Comparative Analysis of Cyclic Sequences: Viroids and other Small Circular RNAs

Axel Mosig[1,2], Ivo L. Hofacker[3] and Peter F. Stadler[4,3,5]
[1]Department of Combinatorics and Geometry (DCG),
MPG/CAS Partner Institute for Computational Biology (PICB),
Shanghai Institutes for Biological Sciences (SIBS) Campus, Shanghai, China
[2]Max Planck Insitute for Mathematics in the Sciences,
Inselstrasse 22, D-04103 Leipzig, Germany
[3]Institute for Theoretical Chemistry, University of Vienna,
Währingerstrasse 17, A-1090 Vienna, Austria
[4]Bioinformatics Group, Department of Computer Science,
and Interdisciplinary Center for Bioinformatics,
University of Leipzig, Härtelstrasse 16-18, D-04107 Leipzig, Germany.
[5]The Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, New Mexico

Abstract: The analysis of small circular sequences requires specialized tools. While the differences between linear and circular sequences can be neglected in the case of long molecules such as bacterial genomes since in practice all analysis is performed in sequence windows, this is not true for viroids and related sequences which are usually only a few hundred basepairs long. In this contribution we present basic algorithms and corresponding software for circular RNAs. In particular, we discuss the problem of pairwise and multiple cyclic sequence alignments with affine gap costs, and an extension of a recent approach to circular RNA folding to the computation of consensus structures.
Keywords: RNA secondary structure, circular RNA, dynamic programming, viroids

1 Introduction

Circular DNA is a common phenomenon in nature. Indeed, bacterial genomes as well as their plasmids are circular. Most organellar genomes of mitochondria and plastids are circular as well. In practice, however, the distinction between linear and circular sequences is irrelevant for bioinformatics at least in the case of long sequences, because the analysis will always focus on individual genes or on short sequence windows. Shorter sequences, on the other hand, with a length of, say, less than 10kb or 20kb, could be investigated as a whole. While mitochondrial genomes (with a length between 15 and 17kb for most metazoan animals and much longer for most other Eukaryote clades) have to be treated at the gene level due to rapid genomic rearrangements [BB98], this is not the case for most virus families. Proper virus genomes can be as short as 2kb, see Tab. 1. This heterogeneous group of sequences stems from various viral families, two of which are retro-transcribing (Hepatitis B Virus and Caulimo-Virus), several have double-stranded circular DNA genomes, and a

Table 1: Natural Short Circular Nucleic Acids Groups

Name	Type	Size(kb)	Seq.*	Remarks
rCarSV	ssRNA	0.2-0.3	13	viroid-like
Viroids	ssRNA	0.3-0.4	659	
Satellite RNAs	ssRNA	0.35	11	9 distinct groups
Cherry SCV	ssRNA	0.45	1	viroid-like
Deltavirus	ssRNA	1.7	51	viroid-like
Circoviridae	ssDNA virus	2	299	
Hepadnaviridae	dsDNA RT-virus	3	893	Hepatitis B
Parvoviridae	ssDNA virus	4-6	84	
Microviridae	ssDNA virus	4-6	92	
Polyomaviridae	dsDNA virus	5	447	
Geminiviridae	ssDNA virus	5	301	1 or 2 chromosomes
Nanovirus	ssDNA virus	6-9	2	4-6 chromosomes
Papillomaviridae	dsDNA virus	7-8	169	
Inoviridae	ssDNA virus	8	42	
Caulimoviridae	dsDNA RT-virus	8	57	
Corticoviridae	dsDNA virus	9	2	
Plasmaviridae	dsDNA virus	12	2	
Fuselloviridae	dsDNA virus	15	9	
Mitochondria	dsDNA virus	$\geq$13	$\sim$1000	rapid rearrangements
Plasmids $\leq$ 20kb	dsDNA		50	

NCBI Query: *Name*`[orgn] and ''complete genome''`, 2006-03-12

few have short, single-stranded DNA genomes.

