Squares in Cycles: Prediction of G-Quadruplexes in Circular RNA Secondary Stuctures

Ronny Lorenz[1,2], **Peter F. Stadler**[1−6]

[1] Bioinformatics Group, Department of Computer Science and
Interdisciplinary Center for Bioinformatics,
Universität Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany,

[2]Institute for Theoretical Chemistry,
University of Vienna, Vienna, Austria

[3]Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany

[4]Center for Non-Coding RNAs in Technology and Health, University of Copenhagen, Denmark

[5]Faculty of Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia

[6]Santa Fe Institute, Santa Fe, NM

{ronny,studla}@bioinf.uni-leipzig.de

Abstract. *RNA secondary structures, determined by non-crossing base pairs, capture many of the salient features of RNA molecules, explain the free energy of structure formation very accurately, and can be computed efficiently given the sequence information only. G-quadruplexes are compact local structures that have been shown to have important biological function and can be integrated into secondary structure prediction. In recent years circular RNAs have gained considerable interest as a biological relevant subclass of RNAs. While algorithms and tools are available that extend secondary structure prediction from linear to circular RNAs, no support is provided for G-quadruplexes. In this contribution we close this gap and describe how the* `ViennaRNA` *package has been extended to include this increasingly relevant case.*

1. Introduction

RNA G-quadruplexes (GQs) are non-canonical structures forming stable four-stranded conformations composed of stacks of guanine tetrads that form Hoogsteen hydrogen bonds. In contrast to DNA, RNA GQs appear to fold into a monomorphic parallel conformation [Zhang et al. 2010, Pandey et al. 2013]. Specialized sequencing approaches such as rG4-seq [Kwok et al. 2016] and G4RP-seq [Yang et al. 2022], makes it possible to map GQs at transcriptome-wide scales [Kwok et al. 2016]. The `G4Atlas` [Yu et al. 2023] collects GQs at transcriptome-wide scales.

Surprisingly, GQs are globally unfolded *in vivo* in many eukaryotic cells counteracting their formation *in vitro* [Guo and Bartel 2016]. Nevertheless, their association with untranslated regions and non-coding RNAs [Li and Zhou 2023], as well as evidence for negative selection [Lee et al. 2020] suggest regulatory functions. In *Arabidopsis* and *Oryza*, an enrichment of G-quadruplexes with two G-quartets and a function in translation regulation have been reported [Yang et al. 2020]. GQs

have been implicated in the regulation of splicing, transcription, and chemical modifications, see [Fay et al. 2017, Dumas et al. 2021, Lyu et al. 2021, Cueny et al. 2022, Sahayasheela and Sugiyama 2023] for recent reviews.

Circular RNAs have been a topic of intensive research during the last decade. By far the most abundant subclass, usually referred to as circRNAs, is produced by back-splicing from many eukaryotic transcripts. Current versions of dedicated databases report millions of circRNAs. A recent careful benchmarking study showed, moreover, that circRNAs are identified reliably in RNA-seq data [Vromman et al. 2023]. The secondary structures of circRNAs appear to play an important role in particular in their interactions [Liu et al. 2019].

GQs have been available as a feature in RNA secondary structure prediction in the `ViennaRNA` package for more than a decade [Lorenz et al. 2012, Lorenz et al. 2013]. The prediction of circular RNAs has also been a long-standing feature in `ViennaRNA` package [Hofacker and Stadler 2006]. However, due to a lack of demand, the combination of the two capabilities has not been implemented so far. Although there is at present no direct evidence for a functional link between circular RNAs and GQs, this is a plausible hypothesis because of the impact of GQs on splicing regulation and the role of splicing in circRNA biogenesis. In order to enable a systematic study of GQs in circular RNAs, we report here on an improved implementation of the of GQ-related features in `ViennaRNA`.

2. Theory and Implementation

2.1. Recursions for Folding Circular RNAs with G-Quadruplexes

RNA secondary structures can be computed exactly by means of dynamic programming given the loop-based standard energy model, which distinguishes hairpin loops, interior loops (including stacked base pairs), and multi-branched loops [Turner and Mathews 2010]. For the latter, a linear approximation is used that depends on both the branching degree, i.e., the number of helices emanating from it and the number of unpaired bases. A detailed benchmarking showed that more sophisticated energy models for multiloops does not improve structure prediction [Ward et al. 2017]. GQs, by definition, are self-enclosed local elements that cover a contiguous sequence interval. Consequently, they can be treated in the same way as substructures enclosed by a base pair when embedded in a larger secondary structure context. In other words, GQs can be incorporated into folding recursion as a straightforward extension [Lorenz et al. 2012]. More precisely, it suffices to introduce extra cases in which a GQ replaces as substructure that is enclosed by a base pair. Fig. 1 summarizes these extensions to the core recursions.

RNA folding distinguishes three types of loops: hairpin loops, in which the closing pair (i, j) is connected by an unpaired stretch of sequence, interior loops, in which the closing pairs (i, j) and a unique base pair (k, l) immediately interior to (i, j) with $i < k < l < j$ are connected by unpaired stretches between i and k, and l and j, respectively, and multiloops in which there are two or more base pairs immediately interior to the closing pair (i, j). Each of these pairs delimits a component of the multiloop.

