Memory Efficient Folding Algorithms for Circular RNA Secondary Structures

Ivo L. Hofacker[1] and Peter F. Stadler[2,1,3,]
[1]Institute for Theoretical Chemistry, University of Vienna,
Währingerstrasse 17, A-1090 Vienna, Austria
[2]Bioinformatics Group, Department of Computer Science,
and Interdisciplinary Center for Bioinformatics,
University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany.
[3]The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, New Mexico

Abstract: A small class of RNA molecules, in particular the tiny genomes of viroids, are circular. Yet most structure prediction algorithms handle only linear RNAs. The most straightforward approach is to compute circular structures from "internal" and "external" substructures separated by a base pair. This is incompatible, however, with the memory-saving approach of the `Vienna RNA Package` which builds a linear RNA structure from shorter (internal) structures only. Here we describe how circular secondary structures can be obtained without additional memory requirements as a kind of "post-processing" of the linear structures.
Keywords: RNA secondary structure, circular RNA, dynamic programming, viroids

1 Introduction

Most RNA molecules are linear. Circular single stranded RNAs, on the other hand, occur only in a few cases. The most prominent class are the "genomes" of viroids, see [FDG+04, TT04] for recent reviews. A related example is the circular RNA genome of Hepatitis Delta virus which contains a viroid-like domain, see e.g. [GCT04, WB02] and the references therein. In addition, alternative splicing may lead to circular RNAs from intronic sequences. This appears to be a general property of nuclear group I introns [NFB+03] and was also observed during tRNA splicing in *H. volcanii* [SSGG03]. Circularized C/D box snoRNAs were recently reported in *Pyrococcus furiosus* [SMJ+04]. Circular nucleic acids, furthermore, have been investigated in the context of *in vitro* selection experiments [KZG+02].

While structure prediction of these fairly rare circular RNAs may appear as a rather esoteric topic, most of the examples above have functional secondary structures. Indeed, viroids were among the first RNAs for which secondary structures have been studied systematically [SHF+84], see also [RWR+99] for more recent work. Since viroid RNAs are short (approx. 200-400 nucleotides), we have to expect significant differences between the folds of linear and circular sequences, see Fig. 2.

It is therefore worthwhile to develop circular variants of at least the most common RNA

folding tools; indeed algorithms for computing minimum energy folding and the computation of suboptimal structure of circular RNAs are implemented in Michael Zuker's `mfold` package [Zuk89, Zuk03]. These algorithms, in fact, treat linear RNAs as exceptional variants of the circular ones. In contrast, the `Vienna RNA Package`[1] [HFS+94, Hof03], optimizes the memory requirements for linear RNAs; this approach saves approximately a factor of 2 in memory as well as some CPU time. Circular RNAs, however, are nontrivial to handle in this framework. In this contribution we demonstrate how circular RNA folding can be implemented efficiently as a kind of "post-processing" step of the forward recursion and as a corresponding "pre-processing" step for the the backtracking part of the folding algorithms without requiring significant additional resources or a redesign of the optimized recursion for the linear RNA case. Circular RNA folding can therefore be included into the `Vienna RNA Package` without duplicating the code or compromising the efficiency of the current implementations.

This contribution is organized as follows: We briefly recall the RNA folding algorithms as implemented in the `Vienna RNA Package`. We then discuss the extension of the minimum free energy folding approach to circular RNAs and describe how the same ideas apply to the computation of the partition function.

2 Folding Linear RNA Molecules

The energy model for RNA folding is based upon carefully measured energy parameters [MSZT99, MDC+04] for the loops of the RNA secondary structure (i.e., the cycles of the unique minimum cycle basis [LS98]). The energy of a loop depends on the sequence near the base pairs that are part of the loop, the length of the loop, and on its type. From the biophysical point of view one distinguishes hairpin loops, stacked base pairs, bulges, true interior loops, and multi(branched) loops. From an algorithmic point of view one can treat bulges, stacked pairs, and true interior loops as subtypes of interior loops.