The overwhelming majority of single-stranded nucleic acid is RNA, and most RNA molecules are linear. The `Subviral RNA DB` [PRPP03, RP06], nevertheless lists more than 1000 circular RNA genomes of viroids and related objects. Viroids are important plant pathogens that induce symptoms similar to those accompanying virus infections. They are composed of a small, nonprotein-coding, single-stranded, circular RNA, with autonomous replication that proceeds through an RNA-based rolling-circle mechanism, see [FHMdA$^+$05] for a recent review. Several different classes of satellite RNAs are also circular [SR99]. In addition, there are three distinct classes of viroid-like sequences: cherry small circular viroid-like RNA [DSDRR97], carnation small viroid-like RNA (CarSV RNA), which is unique in that is has a DNA form and behaves similar to a retrovirus [HDB04], and Hepatitis delta virus [Tay06]. A phylogenetic analysis of viroid and viroid-like satellite RNAs can be found in [EDdlP$^+$01].

Recently, several additional (mutually unrelated) groups of circular RNAs have been discovered. In particular, 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].

Single-stranded nucleic acids, in particular, have to be treated with care when windowing techniques are used. For instance, secondary structure formation is well known to be an inherently global phenomenon. As demonstrated in [HS06], neglecting circularity can have quite dramatic effects. Substantial errors might furthermore result from the inconsistent treatment of "end-gaps" when circular sequences are aligned as if they were linear, even if the cut-point is chosen between homologous positions.

The wide variety of short circular nucleic acids listed above calls for the development of specific computational tools to deal with these exceptional sequences in a consistent way. In particular, for the case of single-stranded nucleic acids one will be interested in determining conserved RNA secondary structures, which have been demonstrated to exist in many viral RNA genomes [HSS04], in viroids [SHF$^+$84, RWR$^+$99], which were among the first RNAs for which secondary structures have been studied systematically, and also in some ssDNA viruses including parvoviridae, circoviridae, and geminiviridae [VCF05].

2 Cyclic Alignments

An alignment $\mathbb{A}$ of two strings x and y of length n and m, resp., is a sequence of pairs of the form (x_i, y_j), $(x_i, -)$, and $(-, y_j)$ that preserves the order of sequence positions in both x and y. A maximal sequence of $(x_i, -)$ pairs is called a *deletion*, while a maximal sequence of $(-, y_j)$ is called an *insertion*. We assume that the (similarity) score $S(\mathbb{A})$ of the alignment $\mathbb{A}$ is the sum of scores for individual substitutions, insertions, and deletions.

In the case of cyclic sequences, insertions and deletions may wrap around the ends, of course. Thus the cyclic score $S_L(\mathbb{A})$ is in general larger than the score $S(\mathbb{A})$ of the linear (representation of the) alignment $\mathbb{A}$. The cyclic shift operator σ that rotates a string or an alignment by one position: $\sigma(x) = (x_2, \ldots, x_{n-1}, x_n, x_1)$. The cyclic score of the alignment is thus

$$S_L(\mathbb{A}) = \max_k S(\sigma^k(\mathbb{A})) \tag{1}$$

under the above additivity assumption on the scoring model.

The cyclic string associated with an ordinary string x conveniently represented as the equivalence class $[x] = \{x, \sigma(x), \sigma^2(x), \ldots, \sigma^{n-1}(x)\}$. The cyclic alignment problem thus consists of finding the optimal linear alignment of $\sigma^p(x)$ and $\sigma^q(y)$ for the optimal choices of p and q [BB93, GT93, Mae90, MVC02]. This problem can be solved in polynomial time in full generality: One simply has to compute the optimal alignment $\mathbb{A}^{(p,q)}$ of $\sigma^p(x)$ and $\sigma^q(y)$, which can done in $\mathcal{O}(nm \max(n+m))$ time and $\mathcal{O}(nm)$ space [Dew01], for all nm combinations of p and q. This quintic algorithm is implemented in the `circal` approach to aligning mitochondrial genome arrangements [FSS06], which uses a complex gap cost function while having to deal with only short "sequences". This approach becomes impractical, however, for sequences with several hundred or thousand characters.