As noted above, GQs can be formally treated just like components, i.e., a GQ from k to l corresponds to a base pair (k, l) immediately interior to the closing pair. Thus, GQs are appended as additional structural alternative wherever the recursions in Fig. 1

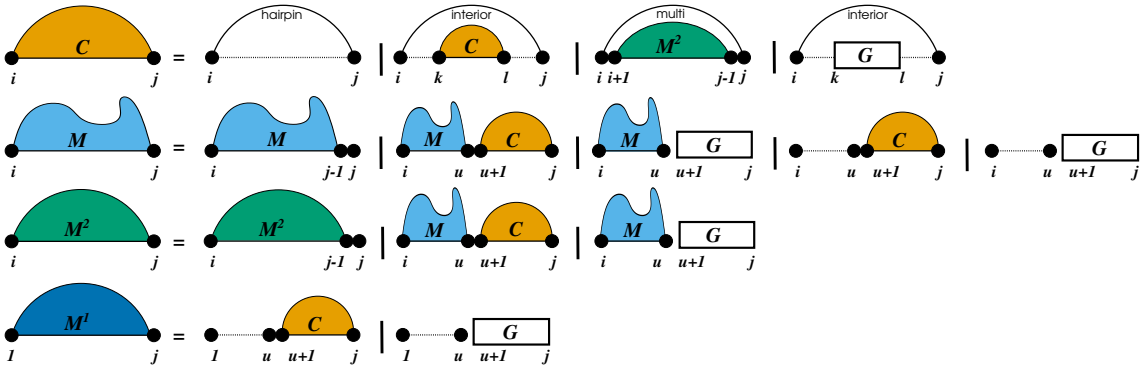

Figure 1. Grammar for the decomposition of "loops". Every structure enclosed by a base pair, C, is either a hairpin loop, an interior loop, or a multi-branch loop. In addition, we need to consider an interior-loop-like conformation in which the inner structure in the loops is a GQ. The linear energy model for multiloops requires two auxiliary variables, referring to arbitrary substructures of a multiloop, M, and to substructures of multiloops with at least two components, M^2. The auxiliary variable M^1 referring to a substructure with exactly one component (last line) is only required for evaluating the external loop, which is treated separately in Fig. 2 below.

explicitly call for contributions deriving from a single component. In particular, a GQ may take on the role of the inner substructure in an interior loop (first line). Similarly, GQs may replace single-component parts within multiloops thus leading to additional cases in the recursions for unconstrained multiloops (M) and multiloop components with at least two substructures M^2. As a consequence, the GQ contributions are accounted for in the C, M, and M^2 arrays. We note that this part of the grammar in Fig. 1 is slightly different from (but equivalent to) the implementation of GQs as described in [Lorenz et al. 2013]. We shall return to this point below.

Linear and circular structures differ only with respect to the *exterior loop*, i.e., the part of the structure containing positions 1 and n. In the case of linear RNAs, the exterior loop is considered unconstrained and itself incurs no energy contribution. That is, it can be decomposed stepwisely, into an unpaired base, a substructure enclosed by a base pair, or a GQ, followed by the remainder of the exterior loop. In grammatical form, this can be expressed as $F \rightarrow \bullet F \mid CF \mid GF$, where $\bullet$ denotes unpaired base, C is a substructure enclosed by a base pair, and G denotes a GQ.

In the circular case, however, the exterior loop, F° contains the covalent bond between the last nucleotide n and the first nucleotide 1 in the input sequence, which we will refer to as $n, 1$-junction. Incorporating GQs entails several additional cases to retain sufficient flexibility in the scoring. We distinguish two cases: (1) decompositions where the $n, 1$-junction is exposed and (2) decompositions where the $n, 1$-junction is covered by a quadruplex.

For case (1), there are eight alternatives, the first four of which arise in the same form in circular folding algorithms without GQs [Hofacker and Stadler 2006], namely the completely unpaired structure, the hairpin, interior loop, and multiloop case. The latter is handled here differently from earlier grammars. Since an exterior multiloop has at least three components, it can be composed of single component described by M^1 array in Fig. 1 and a part with at least two components, M^2 (see below). The remaining four cases