We consider an RNA sequence x of length n. Hairpin loops are uniquely determined by their closing pair k, l. The energy of a hairpin loop is

$$\mathcal{H}(k,l) = \mathcal{H}(x_k, x_{k+1}, \ell, x_{l-1}, x_l)$$

where ℓ is the length of the loop (expressed as the number of its unpaired nucleotides). Each interior loop is determined by the two base pairs enclosing it. Its energy is tabulated as

$$\mathcal{I}(k,l;p,q) = \mathcal{I}(x_k, x_{k+1}; \ell_1; x_{p-1}, x_p; x_q, x_{q+1}; \ell_2; x_{l-1}, x_l)$$

where ℓ_1 is the length of unpaired strand between k and p and ℓ_2 is the length of the unpaired strand between q and l. Symmetry of the energy model dictates $\mathcal{I}(k,l;p,q) = \mathcal{I}(q,p;l,k)$. If $\ell_1 = \ell_2 = 0$ we have a (stabilizing) stacked pair, if only one of ℓ_1 and ℓ_2 vanish we have a bulge. For multiloops, finally, we have an additive energy model with three parameters a, b, c of the form $\mathcal{M} = a + b \times \beta + c \times \ell$ where ℓ is the length of multiloop

[1] Available at `http://www.tbi.univie.ac.at/RNA/`

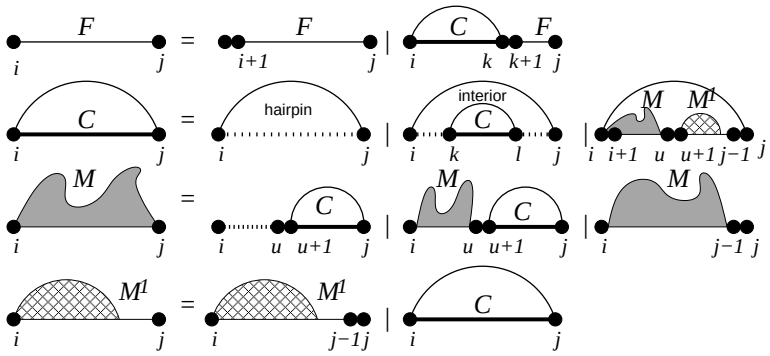

Figure 1: Decomposition of secondary structures underlying the folding algorithms as implemented in the `Vienna RNA Package`. Top: a structure on $[i, j]$ starts either with an unpaired base or with a paired 5' base. 2nd row: A structure enclosed in a base pair is either a hairpin loop, delimited by an interior loop, or branches in a multiloop. The multiloop itself is composed of two parts, one with one or more components (M) and another with exactly one component (M^1). The last two rows further depict the recursions for the two types of multiloop components. Again, the decompositions are into disjoint sets of cases.

(again expressed as the number of unpaired nucleotides), β is the number of branches, not counting the branch in which the closing pair of the loop resides.

RNA folding algorithms are based on decomposing the set of possible structures into sets of smaller structures. This decomposition can be chosen such that each possible structure appears in exactly one of the subcases. In the course of the "normal" RNA folding algorithm for linear RNA molecules as implemented in the `Vienna RNA Package` [HFS+94, Hof03] the following arrays, which correspond to different structural components in Fig. 1, are computed for $i < j$:

- F_{ij} free energy of the optimal substructure on the subsequence $x[i, j]$.
- C_{ij} free energy of the optimal substructure on the subsequence $x[i, j]$ subject to the constraint that i and j form a basepair.
- M_{ij} free energy of the optimal substructure on the subsequence $x[i, j]$ subject to the constraint that $x[i, j]$ is part of a multiloop and has at least one component, i.e., a sub-sequence that is enclosed by a base pair.
- M^1_{ij} free energy of the optimal substructure on the subsequence $x[i, j]$ subject to the constraint that that $x[i, j]$ is part of a multiloop and has exactly one component, which has the closing pair i, h for some h satisfying $i \leq h < j$.

The "conventional" energy minimization algorithm for linear RNA molecules [ZS81,ZS84] can be summarized in the following way, which corresponds to the recursions implemented

in the `Vienna RNA Package` [HFS+94, Hof03]:

$$
\begin{aligned}
F_{ij} &= \min\left\{F_{i+1,j},\ \min_{i<k\le j} C_{ik} + F_{k+1,j}\right\}\\
C_{ij} &= \min\left\{\mathcal{H}(i,j),\ \min_{i<k<l<j} C_{kl} + \mathcal{I}(i,j;k,l),\ \min_{i<u<j} M_{i+1,u} + M^1_{u+1,j-1} + a\right\}\\
M_{ij} &= \min\left\{\min_{i<u<j}(u-i-1)c + C_{u+1,j} + b,\ \min_{i<u<j} M_{i,u} + C_{u+1,j} + b,\ M_{i,j-1} + c\right\}\\
M^1_{ij} &= \min\left\{M^1_{i,j-1} + c,\ C_{ij} + b\right\}
\end{aligned}
\tag{1}
$$

These recursions are directly derived from the structure decomposition shown in Fig. 1. The corresponding recursions for the partition function are obtained by replacing minimum operations with sums and additions with multiplications [McC90].