In the case of linear gap cost functions ($g(k) = k\delta$), [Mae90] introduced a $\mathcal{O}(n^2 \ln n)$ algorithm that is based on the fact that optimal alignment paths cannot cross in this case. The most widely used alignment programs, including `clustalW` [THG94], uses a scor-

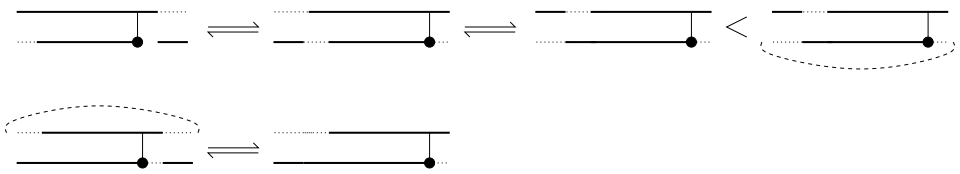

Figure 1: The case in which the longer sequence x ends in a gap need not be considered. The "tail" of $\sigma^k y$ after the last matched position can always be rotated to the beginning of the alignment without decreasing the score. Of the three possible cases for the beginning of the alignment we show only the two non-trivial ones. If the alignment begins with a (mis)match we have the same situation as in the second case. $\leftrightharpoons$ indicates changes that do not influence the score of the alignment, $<$ indicates a possible increase in the score, and a brace means that the alignment is scores as circular alignment rather than as a linear one.

ing scheme with affine gap costs. In this model a sequence of k contiguous insertions (deletions) incurs a cost $g(k) = \delta_o + (k-1)\delta_e$ independent of the inserted or deleted characters. In the case of affine gap function, optimal alignment paths may cross, so that a direct generalization of Maes's idea does not seem to be feasible.

The alignment problem for linear strings with affine gap costs is solved by Gotoh's algorithm [Got82]: Let $S_{i,j}$ be the score of an optimal alignment of the prefixes $x[1 \ldots i]$ and $y[1 \ldots j]$; similarly, E_{ij} and F_{ij} are the optimal alignments on the prefixed subject to the constraint that they end with a deletion or an insertion, respectively. We have the following recursions

$$\begin{aligned} E_{ij} &= \max\{E_{i-1,j} - \delta_e, S_{i-1,j} - \delta_o\} \\ F_{ij} &= \max\{F_{i,j-1} - \delta_e, S_{i,j-1} - \delta_o\} \\ S_{ij} &= \max\{E_{i,j}, F_{i,j}, S_{i-1,j-1} + \sigma(x_i, y_j)\} \end{aligned} \tag{2}$$

with initial conditions $S_{0,0} = 0$, $E_{0,0} = F_{0,0} = -\infty$ and the understanding that terms with negative indices are ignored.

This automatically yields a quartic algorithm for the cyclic case by simply considering all cyclic shifts of x and y. In the case of linear gap costs one easily sees that it is sufficient to use all rotations of the shorter string, which reduces the CPU requirements to $\mathcal{O}(nm \min(n,m))$, i.e. a cubic algorithm. This simplification does not work for more general gap costs, however, since gaps may "wrap around" the ends of the sequence.

In the case of *affine* gap functions we can also obtain a cubic algorithm using the fact that we have to distinguish only three cases: (1) The alignment of x with $\sigma^q(y)$ (1) ends in a match $x_n, (\sigma(y))_m$, it ends in a deletion $(x_n, -)$ or (3) it ends in an insertion $(-, (\sigma^q(y)_m)$. In the first case, the score of the cyclic alignment is the same as in the linear alignment. In case (2), we simply start the recursions with the initialization $S_{0,0} = -\infty$, $E_{0,0} = 0$, $F_{0,0} = -\infty$ and obtain the optimal score in $E_{n,m}$, while in case (3) we initialize $S_{00} = -\infty$, $E_{0,0} = -\infty$, $F_{0,0} = 0$ and obtain the optimal solution from $F_{n,m}$. Since we rotate the sequence y we do not need to consider case (3) where $(\sigma^q(y)_y)$ is unpaired since this situation can always be achieve by rotating the remaining tail after the last match in $(\sigma^q(y)_y)$ to the beginning of the string, see Fig. 1. This yields a simple