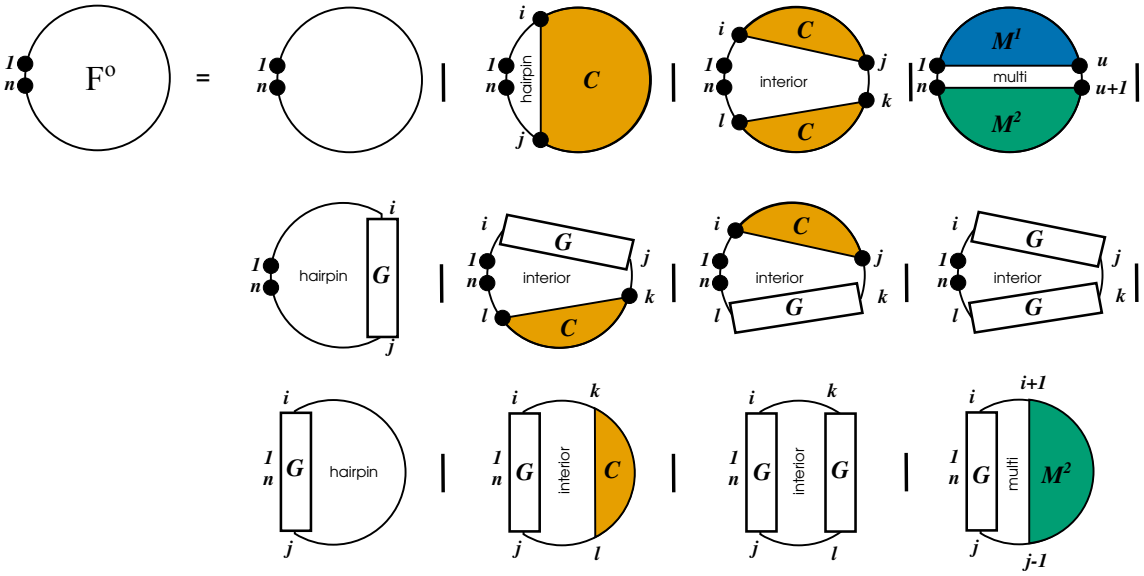

Figure 2. Grammar for the decomposition of the exterior loop of circular structure with GQs. The first four cases are the unpaired structure, and the hairpin, interior loop, and multiloop case. The next four cases explicitly describe structures exclusively composed of one or two GQs not overlapping the $n, 1$-junction. The remaining four alternatives deal with a quadruplex covering the $n, 1$-junction, distinguishing the hairpin, interior loop and multiloop cases.

correspond to a hairpin- and interior-loop-like arrangement comprising structures that are unpaired except for one or two quadruplexes.

The last four alternatives pertain to structures in which the $n, 1$-junction is covered by a GQ. These do not have analogs in either linear folding with GQs or circular folding without GQs. Again, however, they correspond to the three loop types. There are two variants for the interior-loop-like case, since the interior component is either a structure enclosed by a base pair, C, or GQ. In the multiloop case, the GQ covering the $n, 1$-junction counts as a component, thus the remainder of the sequence is covered by multiloop part with at least two components, M^2.

The full set of recursion equations is shown in Tab. 1 for the computation of the minimum free energy structure. Each variable denotes the optimal energy over all structures of a particular type over the sequence interval indicated by the indices. For example, C_{ij} is the energy of the optimal secondary structure on the subsequence from position i to j subject to the condition that (i, j) forms a base pair. A value of $+\infty$ appears whenever a structure is impossible, i.e., $C_{ij} = \infty$ if the bases at position i and j cannot form a base pair. All terms that involve GQs are highlighted in color. Throughout, we write $G_{p,q}$ for the free energy of a GQ with first base p and last base q. The optimal energy of the complete structure is denoted by F°. For ease of presentation, is is given as the minimum over the four alternatives F°_E, F°_H, F°_I, and F°_M depending on the type of the exterior loop. The cases with exposed and covered $n, 1$-junction are collected by loop type in the recursion equations.

Analogous expression for the corresponding partition functions are obtained by replacing the minima by sums, the sums by multiplications, and parameter values such as $\mathcal{H}(i, j)$ by the corresponding Boltzmann factor $\exp(-\mathcal{H}(i, j)/RT)$. Outside recursions

Table 1. Complete recursion for minimum energy folding of circular RNAs with G-quadruplexes. See text for explanation of the variables; calligraphic symbols $\mathcal{H}$, $\mathcal{I}$ refer to energy parameters, superscript G indicates values for G-quadruplexes. Entries in color are extensions due to the inclusion of G-quadruplexes.

$$C_{i,j} = \min \begin{cases} \mathcal{H}(i,j), \\ \min_{i<k<l<j} \left(\mathcal{I}(i,j,k,l) + C_{k,l}\right), \\ a + b + M^2_{i+1,j-1} \\ \min_{i<k<l<j} \left(\mathcal{I}^G(i,j,k,l) + G_{k,l}\right) \end{cases} \tag{1}$$

$$M^2_{i,j} = \min \begin{cases} M^2_{i,j-1} + c, \min_{i<u<j} M_{i,u} + C_{u+1,j} + b, \\ \min_{i<u<j} M_{i,u} + G_{u+1,j} + b \end{cases} \tag{2}$$

$$M_{i,j} = \min \begin{cases} M_{i,j-1} + c, \\ \min_{i<u<j} \left(M_{i,u} + C_{u+1,j} + b\right), \\ \min_{i\le u<j} \left((u-i)\cdot c + C_{u,j} + b\right), \\ \min_{i<u<j} \left(M_{i,u} + G_{u+1,j} + b\right), \\ \min_{i\le u<j} \left((u-i)\cdot c + G_{u,j} + b\right) \end{cases} \tag{3}$$

$$M^1_j = \min \begin{cases} M_{j-1} + c, \min_{1\le i<j} \left((i-1)\cdot c + C_{i,j}\right) \\ \min_{1\le i<j} \left((i-1)\cdot c + G_{i,j}\right) \end{cases} \tag{4}$$

$$F^\circ = \min\left\{F^\circ_E, F^\circ_H, F^\circ_I, F^\circ_M\right\} \tag{5}$$

