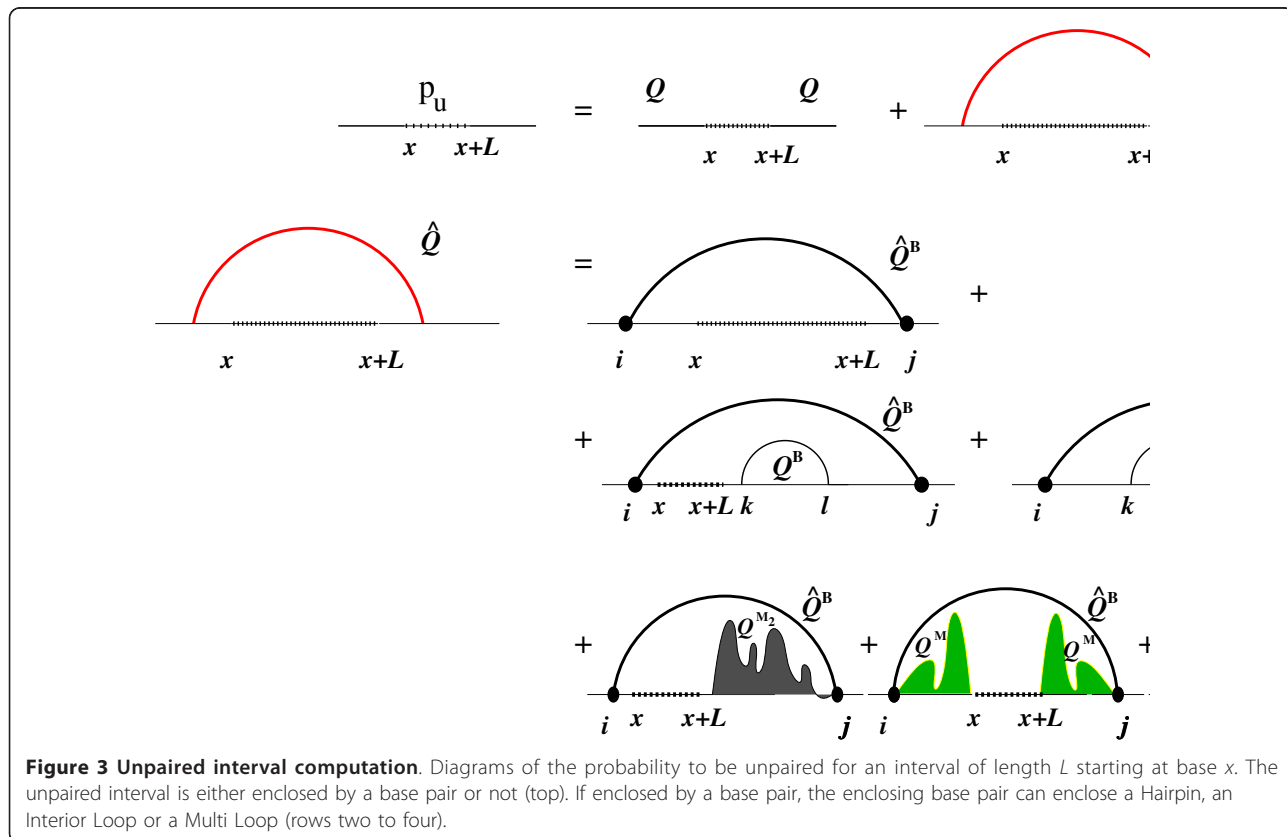

Figure 3 Unpaired interval computation. Diagrams of the probability to be unpaired for an interval of length L starting at base x. The unpaired interval is either enclosed by a base pair or not (top). If enclosed by a base pair, the enclosing base pair can enclose a Hairpin, an Interior Loop or a Multi Loop (rows two to four).