The computation of the minimum free energy structure requires to store only the arrays F, C, and M. In addition, the full M^1 array is required for the more elaborate backtracking procedure of the `RNAsubopt` program [WFHS99] which produces all RNA secondary structures within a given energy interval above the ground state. Similarly, uniqueness of the decomposition is necessary for partition function algorithms, see Sect. 4.

3 Folding Algorithms for Circular RNAs

A straightforward way of dealing with circular RNA molecules is to compute C_{ij} and M_{ij} also for the subsequences of the form $x[j,n]x[1,i]$. This is implemented in the `mfold` package [Zuk03] and described e.g. in [Zuk89]. The disadvantage of this approach is, however, that it doubles the memory requirements (and also the CPU requirements, because more matrix entries need to be computed).

As an alternative, we propose here to extend the linear folding algorithms in such a way that the circular molecules are handled as a kind of "post processing" of the arrays that are computed in the linear case. This is not only memory efficient but also allows us to assess the structural differences between linear and circular sequences with just a single run of the forward recursions. (Recall that the backtracking step for minimum energy folding is fast: $\mathcal{O}(n)$ compared to the $\mathcal{O}(n^3)$ steps for filling the arrays.)

The key observation is that the only difference between the linear and the circular case is the energy of the loop that contains x_n and x_1. In the linear case, there is no energy contribution associated with the "exterior" loop, while it has to be scored like any other loop in the circular case. Hence we have to distinguish the types of "exterior" loops.

Exterior Hairpin. If the exterior loop is a hairpin, then there is a base pair p, q, $1 \le p < q \le n$ such that both $x[1, p-1]$ and $x[q+1, n]$ are unpaired. The optimal energy of such a structure is

$$
F^{\circ}_H = \min_{p<q}\left\{C_{pq} + \mathcal{H}(q,p)\right\} \tag{2}
$$

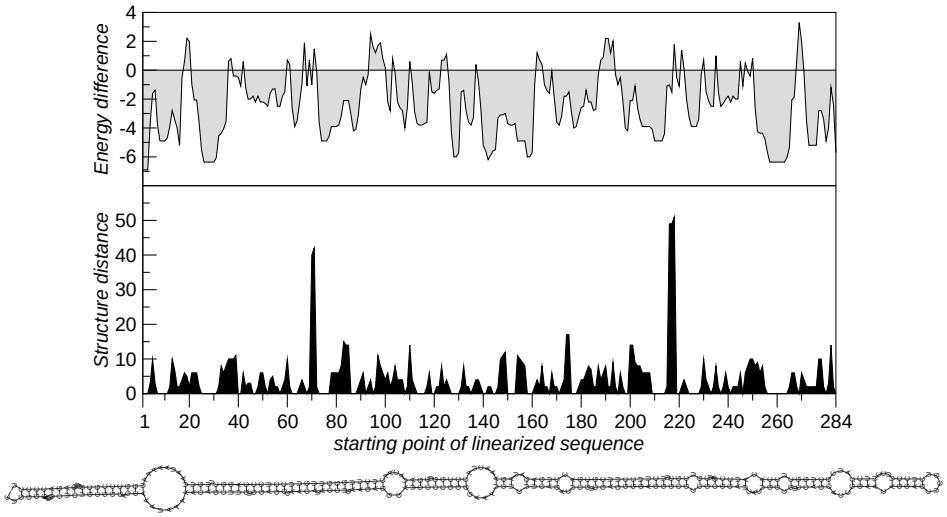

Figure 2: Differences between linear and circular folds of the Citrus Viroid IV (***Acc. No. X14638***) [PRL$^+$91] as a function of cut point in the sequence (relative to the database entry). Structure distance is measured as Hamming distance of the dot-parenthesis strings, differences in folding energy in kcal/mol. Below, the correct circular is shown.

where $\ell = p - 1 + (n - q + 1)$ is length of the hairpin loop and $1 \leq p < q \leq n$.