$$F^\circ_E = \Delta G_{cyc} \tag{6}$$

$$F^\circ_H = \min \begin{cases} \min_{1\le i<j\le n} \left(\mathcal{H}(j,i) + C_{i,j}\right) \\ \min_{1\le i<j\le n} \left(\mathcal{H}^G(i,j) + G_{i,j}\right), \min_{1<i\le n, 1\le j<i} \left(\mathcal{H}^G(i,j) + G_{i,j}\right) \end{cases} \tag{7}$$

$$F^\circ_I = \min \begin{cases} \min_{1\le i<j<k<l\le n} \left(\mathcal{I}(i,j,k,l) + C_{i,j} + C_{k,l}\right) \\ \min_{1\le i<j<k<l\le n} \left(\mathcal{I}(i,j,k,l)^G + G_{i,j} + C_{k,l}\right) \\ \min_{1\le i<j<k<l\le n} \left(\mathcal{I}(i,j,k,l)^G + C_{i,j} + G_{k,l}\right) \\ \min_{1\le i<j<k<l\le n} \left(\mathcal{I}(i,j,k,l)^G + G_{i,j} + G_{k,l}\right) \\ \min_{1<i\le n, 1\le j<i, j<k<l<i} \left(\mathcal{I}^G(i,j,k,l) + G_{i,j} + C_{i,j}\right) \\ \min_{1<i\le n, 1\le j<i, j<k<l<i} \left(\mathcal{I}^G(i,j,k,l) + G_{i,j} + G_{i,j}\right) \end{cases} \tag{8}$$

$$F^\circ_M = \min \begin{cases} \min_{1<u<j} \left(a + M^1_u + M^2_{u+1,n}\right), \\ \min_{1<i\le n, 1\le j<i, j<k<l<i} \left(a + b + G_{i,j} + M^2_{j+1,i-1}\right) \end{cases} \tag{9}$$

for the partition functions, furthermore, can be derived from the "inside grammar" displayed in Figs. 1 and 2 as described in [Höner zu Siederdissen et al. 2015]. Due to space restrictions, we omit the details here. Importantly, the grammar in Figs. 1 and Fig. 2 is unambiguous, i.e., each structure has a unique parse. This is a key requisite for the correct computation of partition functions or the counting of structures [Giegerich 2000].

2.2. Implementation Details

Version 2.7.0 of the `ViennaRNA` package implements both the minimum free energy and the partition function version of circular RNA folding with GQs for single sequences and multiple sequence alignments. Beyond the presentation above, so-called dangling end and terminal mismatch contributions are considered as in other components of the `ViennaRNA` package, see [Lorenz et al. 2011] for details.

There are some changes in the RNA folding grammar compared to previous implementations, which however generate the same language of RNA structures, and thus give the same results. Previously, multiloop contributions were stored in two triangular arrays: (i) an array $M^1_{i,j}$ referring to parts of a multiloop starting with a base pair at position i and covering exactly one component, and (ii) an array $M_{i,j}$ referring to a substructure comprising at least one component. Instead of $M^1_{i,j}$, we now use the array $M^2_{i,j}$ describing a multiloop component with at least two components. This variant of the folding recursions was introduced in [Bernhart et al. 2011] in the context of efficient accessibility computations and will replace the previous decompositions of most of algorithms implemented in `ViennaRNA`. The M^1 array is required here only as a contribution to the exterior loop. Since only entries of the form $M^1_{1,j}$ are used here, a linear array suffices. This modification on the grammar thus does not come at a noticeable addition in memory cost.

The inclusion of GQs spanning the $n, 1$-junction introduces a complication that affects the output format. So far, GQs were presented as four runs of + signs in ASCII output. In linear sequences this can be easily parsed. In the circular case, however, it leads to ambiguities since the start of the first GQ can no longer be identified unambiguously. We therefore extended the representation by introducing a dedicated *stop* character, the tilde sign ~, denoting the most 3'-base of the GQ, i.e. its actual end. This stop character is mandatory for a GQ that spans the $n, 1$-junction, but is optional for GQs anywhere else in the structure.

2.3. Computational Complexity

By definition, GQs are local structural elements with finite size of at most g. The additional effort in each of the interior-loop-like case in Fig. 1 is therefore bounded by $O(ng)$ assuming that arbitrary lengths and position is allowed. As in the other RNA folding algorithms, unpaired stretches in interior loops are limited to a total length of 30. Since $g \leq 65$, this limits the effort for this case to a constant. The other terms in Fig. 1 only require $O(g)$ effort for each of the $O(n^2)$ matrix entries.

For the exterior loop the addition terms with explicit quadruplexes require $O(ng)$, $O(n^2g^2)$, $O(g^2)$, $O(g^2n^2)$, $O(g^2ng)$ and $O(g^2)$ effort, where the factor $O(g^2)$ accounts for the placement of the GQ that covers the $n, 1$-junction. Overall, the complexity in both space and time is the same as the usual RNA folding recursions.