$x + L + 1$. This gives rise to a contribution $Q^H(i, x + L)$. Summing over i yields the complete partition function for the hairpin case:

$$Q^{ub}(x, x + L)[\text{Hairpin}] = \sum_{i \leq x} Q^H(i, x + L)) \tag{6}$$

For the interior loop contributions, we do a similar, but a little more complicated decomposition. We note that our region of interest can be located either in the 5' or 3' side of the interior loop and start by computing the partition function $Q^{I5}(i, k)$ $(Q^{I3}(l, j))$ over all structures in which $i \ldots k$ $(l \ldots j)$ forms the 5' (3') side of an interior loop (see Figure 3 row 3):

$$Q^{I5}(i, k) = \sum_{i < k < l < j} \mathcal{I}(ij, kl)\hat{Q}^B(i, j)Q^B(k, l)$$
$$Q^{I3}(l, j) = \sum_{i < k < l < j} \mathcal{I}(ij, kl)\hat{Q}^B(i, j)Q^B(k, l)$$

In analogy to Q^H in the hairpin case, $Q^I(x, x + L)$ sums over all cases where the position $x - 1$ preceding our region is paired. The final summation takes care of the cases where the interior loop starts before $x - 1$:

$$Q^{ub}(x, x + L)[\text{Interior}] = \sum_{i \leq x} Q^I(i, x + L)$$
$$Q^I(x, x + L) = \sum_{g > x+L} Q^{I5}(x - 1, g) + Q^{I3}(x - 1, g)$$

The multi loop contributions can be split into three parts, depending on where within the multi loop the unpaired interval is situated. If the unpaired interval is between the closing base pair and the first stem of the multi loop (i.e. at the 5' side of the multi loop), we compute:

$$Q^{ub}(x, x + L)[\text{Multi 5}] = \sum_{i \leq x} Q^{5M}(i, x + L)$$
$$Q^{5M}(x, x + L) = \sum_{x+L < j} \hat{Q}^B(x - 1, j)Q^{M2}(x + L + 1, j - 1)$$
$$Q^{M2}(x, j) = \sum_{x < u < j} Q^M(x, u)Q^{M1}(u + 1, j)$$

The terms for unpaired intervals that are located between the components of the multi loop and between the closing base pair and the last stem of the multi loop (3' side) are similar to each other:

$$Q^{ub}(x, x + L)[\text{Multi 3}] = \sum_{i \leq x} Q^{3M}(i, x + L)$$
$$Q^{3M}(x, x + L) = \sum_{x+L < j} \hat{Y}^M(x - 1, j)c^{(j-x-1)}$$
$$Q^{ub}(x, x + L)[\text{Multi between}] = \sum_{i \leq x} Q^{MB}(i, x + L)$$
$$Q^{MB}(x, x + L) = \sum_{x+L < j} (\hat{Y}^M(x - 1, j) + \hat{X}^M(x - 1, j))Q^M(x + L + 1, j - 1)c^{L+1}$$

Here, the contributions $\hat{X}^M(x - 1, j)$ and $\hat{Y}^M(x - 1, j)$ stand for the partition function over all partial structures where $x - 1$ is paired with some $g < x - 1$, closing a multi loop component (i.e. $(g, x - 1)$ is one of the interior pairs of the multi loop). The multi loop is closed by a pair (i, j), where $i < g$. The region between $x \ldots j - 1$ is a part of the multi loop not yet determined. In the case of $\hat{X}^M(x - 1, j)$, $x \ldots j - 1$ has to contain at least one other multi loop component. In $\hat{Y}^M(x - 1, j)$, at least one other multi loop component exists between g and i, so that no additional multi loop component is needed.

$$\hat{X}^M(x, j) = \sum_{1 < g < x} Q_C^M(g, j)Q^B(g, x)b$$
$$\hat{Y}^M(x, j) = \sum_{1 < g < x} Q_C^M(g, j)Q^B(g, x)b$$

Using the matrices from above, it is easy to see that the total computation for a single interval is linear in time:

$$p^u(x, x + L) = \frac{1}{Q(1, n)} \left(Q(1, x - 1)Q(x + L + 1, n) + \sum_{i \leq x} Q^H(i, x + L) + \sum_{i \leq x} Q^I(i, x + L) + \right.$$
$$\left. + \sum_{i \leq x} Q^{M5}(i, x + L) + \sum_{i \leq x} Q^{M3}(i, x + L) + \sum_{i \leq x} Q^{MB}(i, x + L) \right)$$