Exterior Interior Loop. In this case, the "exterior loop" contains the closing pairs k, l and p, q of exactly two components. Thus

$$F_I^\circ = \min_{k<l<p<q} \{C_{pq} + C_{kl} + \mathcal{I}(q, p, l, k)\}$$

where $\ell_1 = n - q + k - 1$ and $\ell_2 = p - l - 1$. In practice, the size $\ell = \ell_1 + \ell_2$ of an interior loop is limited to $\ell \leq m$, typically $m = 30$. Thus F_I° can be computed in $\mathcal{O}(n^3)$ time without additional memory requirements.

Exterior Multi-Loop. Generalizing the approach for the interior loops, we can view an exterior multiloop as a multi-loop with at least 3 branches on the sequence interval form 1 to n. Starting from M_{ij}^1 we compute the linear auxiliary array M_{kn}^2 containing the optimal energy of $x[k, n]$ given that the sequence interval is contained in a multiloop, has exactly *two* components, and starts with a base pair k, h. We obtain

$$M_{kn}^2 = \min_{k<u<n} \left(M_{ku}^1 + M_{u+1,n}^1\right) \tag{3}$$

This array requires only $\mathcal{O}(n)$ memory and can be computed in $\mathcal{O}(n^2)$ time. A multiloop with at least 3 components can now be constructed from a piece with at least one component at the beginning of the sequence and a piece that contains exactly two components (with first closing pair $k + 1, v$, for some $k < v < n - 2$):

$$F_M^\circ = \min_{1<k<n} \left\{M_{1,k} M_{k+1,n}^2 + a\right\} \tag{4}$$

The multiloop case thus can be dealt with in quadratic time with only linear memory overhead.

The minimum free energy structure of the folded circular molecule is therefore

$$F^\circ = \min\{F_H^\circ, F_I^\circ, F_M^\circ\} \tag{5}$$

Backtracking. Backtracking is straightforward with this approach: First we determine whether the optimal "exterior loop" is a hairpin ($F^\circ = F_H^\circ$), an interior loop ($F^\circ = F_I^\circ$), or a multiloop ($F^\circ = F_M^\circ$). Depending on the result we determine either

(1) p, q such that $F_H^\circ = C_{pq} + \mathcal{H}(q, p)$, or

(2) k, l and p, q such that
$F_I^\circ = C_{pq} + C_{kl} + \mathcal{I}(q, p; l, k)$, or

(3) (a) k such that $F_M^\circ = M_{1,k} + M^2_{k+1,n} + a$, and then
(b) u such that $M^2_{kn} = M^1_{k,u} + M^1_{u+1,n}$.

The next step already follows the normal backtracking procedure of the linear folding problem.

Dangling Ends. The `Vienna RNA Package` implements three different models for handling the so-called dangling-end contributions that arise when an unpaired nucleotide stacks with an adjacent base pair. These contributions can be (a) ignored, (b) taken into account for every combination of adjacent bases and base pairs, or (c) a more complex model can be used in which the unpaired base can stack with at most one base pair. The latter model strictly speaking violates the secondary structure model in that an unpaired bases x_i between two base pairs (x_p, x_{i-1}) and (x_{i+1}, x_q) has three distinct states with different energies: x_i does not stack to its neighbors, x_i stacks to x_{i-1}, or x_{i+1}. The algorithm then minimizes over these possibilities. In cases (a) and (b) one can absorb the dangling end contributions in the loop energies. In case (c), however, they have to be treated explicitly, which is done in the forward recursions already for all cases with the exception of the dangling end contribution reaching across the "gap" 1-n. The cases unpaired x_1 stacks to paired x_n and unpaired x_n stacks to paired x_1 need to be treated separately, adding two additional sub-cases to the multi-loop recursion above. Even more sub-classes are needed if one wants to allow also for co-axial stacking of helices in the multi-loop.