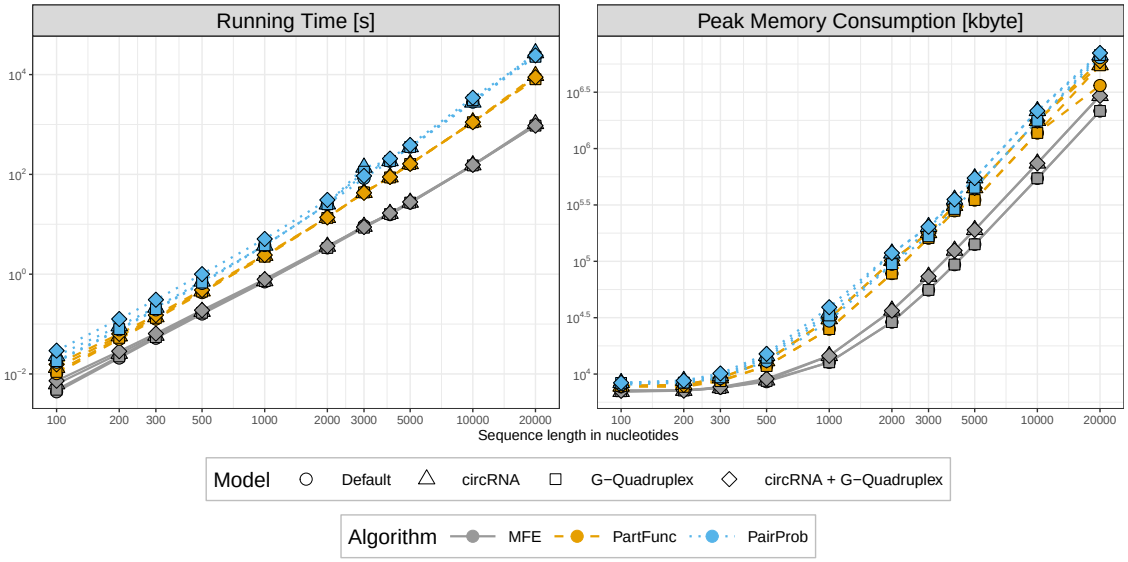

Figure 3. Comparison of `RNAfold` running time and memory requirements for random input sequences of different lengths. Shown are predictions of MFE, partition function and base pair probabilities for linear RNAs (Default), circular RNAs (circRNA), linear RNAs with GQ-support (G-Quadruplex), and circular RNAs with GQ-support (circRNA + G-Gquadruplex).

The `ViennaRNA Package` already supports GQs for linear RNAs and implements predictions without GQ support for circular RNA sequence input. When compared against these existing implementations, the novel support for circular RNAs with GQs as described above does not impose any mentionable overhead in terms of running time and memory requirements (see Fig. 3). The additional cases in the grammar required to compute probabilities for GQs and base pairs that reside in a loop with at least one GQ (red entries in Tab. 1), however, result in a small but noticeable increase in running time. Nevertheless, this overhead for circular RNAs virtually vanishes for longer sequences. Our transition to use the M^2 matrix instead of the previously implemented decomposition scheme even renders computations slightly faster compared to versions prior to $2.7.0$ (data not shown). For linear RNA sequences, we make use of a proper interleaving scheme for the recursions such that only two lines of the quadratic M^2 matrix have to be held in memory. In contrast, structure prediction for circular RNAs (with GQs) actually does require the full M^2 and M^1 matrices, hence some memory overhead can be observed in these cases.

2.4. Energy Parameters

The inclusion of GQs requires an extension of the standard energy model in two respects. First, energies need to be assigned to the GQs themselves. As in [Lorenz et al. 2012, Lorenz et al. 2013], we use a very simple energy function of the form $G = a(L-1) + b\ln(\ell - 2)$, where L is the number of G-quartets (tetrads) and ℓ is the total length of the three linker sequences. The default values $a = -18$ kcal/mol and $b = 12$ kcal/mol were obtained by fitting the data from [Zhang et al. 2011].

Due to the lack of available data we currently limit the number of stacking layers (G-quartets) to at least two and at most five. The linker lengths connecting the individual runs of Gs can adopt values from 1 to at most 15. This imposes the following limits to the

minimum and maximum span of a GQ: $11 \leq g \leq 65$.

In order to assign energies for loops that include GQs as an interior component we used the following approximations: Multiloop-like cases are scored in the same way as multiloops, i.e., a GQ is treated like any other component of the multiloop. In cases where a GQ is the only component enclosed by a base pair, i.e. the interior loop-like case, a (stabilizing) terminal mismatch contribution for the enclosing base pair and its adjacent nucleotides as well as an entropic penalty for the number of unpaired bases surrounding the GQ is added, similar to regular interior loops composed of two base pairs.

For circular RNAs, two more special configurations arise, where the structure does not consist of any regular base pairs but exactly one or two GQs. We score both similar to the hairpin- and interior-loop-like cases, respectively, but omit the terminal mismatch contribution(s) due to the lack of any supporting data.