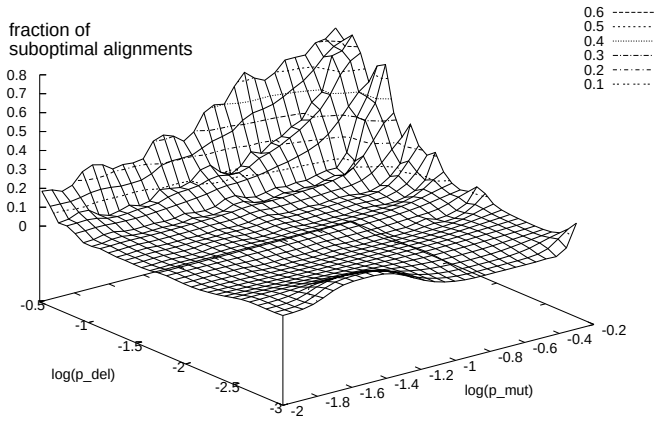

Figure 2: Fraction of non-optimal alignments computed using the best local heuristic as a function of the fraction of substitutions (0-50%) and indels (0-30%) for sequence of length $N = 100$. The average score of the heuristic alignments is above 97% of the optimal score in the entire range.

$\mathcal{O}(nm \times \min(n, m))$ algorithm for the cyclic alignment problem with affine gap costs.

The close relationship between cyclic and linear alignment suggests a number of plausible heuristics. For instance, one may search for the best local alignment of x and y and use a central match from this local alignment to "anchor" the alignment of cyclic sequences. Fig. 2 summarizes the performance of this heuristic approach: as long as there is only a moderate fraction of insertions and deletions, it yields the correct solution in most cases, and only slightly sub-optimal alignments in the remaining cases.

The $\mathcal{O}(n^3)$ exact algorithm as well as the $\mathcal{O}(n^2)$ heuristic algorithms are implemented in the `cyclope` package[1].

3 Multiple Alignments and Phylogenetics

Progressive multiple alignments can be constructed by generalizing the pairwise algorithms described above to profiles in the same way as in `clustalW` [THG94]. To this end, we use the sum-of-pairs score to measure the similarity of two profiles and employ iterative clustering to construct the guide tree. After each step, similarity scores to the newly joined profile are computed explicitly rather than estimated by an averaging procedure such as WPGMA.

An issue common to all progressive alignment methods, independent of whether linear or cyclic sequences are considered, is the assignment of appropriate weights to the individual sequences. Without a weighting scheme, groups of highly similar sequences tend to be overemphasized and dominate the alignment, while individual highly divergent sequences

[1] Available from `www.bioinf.uni-leipzig.de/Software/cyclope/`.