Implementation

The matrices needed for the computation of the accessibilities mostly contain terms that are computed anyway during the inside or outside recursions. These matrices can thus be committed to memory with little additional cost. The two contributions Q^{I5} and Q^{I3} are saved during the computation of the base pair probabilities in Eqn 4. The multi loop terms $\hat{X}^M(x - 1, j)$ and $\hat{Y}^M(x - 1, j)$ are also computed during the computation of the pair probabilities in Eqn 5, while Q^{M2} is saved during the forward recursion Eqn 2. The computation of the accessibilities is thus conveniently performed after the outside recursion. Some of the matrices needed for the computation of Q^{ub} can be computed recursively:

$$Q^H(x, x + L) = Q^H(x, x + L + 1) + \mathcal{H}(x - 1, x + L + 1)$$
$$Q^I(x, x + L) = Q^I(x, x + L + 1) + Q^{I5}(x - 1, x + L + 1) + Q^{I3}(x - 1, x + L + 1)$$
$$Q^{3M}(x, x + L) = Q^{3M}(x, x + L + 1) + \hat{Y}^M(x - 1, x + L + 1)$$

and thus require negligible additional computational costs. Due to the layout of this algorithm, it is easily possible to split the terms and e.g. find the probability of an interval to be within a hairpin loop or an interior loop. This can be useful if special types of RNA

interactions, for example kissing hairpin interactions, are to be considered.

Minimum free energy version

The minimum free energy (MFE) version of this algorithm can be used to compute one optimal structure that has an unpaired stretch $x \ldots x + L$ for every single interval $x \ldots x + L$. In principle, the computation of the optimal energy and the backtracking procedure of its structure are similar to the partition function version described above. Because ambiguity is of no concern, only one matrix for multi loop contributions has to be filled. Thus, after computing the "usual" matrices for RNA minimum free energy prediction (the minimum free energy $F(i, j)$, the minimum energy given that i and j form a base pair $C(i, j)$, and the minimum free energy for multi loop segments $M(i, j)$), we need to fill the following matrices:

$$M_2(i, j) = \min_{i<k<j} M(i, k) + M(k + 1, j)$$

$$I_3(i, j) = \min_{m<k<i} \hat{C}(j + 1, m) + \mathrm{I}(i - 1, k; m, j + 1)$$

$$I_5(i, j) = \min_{j<m, j<k<m} \hat{C}(m, i - 1) + \mathrm{I}(i - 1, m; j + 1, k)$$

$$M_M(i, j) = \min_{k<i<j} \hat{C}(k, j + 1) + M(k + 1, i - 1)$$

Here, I is the energy of an interior loop. The minimum energy $f_a(x, x + L)$ with an accessible interval between x and $x + L$ is then:

$$p^u(x, x+L) = \frac{1}{Q(1, n)} \left(Q(1, x-1)Q(x+L+1, n) + \sum_{i \leqslant x} Q^H(i, x+L) \right.$$
$$+ \sum_{i \leqslant x} Q^I(i, x+L) + \sum_{i \leqslant x} Q^{M5}(i, x+L)$$
$$\left. + \sum_{i \leqslant x} Q^{M3}(i, x+L) + \sum_{i \leqslant x} Q^{MB}(i, x+L) \right)$$

In our current implementation, computation of the structure on the outside of a base pair is done by doubling the sequence, doing a simple cofolding and backtracking $j, i + n$ for the base pair (i, j). This leaves room for improvement, as the memory consumption is twice as high as is strictly necessary. Backtracking the $f_a(i, j)$s gives the secondary structures to the energies.