Finally, we introduce a change in the energy computation for circular RNAs. The structure without base pairs conceptually corresponds to the open chain in the linear case. Biophysically, the latter is interpreted as a random coil and used as reference state with free energy 0. In the circular case, however, the ends are constrained by the $n,1$-junction. Jacobson-Stockmeyer theory implies an asymptotic entropy for circularization of $S_{cyc} \sim -3/2 \ln n$. Corrections for steric hindrance can be taken into account in lattice models by setting $S_{cyc} \sim \ln p_n/w_n$, where p_n and w_n are the the number of self-avoiding polygons and walks, respectively, which yields prefactors different from $3/2$, see e.g. [Freed 2012] and the references therein. A pragmatic estimate is taken from equ.(6) in [Kimchi et al. 2019],

$$\Delta G_{cyc} = RT\left(\alpha_0 + (3/2)\ln n\right) \ . \tag{10}$$

For the additive constant, a value of $\alpha_0 \approx 6.33$ has been estimated from simulations using hairpins. We suspect that this is an overestimate for the entropy loss due to the rigid geometry of the closing base pair. In order to obtain a more plausible estimate we assume that a circular sequence of length 12 can form a helix of length 3 that is at least marginally stable compared to the unpaired structure. The most stable among such structures has an energy of about $+5.0$ kcal/mol in the standard energy model. We therefore choose $\alpha_0 = 4.385$ kcal/mol, which shifts the unpaired circle of length 12 to 5.0 kcal/mol.

3. Examples and Applications

Due to the lack of circRNA sequences with known GQs we resort here to an artificial example to demonstrate the new features of the `ViennaRNA` package. The precursor of the microRNA miRNA-92B is known to harbor a three-layer GQ that regulates miRNA maturation in dependence of the potassium ion (K^+) concentration [Arachchilage et al. 2015]. At sufficiently high concentrations of K^+ the GQ becomes more stable than the canonical stem-loop structure that is processed by Dicer, leading to partial unwinding of the stem-loop and inhibition of Dicer-mediated maturation. We circularly rotated the sequence by -12 nt to move the GQ across the $n,1$-junction to demonstrate the modified output style that is required in such cases. Fig. 4 shows the MFE-, centroid-, and maximum expected accuracy (MEA) structures, as well as base pair- and GQ probabilities. With our parametrization, the GQ is predicted to have a very high equilibrium probability and it is present in all three structure representatives as part of a three component multibranch loop.

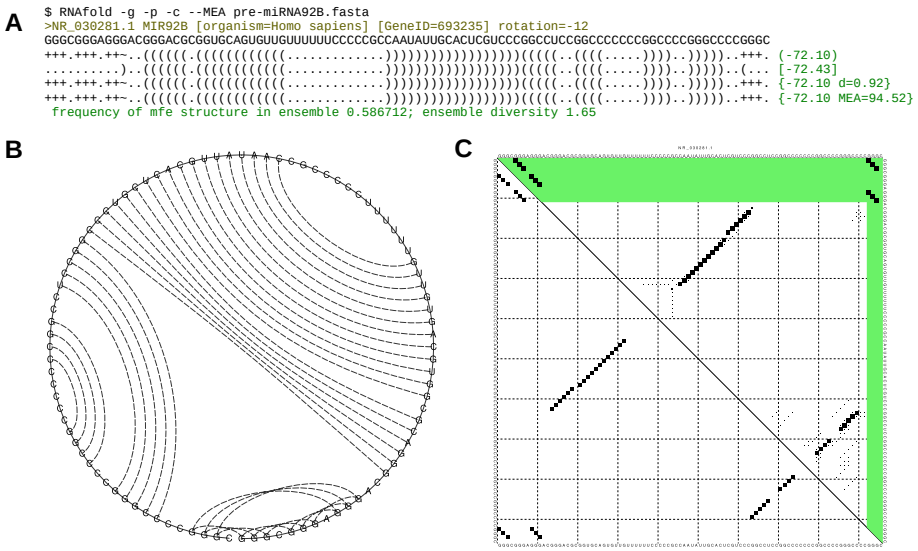

Figure 4. Example output for a short artificial circular RNA sequence that forms a GQ. `RNAfold` output for the pre-cursor miRNA-92B as if it were circular, i.e., with a covalently linked 5' and 3' end. (**A**) Output of the command line call. The first structure in dot-bracket notation is the MFE structure, the second structure output represents the pairing tendency, and the last two structures are the centroid and maximum expected accuracy (MEA) structures, respectively. (**B**) Circular plot of the MFE structure. (**C**) Base pairing and GQ probabilities shown as dot-plot. In the latter, the $n,1$-junction spanning GQ is highlighted in green.

As a real-life application we re-computed the structures of the circRNAs studied by [Liu et al. 2019] with our new GQ support. To this end, we retrieved the 827 high-confidence circRNAs described in the supplemental data Table S1. From this set, we selected all 649 sequences with a length of at most $20,000$ nt. The resulting data set consists of sequences from 208 nt to $19,863$ nt and an average length of about $6,500$ nt. MFE predictions for these sequences resulted in 193 sequences ($\approx 30\%$) that harbor at least one GQ, while the total number of GQs predicted among the entire data set is 298. Many circular RNA sequences thus show the potential to form more than one GQ that can compete against canonical base pairing. The majority of predicted GQs consist of two layers (254). However, there are also 33 three-layer GQs and 11 four-layer GQs. One sequence even showed a stable GQ with two layers spanning the n,1 junction. This analysis shows, that naturally occurring circRNAs seem to frequently have the potential to form stable GQs. Whether or not those are functional *in vivo* remains to be shown. To our knowledge, no studies exist so far that investigate putative GQs in circRNAs.