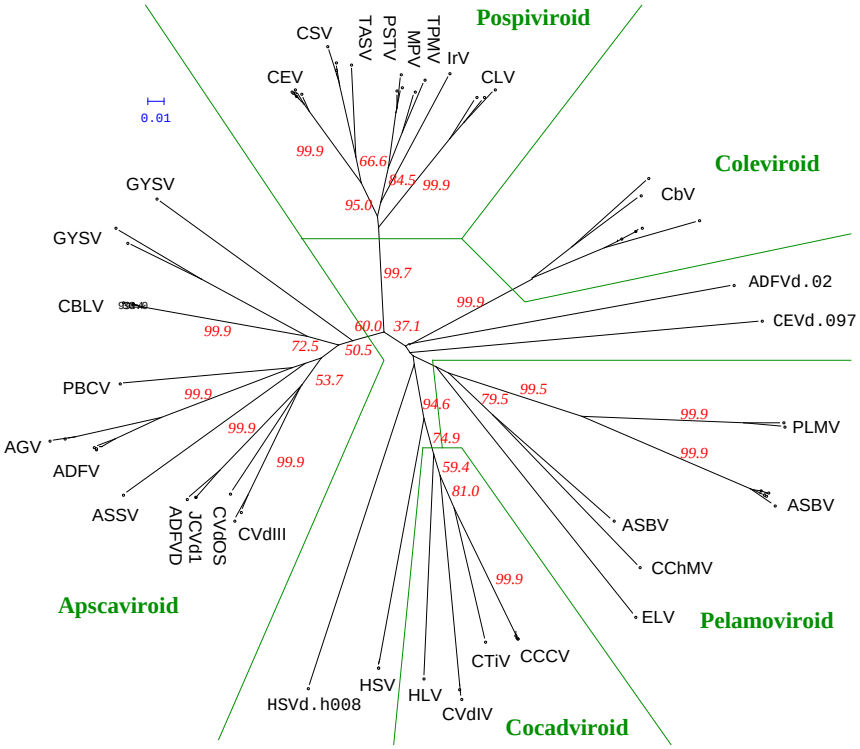

Figure 3: Neighbor-Joining tree obtained from an alignment of sequences from the `Subviral RNA DB`. Starting with 987 viroid sequences from 25 known classes of viroids, we applied `cyclope` to compute one alignment for the sequences of each of the 25 classes. These 25 alignments were used to derive 70 representative sequences, an alignment of which was taken as input to obtained the neighbor joining tree shown here. The circular multiple alignment was computed using `cyclope`, followed by a realignment step using `clustalW`. Major edges in the tree are shown along with their bootstrap values, as computed by the `SplitsTree` package.

are neglected. Avoiding the costly computation of all distances in a neighbor joining tree, `cyclope` supports a simple, yet effective weighting scheme that works as follows: Initially, each individual sequence is assigned weight 1. After two alignments A and B with K and L many sequences, respectively, have been aligned to one alignment C with $M = K + L$ sequences, the weights are readjusted. Let $v_1, \ldots, v_K$ and $w_1, \ldots, w_L$ denote the original weights of the sequences in A and B. We compute the *alignment distance* (i.e., the relative number of of matching nucleotides in the alignment of A and B) $D := \mathbf{d}(A, B)$ between A and B, which then allows us to update the weights by setting $v_i := (v_i + D)/(2K)$ as well as $w_i := (w_i + D)/(2L)$. This essentially corresponds to setting branches in the guide tree to the midpoint at each vertex, while the weights are adjusted such that small and highly divergent classes receive higher scores than large groups of similar sequences. During any alignment step, those weights are normalized such that the largest weight will be 1.

In order to utilize the much more sophisticated features of standard linear alignment packages such as `clustalW`, one can use `cyclope` to obtain a rough preliminary multiple alignment and realign those with linear alignment packages. Note that the heuristic for shifting implemented by `cyclope` is based on conserved blocks, so that the likeliness of gaps at starting positions – which linear alignment programs are not capable of handling – is kept minimal.

As a demonstration of `Cyclope` we constructed a multiple alignment of representative sequences from the `Subviral RNA DB`. Neighbor-joining was then used to infer the phylogenetic tree in Fig. 3. In general, the results are in good agreement with an earlier study [EDdlP+01]. A few details, however, deviate. For example, our data places Avocado Sunbloch Viroid ASB-Viroid within the Pelamoviroid group.

4 RNA Folding Algorithms for Circular Sequences

Michael Zuker's approach to computing both the minimum energy structure and a certain class of suboptimal folds for an RNA sequence is directly applicable to circular RNAs. In fact, `mfold` treats linear RNAs as exceptional variants of the circular ones [Zuk89, Zuk03]. In contrast, the `Vienna RNA Package`[2] [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 can be treated 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 re-design of the recursions that are optimized for the linear RNA case, see [HS06] for details.