As one possible application, the $f_a(i, n)$ subset of the secondary structures can be viewed as the minimum free energy structures during transcription of a RNA molecule. The unpaired interval in this case is regarded as the part of the molecule that is not yet transcribed. As an example, we show the ydhL Adenine riboswitch in Figure 4.
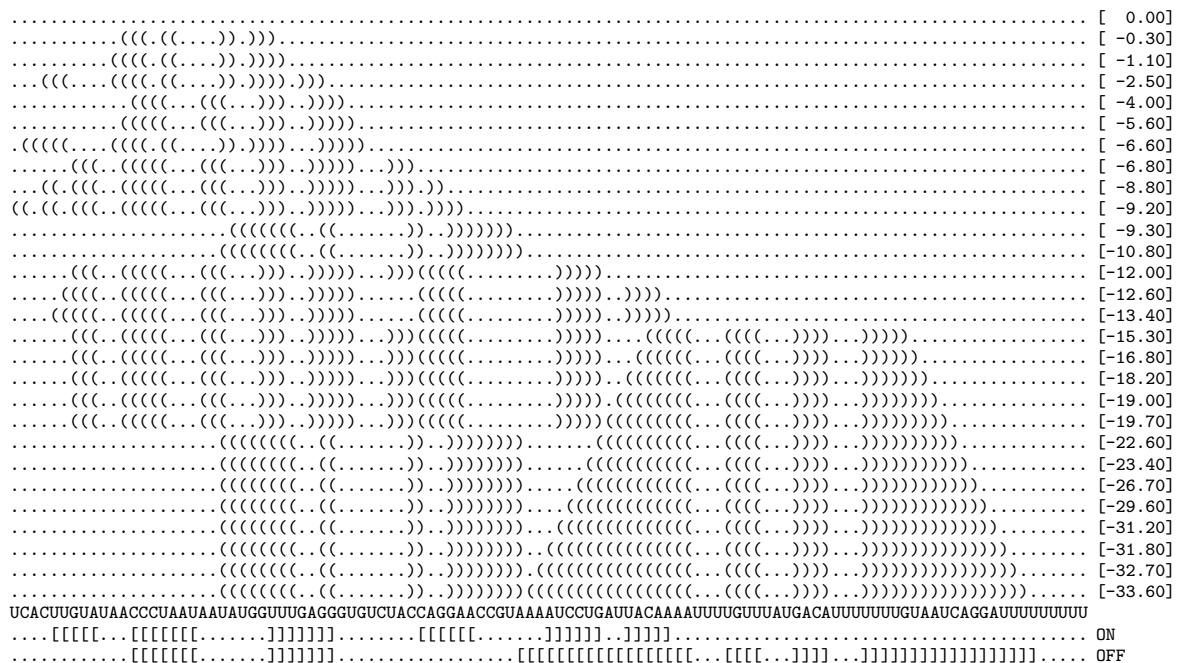
```
.......................................................................................................... [  0.00]
...........(((.((.(....)).)))).............................................................................. [ -0.30]
..........((((.((.(....)).))))).............................................................................. [ -1.10]
...(((....((((.((.(....)).))))).)))......................................................................... [ -2.50]
..........(((((...(((...))).))))).......................................................................... [ -4.00]
..........(((((...(((...))).)))))).......................................................................... [ -5.60]
.(((((....(((((.((.(....)).))))...))))).................................................................... [ -6.60]
......(((..(((((...(((...))).)))))...)))..................................................................... [ -6.80]
...((.(((..(((((...(((...))).)))))...))).)).................................................................. [ -8.80]
((.((.(((..(((((...(((...))).)))))...))).)))).............................................................. [ -9.20]
.....................(((((((..((.......))..)))))))).......................................................... [ -9.30]
.....................(((((((..((.......))..))))))))......................................................... [-10.80]
......(((..(((((...(((...))).)))))...)))(((((.........)))))................................................. [-12.00]
.....(((((..(((((...(((...))).)))))...)))))......(((((.........))))).)))).................................. [-12.60]
....((((((..(((((...(((...))).)))))...))))).....(((((.........)))))..))))).................................. [-13.40]
......(((..(((((...(((...))).)))))...)))(((((.........))))....(((((...(((...)))...)))))...................... [-15.30]
......(((..(((((...(((...))).)))))...)))(((((.........)))))..(((((...(((...)))...))))).................... [-16.80]
......(((..(((((...(((...))).)))))...)))(((((.........)))))..(((((...(((...)))...))))))).................. [-18.20]
......(((..(((((...(((...))).)))))...)))(((((.........)))))).(((((((...(((...)))...)))))))............... [-19.00]
......(((..(((((...(((...))).)))))...)))(((((.........)))))).(((((((...(((...)))...))))))))............. [-19.70]
.....................(((((((..((.......))..)))))))).......(((((((((...(((...)))...)))))))))............... [-22.60]
.....................(((((((..((.......))..))))))))......(((((((((...(((...)))...)))))))))............... [-23.40]
.....................(((((((..((.......))..))))))))....(((((((((((...(((...)))...)))))))))))............ [-26.70]
.....................(((((((..((.......))..))))))))....(((((((((((...(((...)))...)))))))))))........... [-29.60]
.....................(((((((..((.......))..))))))))...(((((((((((...(((...)))...)))))))))))........... [-31.20]
.....................(((((((..((.......))..)))))))).(((((((((((((...(((...)))...)))))))))))))......... [-31.80]
.....................(((((((..((.......))..)))))))).(((((((((((((...(((...)))...))))))))))))))........ [-32.70]
.....................(((((((..((.......))..))))))))(((((((((((((...(((...)))...))))))))))))))...... [-33.60]
UCACUUGAUAUAACCCUAAUAAUAAUGGUUUGAGGGUGUCUACCAGGAACCGUAAAAUCCUGAUUACAAAAUUUUGUUUAUGACAUUUUUUUGUAAUCAGGAUUUUUUUUU
....[[[[[...[[[[[[......]]]]]]]]]]........[[[[[......]]]]]..]]]]]]............................................ ON
...........[[[[[[......]]]]]]]................[[[[[[[[[[[[[[[[...[[[[...]]]]...]]]]]]]]]]]]]]]]]].... OFF
```