An important observation about the recursions (2-4) is that each possible secondary structure is counted exactly once, i.e., the recursions are non-redundant. This is important when one is interested in enumerating structures as e.g. in the `RNAsubopt` program. This property is also crucial for the partition function calculations discussed in the next section. For the purpose of energy minimization, however, it is not necessary. One can therefore replace eq.(3) by

$$M^2_{kn} = \min_{k<u<n} (M_{ku} + M_{u+1,n}) \tag{6}$$

and reinterpret M^2 as the contribution of segments with *at least* two branches in a multiloop. As a consequence, the M^1 array does not need to be stored and the memory requirements of the minimum free energy folding are the same as in the linear case up to a the auxiliary array M^2 of size n.

4 Partition Function

It is straightforward to translate the recursions (2-4) into recursions for the partition function because they already provide a partition of the set of all secondary structures that can be formed by the sequence x. In the following we suppress the factor $1/RT$ in the Boltzmann factors of the energy parameters, i.e., we assume that the energy parameters are already scaled relative to the thermal energy. Eq.(2-4) then become

$$\begin{aligned} Z^{M2}_{kn} &= \sum_u Z^{M1}_{ku} Z^{M1}_{u+1,n} \\ Z^{\circ}_{H} &= \sum_{p<q} Z^{B}_{pq} e^{-\mathcal{H}(q,p)} \\ Z^{\circ}_{I} &= \sum_{k<l<p<q} Z^{B}_{kl} Z^{B}_{pq} e^{-\mathcal{I}(k,l,p,q)} \\ Z^{\circ}_{M} &= \sum_{k} Z^{M}_{1,k} Z^{M2}_{k+1,n} e^{a} \\ Z^{\circ} &= Z^{\circ}_{H} + Z^{\circ}_{I} + Z^{\circ}_{M} \end{aligned} \tag{7}$$

The probability P_{kl} of a base pair kl can be represented, in the simplified version of the Nussinov algorithm [NPGK78], as

$$P_{kl} = P^{\circ}_{kl} + \sum_{p<k;q>l} P_{pq} \frac{Z_{p+1,k-1} Z^{B}_{k,l} Z_{k+1,q-1}}{Z^{B}_{pq}} e^{-x} \tag{8}$$

see Fig. 3. Here P°_{kl} is the probability of that kl is a closing pair contained in the exterior loop. This is the only term that differs from the linear case. For the full energy model we can use the same logic, but we need to consider the individual loop types separately. In detail we obtain [McC90]:

$$\begin{aligned} P_{kl} = P^{\circ}_{kl} + \sum_{p<k;q>l} P_{pq} \frac{Z^{B}_{k,l}}{Z^{B}_{p,q}} \Bigg\{ & e^{-\mathcal{I}(p,q,k,l)} \\ & + \left(\sum_{p<u<k} Z^{M}_{p+1,u} Z^{M1}_{u+1,k-1}) \right) e^{-(a+(q-l-1)c)} \\ & + \left(\sum_{l<u<q} Z^{M}_{l+1,u} Z^{M1}_{v+1,q-1}) \right) e^{-(a+(k-p-1)c)} \\ & + Z^{M}_{p+1,k-1} Z^{M}_{l+1,q-1} \Bigg\} \end{aligned} \tag{9}$$

The first term covers the case where p, q and k, l delimit an interior loop. The remaining three terms cover the multi-loop case with the three sub-cases that kl delimits the most 3', the most 5', or an intermediate branch, respectively.

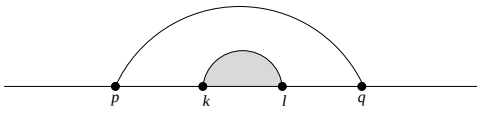

Figure 3: Backward recursion. In order to compute P_{kl} we have to consider all configurations in which the pair kl is immediately interior to a pair pq. This basepair in turn is formed with probability P_{pq}.

The contribution P°_{kl} covers the cases in which the basepair kl is part of the "exterior" loop. In the linear case we have simply

$$P^{\mathrm{lin}}_{kl} = \frac{Z_{1,k-1} Z^B_{kl} Z_{k+1,n}}{Z_{1n}} \tag{10}$$

In the circular case we have to consider the three possible loop types for the "exterior" loop separately. This yields:

$$\begin{aligned} P^{\circ}_{kl} = \frac{Z^B_{kl}}{Z^{circ}} \Bigg\{ & \underbrace{e^{-\mathcal{H}(l,k)}}_{\text{hairpin}} \\ & + \underbrace{\sum_{p,q:p<q<k<l} Z^B_{pq} e^{-\mathcal{I}(q,p,l,k)}}_{\text{Interior left}} + \underbrace{\sum_{p,q:k<l<p<q} Z^B_{pq} e^{-\mathcal{I}(p,q,l,k)}}_{\text{Interior right}} \\ & + \underbrace{Z^M_{1,k-1} Z^M_{l+1,n} e^{-a}}_{\text{Multi middle}} + \underbrace{\sum_{j<k} Z^M_{1j} Z^{M1}_{j+1,k-1} e^{-(a+(n-q)c)}}_{\text{Multi left}} \\ & + \underbrace{\sum_{j>l} Z^{M1}_{l+1,j} Z^M_{j+1,n} e^{-(a+(k-1)c)}}_{\text{Multi right}} \Bigg\} . \end{aligned} \tag{11}$$

For given k and l this expression can be evaluated in linear time without additional memory requirements. It follows that the base pairing probability matrix P° for the case of circular RNAs can be computed with a constant additional factor in CPU time and negligible additional memory requirements.

5 Concluding Remarks

Circular RNA folding is being added as an additional feature to the `Vienna RNA Package`. The energy minimization is already available via `cvs`, the implementation of the circular version of `RNAalifold` [HFS02] is in progress. This tool computes the consensus structure of a set of aligned RNA sequences. Algorithmically, it is very similar to the energy minimization described above.

The main applications for these features are a more systematic analysis of viroid structures and circular snoRNAs. In conjunction with alignment algorithms for circular sequences [GT93, Mae90] one can use circular `RNAalifold` to obtain consensus structures. The `alidot` tool [HFF+98, HS99] can be applied without changes to the problem of identifying evolutionarily conserved RNA secondary structure motifs in otherwise structurally variable RNA motifs. The circular version of `RNAsubopt` [WFHS99] will be of particular interest for a detailed understanding of the structural changes in viroid RNAs.

Acknowledgements. This work was supported in part by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Project No. P15893, and by the German *DFG* Bioinformatics Initiative BIZ-6/1-2.

References

[FDG+04] Ricardo Flores, Sonia Delgado, María-Eugenia Gas, Alberto Carbonell, Diego Molina, Selma Gago und Marcos De la Peña. Viroids: the minimal non-coding RNAs with autonomous replication. *FEBS Lett.*, 567:42–48, 2004.

[GCT04] S. O. Gudima, J. Chang und J. M. Taylor. Features affecting the ability of hepatitis delta virus RNAs to initiate RNA-directed RNA synthesis. *J. Virol.*, 78:5737–5744, 2004.

[GT93] J. Gregor und M. G. Thomason. Dynamic programming alignment of sequences representing cyclic patterns. *IEEE Trans. Patt. Anal. Mach. Intell.*, 15:129–135, 1993.

[HFF+98] Ivo L. Hofacker, Martin Fekete, Christoph Flamm, Martijn A. Huynen, Susanne Rauscher, Paul E. Stolorz und Peter F. Stadler. Automatic Detection of Conserved RNA Structure Elements in Complete RNA Virus Genomes. *Nucl. Acids Res.*, 26:3825–3836, 1998.

[HFS+94] Ivo L. Hofacker, Walter Fontana, Peter F. Stadler, Sebastian Bonhoeffer, Manfred Tacker und Peter Schuster. Fast Folding and Comparison of RNA Secondary Structures. *Monatsh. Chemie*, 125(2):167–188, 1994.

[HFS02] Ivo L. Hofacker, Martin Fekete und Peter F. Stadler. Secondary Structure Prediction for Aligned RNA Sequences. *J. Mol. Biol.*, 319:1059–1066, 2002.

[Hof03] Ivo L. Hofacker. Vienna RNA secondary structure server. *Nucl. Acids Res.*, 31:3429–3431, 2003.

[HS99] Ivo L. Hofacker und Peter F. Stadler. Automatic Detection of Conserved Base Pairing Patterns in RNA Virus Genomes. *Comp. & Chem.*, 23:401–414, 1999.

[KZG+02] X. D. Kong, S. Z. Zhu, X. J. Gou, X. P. Wang, H. Y. Zhang und J. Zhang. A circular RNA-DNA enzyme obtained by in vitro selection. *Biochem. Biophys. Res. Commun.*, 292:1111–1115, 2002.