Using the same algorithmic approach, it is straightforward to generalize `RNAalifold` [HFS02] to circular sequences. This option, which allows the computation of the consensus structure of an alignment of circular single-stranded RNA or DNA molecules, is implemented in the current version of `Vienna RNA Package`.

As an example we show the consensus structure for eggplant latent viroid in figure 4. The structure is very stable and moreover supported by several consistent and compensatory mutations. Unusual structural stability is sometimes used as a marker for *functional* noncoding RNAs. A well-tested measure for structural stability is the z-score comparing the consensus folding energy of an alignment with randomized alignments obtained by shuffling columns, which is computed by the `alifoldz` program [WH04]. Interestingly, viroids exhibit the strongest signals for structural conservation of all RNAs investigated so far. While most "classical" ncRNA can be detected by `alifoldz` at a z-score cut-off of -4, the ELVd alignment exhibits a z-sore of -13.8.

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

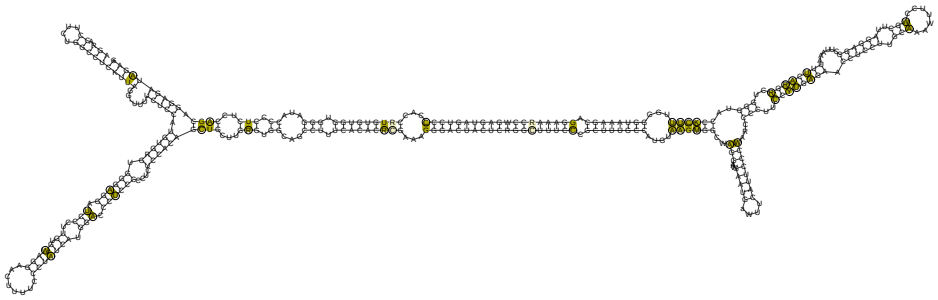

Figure 4: Consensus secondary structure for eggplant latent viroid as predicted from a `cyclope` alignment by `RNAalifold -circ`. The alignment comprises four sequence selected to have have less than 95% sequence identity. Base pairs shown in ochre are supported by a consistent or compensatory mutation with circles marking the sites of variation.

5 Concluding Remarks

So far, circular RNA and DNA sequences have been considered as a rather exotic side-line that do not warrant specialized tools. With increasing sequence information becoming available, the (ab)use of methods that are designed for linear sequences becomes increasingly tedious as it requires manual corrections of both alignments and subsequent analysis. In this contribution we presented a dedicated alignment tool for (short) circular sequences, which is particularly geared towards viroids and small virus RNAs. While the time complexity is higher than classical alignment algorithms, it is efficient enough in practice for use with viroid and other subviral sequences. Based on the pairwise dynamic programming alignments, `cyclope` also features a `clustal`-like progressive alignment tool. These alignments can be used without further processing for phylogeny reconstruction or RNA secondary structure analysis.

In the case of viroid phylogeny, previous studies were essentially confirmed based on extended data sets. An RNA consensus folding program for circular RNAs, which combines the `RNAalifold` approach with a recent algorithm for folding circular RNAs, shows that viroids have exceptionally strong signals for structural stability when compared to other functionals ncRNAs.

Acknowledgments. This work has been funded, in part, by the Austrian GEN-AU projects "bioinformatics integration network II" and "non coding RNA", and by the German *DFG* Bioinformatics Initiative BIZ-6/1-2. PFS thanks PICB for its hospitality during a visit in March 2006, during which a significant part of this work was conceived.

References

[BB93] H. Bunke und U. B¨uhler. Applications of approximate string matching to 2D shape recognition. *Patt. Recogn.*, 26:1797–1812, 1993.

[BB98] J. L. Boore und W. M. Brown. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opinion Gen. Devel.*, 8:668–674, 1998.

[Dew01] T. G. Dewey. A sequence alignment algorithm with an arbitrary gap penalty function. *J. Comp. Biol.*, 8:177–190, 2001.