4. Concluding Remarks

We introduced an exact dynamic programming algorithm that predicts secondary structures with G-quadruplexes in circular RNAs. To this end, the standard grammar describing RNA secondary structure was extended making use of the fact that GQs are local structures that can be handled formally like closed substructure. The most extensive changes concern the exterior loop. The variant of RNA folding as presented here has been implemented in the `ViennaRNA` package and is available from version 2.7.0 onwards.

At present, only a very simple energy model for RNA quadruplexes is available. It entails a very simple energy function for the GQs themselves, depending only

on the total lengths of the linkers and number of stacked layers. Since the publication of the simple energy model [Lorenz et al. 2012], only a small number of energy measurements are available [Matsumoto et al. 2020]. Melting temperatures measured e.g. in [Pandey et al. 2013], on the other hand, cannot readily be used to gauge energy parameters. The stability of G-quadruplexes moreover depends strongly on the potassium concentration [Joachimi et al. 2009]. It would therefore be of significant interest to include both the general dependence of monovalent ions (see [Yao et al. 2023] for regular secondary structures) and the specific dependence on K^+. This remains a question for future research, however.

The `ViennaRNA` package provides partial support for co-axial stacking of adjacent helices in its MFE computations. Partition function computations in `ViennaRNA` neglect such potentially stabilizing interactions. An extended grammar handling the full diversity of coaxial stacking recently has been described in [Courtney et al. 2023] for the `memerna` software. It is not unreasonable to assume that multiple adjacent GQs exhibit a similar stacking behavior. To our knowledge, however, no data is available in this regard, hence we did not attempt to include such effects in our current model.

To-date GQs and circRNAs have been studied independently of each other. The extension of the `ViennaRNA` package described here makes it possible to systematically consider the interplay of these two important features of transcripts.

Acknowledgments. Discussions with Ivo L. Hofacker on the topic are gratefully acknowledged. This work has been funded in parts by the German Federal Ministery of Education in extension of the German Network for Bioinformatics Infrastructure de.NBI/RBC and the Austrian Science Fund (FWF), grant no. I-4520. Within Leipzig University, PFS also holds affiliations with the School for Embedded and Composite Artificial Intelligence (SECAI) Dresden-Leipzig, the Competence Center for Scalable Data Services and Solutions Dresden-Leipzig, German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, the Leipzig Research Center for Civilization Diseases, and Centre for Biotechnology and Biomedicine at Leipzig University.

References

Arachchilage, G. M., Dassanayake, A. C., and Basu, S. (2015). A potassium ion-dependent RNA structural switch regulates human pre-miRNA 92b maturation. *Chemistry & Biology*, 22(2):262–272.

Bernhart, S. H., Mückstein, U., and Hofacker, I. L. (2011). RNA Accessibility in cubic time. *Algorithms Mol Biol*, 6(1):3.

Courtney, E., Datta, A., Mathews, D. H., and Ward, M. (2023). memerna: Sparse RNA folding including coaxial stacking. Technical Report 551958, bioRxiv.

Cueny, R. R., McMillan, S. D., and Keck, J. L. (2022). G-quadruplexes in bacteria: insights into the regulatory roles and interacting proteins of non-canonical nucleic acid structures. *Crit Rev Biochem Mol Biol*, 57:539–561.

Dumas, L., Herviou, P., Dassi, E., Cammas, A., and Millevoi, S. (2021). G-quadruplexes in RNA biology: Recent advances and future directions. *Trends Biochem. Sci.*, 46:270–283.

Fay, M. M., Lyons, S. M., and Ivanov, P. (2017). RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.*, 429:2127–2147.

Freed, K. F. (2012). Influence of small rings on the thermodynamics of equilibrium self-assembly. *J. Chem. Phys.*, 136:244904.

Giegerich, R. (2000). Explaining and controlling ambiguity in dynamic programming. In Giancarlo, R. and Sankoff, D., editors, *Combinatorial Pattern Matching. CPM 2000*, volume 1848 of *Lect. Notes Comp. Sci.*, pages 46–59, Berlin, Heidelberg. Springer.

Guo, J. U. and Bartel, D. P. (2016). RNA G-quadruplexes are globally unfolded in eukaryotic cells and depleted in bacteria. *Science*, 353:aaf5371.

Hofacker, I. L. and Stadler, P. F. (2006). Memory efficient folding algorithms for circular RNA secondary structures. *Bioinformatics*, 22:1172–1176.

Höner zu Siederdissen, C., Prohaska, S. J., and Stadler, P. F. (2015). Algebraic dynamic programming over general data structures. *BMC Bioinformatics*, 16:19:S2.

Joachimi, A., Benz, A., and Hartig, J. S. (2009). A comparison of DNA and RNA quadruplex structures and stabilities. *Bioorg Med Chem*, 17:6811–6815.

Kimchi, O., Cragnolini, T., Brenner, M. P., and Colwell, L. J. (2019). A polymer physics framework for the entropy of arbitrary pseudoknots. *Biophys. J.*, 117(3):520–532.

Kwok, C. K., Marsico, G., Sahakyan, A. B., Chambers, V. S., and Balasubramanian, S. (2016). rG4-seq reveals widespread formation of G-quadruplex structures in the human transcriptome. *Nat. Methods*, 13:841–844.