[LS98] Josef Leydold und Peter F. Stadler. Minimal Cycle Basis of Outerplanar Graphs. *Elec. J. Comb.*, 5:209–222 [R16: 14 p.], 1998. See `http://www.combinatorics.org/` R16 and Santa Fe Institute Preprint 98-01-011.

[Mae90] M. Maes. On a cyclic string-to-string correction problem. *Inform. Process. Lett.*, 35:73–78, 1990.

[McC90] John S. McCaskill. The Equilibrium Partition Function and Base Pair Binding Probabilities for RNA Secondary Structure. *Biopolymers*, 29:1105–1119, 1990.

[MDC+04] David H. Mathews, Matthew D. Disney, Jessica L. Childs, Susan J. Schroeder, Michael Zuker und Douglas H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction ofRNA secondary structure. *Proc. Natl. Acad. Sci. USA*, 101:7287–7292, 2004.

[MSZT99] D.H. Mathews, J. Sabina, M. Zuker und H. Turner. Expanded Sequence Dependence of Thermodynamic Parameters Provides Robust Prediction of RNA Secondary Structure. *J. Mol. Biol.*, 288:911–940, 1999.

[NFB+03] H. Nielsen, T. Fiskaa, A. B Birgisdottir, P. Haugen, C. Einvik und S. Johansen. The ability to form full-length intron RNA circles is a general property of nuclear group I introns. *RNA*, 9:1464–1475, 2003.

[NPGK78] Ruth Nussinov, George Piecznik, Jerrold R. Griggs und Daniel J. Kleitman. Algorithms for Loop Matching. *SIAM J. Appl. Math.*, 35(1):68–82, 1978.

[PRL+91] H Puchta, K Ramm, R. Luckinger, R. Hadas, M. Bar-Joseph und Sänger H L. Primary and secondary structure of citrus viroid IV (CVd IV), a new chimeric viroid present in dwarfed grapefruit in Israel. *Nucl. Acids Res.*, 19:6640, 1991.

[RWR+99] D. Repsilber, S. Wiese, M. Rachen, A. W. Schroder, D. Riesner und G. Steger. Formation of metastable RNA structures by sequential folding during transcription: time-resolved structural analysis of potato spindle tuber viroid (-)-stranded RNA by temperature-gradient gel electrophoresis. *RNA*, 5:574–584, 1999.

[SHF+84] G. Steger, H. Hofmann, J. Fortsch, H. J. Gross, J W Randles, H L Sanger und D. Riesner. Conformational transitions in viroids and virusoids: comparison of results from energy minimization algorithm and from experimental data. *J. Biomol. Struct. Dyn.*, 2:543–571, 1984.

[SMJ+04] Natalia G. Starostina, Sarah Marshburn, L. Steven Johnson, Sean R. Eddy, Rebecca M. Terns, und Michael P. Terns. Circular box C/D RNAs in *Pyrococcus furiosus*. *Proc. Natl. Acad. Sci. USA*, 101:14097–14101, 2004.

[SSGG03] Shilpa R. Salgia, Sanjay K. Singh, Priyatansh Gurha und Ramesh Gupta. Two reactions of *Haloferax volcanii* RNA splicing enzymes: Joining of exons and circularization of introns. *RNA*, 9:319–330, 2003.

[TT04] M. Tabler und M. Tsagris. Viroids: petite RNA pathogens with distinguished talents. *Trends Plant Sci.*, 9:339–348, 2004.

[WB02] T. S. Wadkins und M. D. Been. Ribozyme activity in the genomic and antigenomic RNA strands of hepatitis delta virus. *Cell Mol Life Sci.*, 59:112–125, 2002.

[WFHS99] S. Wuchty, W. Fontana, I. L. Hofacker und P. Schuster. Complete Suboptimal Folding of RNA and the Stability of Secondary Structures. *Biopolymers*, 49:145–165, 1999.

[ZS81] M. Zuker und P. Stiegler. Optimal computer folding of larger RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Research*, 9:133–148, 1981.

[ZS84] M. Zuker und D. Sankoff. RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46:591–621, 1984.

[Zuk89] M. Zuker. On finding all suboptimal foldings of an RNA molecule. *Science*, 244:48–52, 1989.

[Zuk03] Michael Zuker. `Mfold` web server for nucleic acid folding and hybridization prediction. *Nucl. Acids Res.*, 31:3406–3415, 2003.