Figure 4 Structures of a growing RNA molecule. Minimum free energy structures of the growing chain of the ydhL Adenine riboswitch. The list of all optimal $f_a(i, n)$ structures is shown. The 3' dots correspond to the unpaired intervals $i \ldots n$. The bottom two (square bracketed) structures show the ON and the OFF state of the riboswitch taken from [16]. As can be seen, the ON state corresponds to an intermediate mfe structure (-13.40 kcal/mol, green 5' box), while the terminator loop is part of the overall minimum free energy structure (red 3' box).

Application

The method presented here is currently available in the RNAplfold program of the Vienna RNA package, versions 1.8.x and newer. In addition to global folding the program allows to compute accessibilities for local structures using a sliding window approach. In this case accessibilities for a region are averaged over all sequence windows containing this region. Technically, all that has to be done for the local folding version, is to replace the outside partition function $\hat{Q}^B$ by averaged versions $\hat{Q}^B$.

The computation of the $\hat{Q}^B$s has to reflect the fact that sub-sequences of different length will appear in a different number of windows. For example, using a window size W, the contribution of $\hat{Q}^\beta(k, l)$ to the computation of $\hat{Q}^\beta(i, j)$, $k < i < j < l$, would be:

$$
\begin{aligned}
\hat{Q}^\beta(i,j)[\text{enclosed by pair } (k,l)] &= \frac{W-(k-l)+1}{W-(i-j)+1}\hat{Q}^\beta(k,l)\mathcal{I}(kl,ij) \\
&+ \frac{W-(k-l)+1}{W-(i-j)+1}\hat{Q}^\beta(k,l)Q^M(k+1,i-1)Q^M(j+1,l-1)ab \\
&+ \frac{W-(k-l)+1}{W-(i-j)+1}\hat{Q}^\beta(k,l)Q^M(j+1,l-1)abc^{(i-k-1)} \\
&+ \frac{W-(k-l)+1}{W-(i-j)+1}\hat{Q}^\beta(k,l)Q^M(k+1,i-1)abc^{(l-j-1)}
\end{aligned}
$$

The $\hat{Q}^\beta(i, j)$ are subsequently also used in the computation of the accessibilities. This makes the program applicable to even the largest sequences, such as complete chromosomes or all mRNAs of an organism. Several programs, like RNAxs [2], IntaRNA [7], and RNAplex [15] are already using the accessibility computations implemented in RNAplfold with great success to rapidly predict accessibilities of putative target sites on mRNAs.