Lee, D. S. M., Ghanem, L. R., and Barash, Y. (2020). Integrative analysis reveals RNA G-quadruplexes in UTRs are selectively constrained and enriched for functional associations. *Nat Commun*, 11:527.

Li, F. and Zhou, J. (2023). G-quadruplexes from non-coding RNAs. *J Mol Med (Berl)*, 101:621–635.

Liu, C.-X., Nan, F., Jiang, S., Gao, X., Guo, S.-K., Xue, W., Cui, Y., Dong, K., Ding, H., Qu, B., Zhou, Z., Shen, N., Yang, L., and Chen, L.-L. (2019). Structure and degradation of circular RNAs regulate PKR activation in innate immunity. *Cell*, 177:865–880.

Lorenz, R., Bernhart, S. H., Externbrink, F., Qin, J., Höner zu Siederdissen, C., Amman, F., Hofacker, I. L., and Stadler, P. F. (2012). RNA folding algorithms with G-quadruplexes. In de Souto, M. C. P. and Kann, M. G., editors, *BSB 2012*, volume 7409 of *Lect. Notes Bioinf.*, pages 49–60, Heidelberg. Springer.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., and Hofacker, I. L. (2011). ViennaRNA Package 2.0. *Alg. Mol. Biol.*, 6:26.

Lorenz, R., Bernhart, S. H., Qin, J., Höner zu Siederdissen, C., Tanzer, A., Amman, F., Hofacker, I. L., and Stadler, P. F. (2013). 2D meets 4G: G-quadruplexes in RNA secondary structure prediction. *IEEE Trans. Comp. Biol. Bioinf.*, 10:832–844.

Lyu, K., Chow, E. Y.-C., Mou, X., Chan, T.-F., and Kwok, C. K. (2021). RNA G-quadruplexes (rG4s): genomics and biological functions. *Nucleic Acids Res.*, 49:5426–5450.

Matsumoto, S., Tateishi-Karimata, H., Takahashi, S., Ohyama, T., and Sugimoto, N. (2020). Effect of molecular crowding on the stability of RNA G-quadruplexes with various numbers of quartets and lengths of loops. *Biochemistry*, 59:2640–2649.

Pandey, S., Agarwala, P., and Maiti, S. (2013). Effect of loops and G-quartets on the stability of RNA G-quadruplexes. *J. Phys. Chem. B*, 117:6896–6905.

Sahayasheela, V. J. and Sugiyama, H. (2023). RNA G-quadruplex in functional regulation of noncoding RNA: Challenges and emerging opportunities. *Cell Chemical Biology*, 31:53–70.

Turner, D. H. and Mathews, D. H. (2010). NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucl. Acids Res.*, 38:D280–D282.

Vromman, M., Anckaert, J., Bortoluzzi, S., Buratin, A., Chen, C.-Y., Chu, Q., Chuang, T.-J., Dehghannasiri, R., Dieterich, C., Dong, X., Flicek, P., Gaffo, E., Gu, W., He, C., Hoffmann, S., Izuogu, O., Jackson, M., Jakobi, T., Lai, E., Nuytens, J., Salzman, J., Santibanez-Koref, M., Stadler, P. F., Thas, O., Vanden Eynde, E., Verniers, K., Wen, G., Westholm, J., Yang, L., Ye, C.-Y., Yigit, N., Yuan, G.-H., Zhang, J., Zhao, F., Vandesompele, J., and Volders, P.-J. (2023). Large-scale benchmarking of circRNA detection tools reveals large differences in sensitivity but not in precision. *Nature Methods*, 20:1159–1169.

Ward, M., Datta, A., Wise, M., and Mathews, D. H. (2017). Advanced multi-loop algorithms for RNA secondary structure prediction reveal that the simplest model is best. *Nucleic Acids Res.*, 45:8541–8550.

Yang, S. Y., Monchaud, D., and Wong, J. M. Y. (2022). Global mapping of RNA G-quadruplexes (G4-RNAs) using G4RP-seq. *Nat. Protoc.*, 17:870–889.

Yang, X., Cheema, J., Zhang, Y., , Deng, H., Duncan, S., Umar, M. I., Zhao, J., Liu, Q., Cao, X., Kwok, C. K., and Ding, Y. (2020). RNA G-quadruplex structures exist and function *in vivo* in plants. *Genome Biol.*, 21:226.

Yao, H.-T., Lorenz, R., Hofacker, I. L., and Stadler, P. F. (2023). Mono-valent salt corrections for RNA secondary structures in the ViennaRNA package. *Alg. Mol. Biol.*, 18:8.

Yu, H., Qi, Y., Yang, B., Yang, X., and Ding, Y. (2023). G4Atlas: a comprehensive transcriptome-wide G-quadruplex database. *Nucleic Acids Res.*, 51:D126–D134.

Zhang, A. Y., Bugaut, A., and Balasubramanian, S. (2011). A sequence-independent analysis of the loop length dependence of intramolecular RNA G-quadruplex stability and topology. *Biochemistry*, 50:7251–7258.

Zhang, D. H., Fujimoto, T., Saxena, S., Yu, H. Q., Miyoshi, D., and Sugimoto, N. (2010). Monomorphic RNA G-quadruplex and polymorphic DNA G-quadruplex structures responding to cellular environmental factors. *Biochemistry*, 49:4554–4563.