[DSDRR97] F Di Serio, J A Daros, A Ragozzino und Flores R. A 451-nucleotide circular RNA from cherry with hammerhead ribozymes in its strands of both polarities. *J Virol.*, 71:6603–6610, 1997.

[EDdlP+01] Santiago F. Elena, Joaquín Dopazo, Marcos de la Peña, Ricardo Flores, Theodor O. Diener und Andrés Moya. Phylogenetic Analysis of Viroid and Viroid-Like Satellite RNAs from Plants: A Reassessment. *J. Mol. Evol.*, 53:155–159, 2001.

[FHMdA+05] R Flores, C Hernandez, A E Martinez de Alba, J A Daros und F. Di Serio. Viroids and viroid-host interactions. *Annu. Rev. Phytopathol.*, 43:117–139, 2005.

[FSS06] Guido Fritzsch, Martin Schlegel und Peter F. Stadler. Alignments of Mitochondrial Genome Arrangements: Applications to Metazoan Phylogeny. *J. Theor. Biol.*, 2006. doi:10.1016/j.jtbi.2005.10.010, in press.

[Got82] O. Gotoh. An improved algorithm for matching biological sequences. *J. Mol. Biol.*, 162:705–708, 1982.

[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.

[HDB04] K Hegedus, G Dallmann und E. Balazs. The DNA form of a retroviroid-like element is involved in recombination events with itself and with the plant genome. *Virology*, 325:277–286, 2004.

[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.

[HS06] Ivo L. Hofacker und Peter F. Stadler. Memory Efficient Folding Algorithms for Circular RNA Secondary Structures. *Bioinformatics*, 22:1172–1176, 2006. **See also:** Proceedings of the German Conference on Bioinformatics. GCB 2003 Torda, A., Kurtz, S. and Rarey, M., Lecture Notes in Informatics P-71, pp.15-25 (2005).

[HSS04] Ivo L. Hofacker, Roman Stocsits und Peter F. Stadler. Conserved RNA Secondary Structures in Viral Genomes: A Survey. *Bioinformatics*, 20:1495–1499, 2004. **See also:** Proceedings of the German Conference on Bioinformatics. GCB 2003 vol. **1**; Mewes, H.-W., Heun, V., Frishman, D. and Kramer, S. (eds.); belleville Verlag Michael Farin, M¨unchen, D (2003), pp. 57-62.

[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.

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

[MVC02] R. A. Mollineda, E. Vidal und F. Casacuberta. Cyclic sequence alignments: approximate versus optimal techniques. *Int. J. Pattern Rec. Artif. Intel.*, 16:291–299, 2002.

[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.

[PRPP03] M. Pelchat, L. Rocheleau, J. Perreault und J-P Perreault. SubViral RNA: a database of the smallest known auto-replicable RNA species. *Nucleic Acids Res.*, 31:444–445, 2003.

[RP06] Lynda Rocheleau und Martin Pelchat. The Subviral RNA Database: a toolbox for viroids, the hepatitis delta virus and satellite RNAs research. *BMC Microbiology*, 6:24, 2006. doi:10.1186/1471-2180-6-24.

[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.

[SR99] R H Symons und J W. Randles. Encapsidated circular viroid-like satellite RNAs (virusoids) of plants. *Curr. Top. Microbiol. Immunol.*, 239:81–105, 1999.

[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.

[Tay06] J. M. Taylor. Hepatitis delta virus. *Virology*, 344:71–76, 2006.

[THG94] J. D. Thompson, D. G. Higgs und T. J. Gibson. CLUSTALW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties, and weight matrix choice. *Nucl. Acids Res.*, 22:4673–4680, 1994.

[VCF05] Ramachandran Vanitharani, Padmanabhan Chellappan und Claude M. Fauquet. Geminiviruses and RNA silencing. *Trends Plant Sci.*, 10:144–151, 2005.

[WH04] Stefan Washietl und Ivo L. Hofacker. Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J. Mol. Biol.*, 342:19–39, 2004.

[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.