Authors' contributions
UM developed the first versions of the algorithm, SHB developed the current algorithm and integrated it into RNAplfold. ILH initiated and supervised the work. Regrettably, UM died during the preparation of this manuscript. She will always be remembered fondly. ILH and SHB jointly wrote and approved the manuscript.

Competing interests
The authors declare that they have no competing interests.

References
1. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition.** *Nat Genet* 2007, **39**(10):1278-84.
2. Tafer H, Ameres S, Obernosterer G, Gebeshuber C, Schroeder R, Martinez J, Hofacker I: **The impact of target site accessibility on the design of effective siRNAs.** *Nat Biotechnol* 2008, **26**(5):578-83.
3. Wikström P, Lind L, Berg D, Björk G: **Importance of mRNA folding and start codon accessibility in the expression of genes in a ribosomal protein operon of Escherichia coli.** *J Mol Biol* 1992, **224**(4):949-66.
4. Hackermüller J, Meisner N, Auer M, Jaritz M, Stadler P: **The effect of RNA secondary structures on RNA-ligand binding and the modifier RNA mechanism: a quantitative model.** *Gene* 2005, **345**:3-12.
5. Mückstein U, Tafer H, Hackermüller J, Bernhart S, Stadler P, Hofacker I: **Thermodynamics of RNA-RNA binding.** *Bioinformatics* 2006, **22**(10):1177-82.
6. Mückstein U, Tafer H, Bernhart SH, Hernandez-Rosales M, Vogel J, Stadler PF, Hofacker IL: **Translational Control by RNA-RNA Interaction: Improved Computation of RNA-RNA Binding Thermodynamics.** In *Bioinformatics Research and Development, Volume 13 of Communications in Computer and Information Science.* Edited by: Elloumi M, Küng J, Linial M, Murphy R, Schneider K, Toma C. Springer; 2008:114-127.
7. Busch A, Richter A, Backofen R: **IntaRNA: efficient prediction of bacterial sRNA targets incorporating target site accessibility and seed regions.** *Bioinformatics* 2008, **24**(24):2849-56.
8. Kertesz M, Iovino N, Unnerstall U, Gaul U, Segal E: **The role of site accessibility in microRNA target recognition.** *Nat Genet* 2007, **39**(10):1278-84.
9. Ding Y, Lawrence CE: **Statistical prediction of single-stranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond.** *Nucleic Acids Res* 2001, **29**(5):1034-46.
10. Bompfünewerer A, Backofen R, Bernhart S, Hertel J, Hofacker I, Stadler P, Will S: **Variations on RNA folding and alignment: lessons from Benasque.** *J Math Biol* 2008, **56**(1-2):129-44.
11. Ding Y, Lawrence C: **A statistical sampling algorithm for RNA secondary structure prediction.** *Nucleic Acids Res* 2003, **31**(24):7280-301.
12. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer S, Tacker M, Schuster P: **Fast Folding and Comparison of RNA Secondary Structures.** *Monatsh Chemie* 1994, **125**(2):167-188.
13. McCaskill JS: **The Equilibrium Partition Function and Base Pair Binding Probabilities for RNA Secondary Structure.** *Biopolymers* 1990, **29**:1105-1119.
14. Turner DH, Mathews DH: **NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure.** *Nucleic Acids Res* 2010, , **38** Database: D280-2.
15. Tafer H, Amman F, Eggenhofer F, Stadler PF, Hofacker IL: **Fast Accessibility-Based Prediction of RNA-RNA Interactions.** *Bioinformatics* 2011.
16. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ: **Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs.** *Chem Biol* 2004, **11**(12):1729-1741[http://www.hubmed.org/display.cgi?uids=15610857].