# Introduction to Special Issue on RNA

Peter Clote*

> "*There is now an enormous amount of information on the events that take place in living systems at the molecular level. However, much of this information is qualitative and descriptive, even when the components involved are known and the structures of many of them (proteins and nucleic acids) have been determined. Many ingenious experiments have been done to establish which phenomena take place, but most of them do not address the question of why things happen the way they do. This is where the physical sciences, including thermodynamics, can make an essential contribution to biology.*" From "Microscopic basis of macromolecular thermodynamics" by Themis Lazaridis and Martin Karplus, in *Thermodynamics in Biology*, edited by Enrico di Cera, Oxford University Press (2000).

In *Nature* **447**(7146):799–816, 2007 [48], the ENCODE Consortium presented data, which showed that RNA is "pervasively expressed" in the human genome, with approximately 15% of genomic DNA transcribed, much of it into RNA of no known function, quite distinct from messenger RNA, transfer RNA, ribosomal RNA, and small RNAs (microRNA, piRNA, etc.). Has natural selection evolved cells of *Homo sapiens* only to squander a large part of their energy by transcribing nonfunctional RNA? Or are there yet undiscovered roles played by the many intermediate sized RNA transcripts, whose structure has been evolved to regulate processes not yet discovered? This special issue of *Journal of Mathematical Biology* presents new methods that may ultimately help us to understand some of this recently discovered RNA *dark matter*.

*Biology Department, Boston College, Chestnut Hill, MA 02467 clote@bc.edu.
This work is funded in part by the National Science Foundation under Grant No. DBI-0543506.

Indeed, the articles in this special issue concern mathematical and computational aspects of RNA structure (secondary structure, pseudoknots and tertiary structure). Such aspects include structural alignment algorithms, noncoding RNA gene finders, features of the low energy ensemble of structures, canonical secondary structure prediction, prediction of pseudoknotted structures, motif detection, and resolution of steric constraints in 3-dimensional crystal structures including hydrogen bonds. Development of statistical and physico-chemical models for the structure shared by a class of functionally important RNAs allows the subsequent design of *noncoding RNA* gene finders; i.e. gene finders for genomic DNA which is transcribed into functional RNA other than protein-coding messenger RNA. Although such gene finders may find the broadest general interest among biologists, other aspects presented in this special issue are of great interest and applicability for structural biology.

Rather than present articles in alphabetical order, we have grouped the contributions in the following manner. Chapters that constitute for the most part a review of the current field are followed by chapters concerned with RNA structure prediction and analysis (secondary structure and pseudoknots), followed by chapters concerned with structural alignment and gene finders. Since tertiary structure involves quite distinct considerations, when compared with secondary structure and pseudoknots, this issue concludes with chapters concerned with 3-dimensional motif detection and the resolution of hydrogen bond steric clashes in 3-dimensional structures of X-ray structures.

# 1 Brief introduction to RNA

The previously mentioned *Nature* article [48] of the ENCODE Consortium caused quite a stir, even in the popular press. Since the molecular biology community has until now focused most of its attention on DNA and proteins, while generally considering RNA to be a helper molecule of little intrinsic interest, the *Economist* [1] stated that:

> "Molecular biology is undergoing its biggest shake-up in 50 years, as a hitherto little-regarded chemical called RNA acquires an unsuspected significance. It is beginning to dawn on biologists that they may have got it wrong. Not completely wrong, but wrong enough to be embarrassing."

In the past few years, it has emerged that RNA plays a surprising and previously unsuspected role in many biological processes, including *retranslation* of the genetic code (selenocysteine insertion [12, 8], ribosomal frameshift) [5, 42], catalysis (self-splicing and peptide bond formation [17, 44]), highly specific binding of metabolites ($K_D \approx 5nM$) [35, 52], RNA-guided chemical modification of specific nucleotides in the ribosome [56, 47], etc. Of particular importance is the fact that RNA is involved in *gene regulation*; indeed, micro RNAs in both plants and animals can silence genes by post-transcriptional regulation [32], while bacterial riboswitches can regulate genes by allostery, thereby effecting transcriptional or translational control [35, 52]. Recently, alternative splicing in the eukaryote *Neurospora crassa* has been found to be controled by a riboswitch [9]!

While the genomics of DNA and proteins depends to a great extent on *pattern recognition* of strings – e.g. proteins capable of recognizing specific amino acid or nucleic acid sequences (RNA polymerase recognizes and binds the TATA-box) – interactions with RNA depend largely on the secondary and tertiary structure.[1] Examples of control and regulation by RNA structure are given in the following.

The stem-loop structure of a SECIS element[2] is responsible for a singular *retranslation* event, where the UGA *stop* codon is retranslated to selenocysteine, the so-called 21-st amino acid – see [12] for a review of selenocysteine insertion, as well as [8, 26, 23].

A particular type of *pseudoknot* structure in certain mRNA, in combination with a *slippery sequence*, the heptamer `X XXY YYZ`, causes a $-1$ frameshift in ribosomal translation [5, 42], another odd example of retranslation event caused by an RNA structural motif. (The $+1$ frameshift occurs much less frequently, and is explained by a somewhat different model [42].)

Metabolite-sensing $5'$-UTR (untranlated regions) of certain mRNAs, called *riboswitches*, have been discovered to undergo a conformational change upon ligand-binding, which thereby can up- or down-regulate the corresponding protein product [4]. For instance, upon the binding of nucleotide guanine, the G-box riboswitch in the $5'$ UTR of the XPT gene of *Bacillus subtillis*

[1] An exception to this general statement is afforded by micro RNAs, small interfering RNAs, and other small RNAs, which prevent translation of messenger RNA into protein by hybridizing target mRNA.

[2] A selenocysteine insertion sequence (SECIS) element is a small (roughly 32-65 nt.) portion of the $3'$ UTR (untranslated region) in mRNA which forms a characteristic stem-loop structure and is responsible for selenocysteine incorporation [12, 8].

undergoes a conformational change to create a *terminator loop*, thereby prematurely terminating transcription of the XPT gene. Since XPT is involved in guanine metabolism, this is an example of negative autoregulation by a riboswitch [52]. Although riboswitches have been postulated to be an ancient genetic regulatory system, first developed in bacteria, the remarkable discovery of Cheah et al. [9] suggests that eukaryotes may have co-opted riboswitches to control alternative splicing of genes.

These examples point to the imporance of *structure*, rather than sequence, in understanding the function of RNA. We now briefly describe secondary and tertiary structure of RNA, a topic central to most of the articles in this special issue.

## 1.1 RNA secondary structure

A secondary structure for an RNA sequence is a well-balanced parenthesis expression with dots; a dot at position $i$ indicates that $i$ is not base-paired,[3] while balanced left and right parentheses at positions $i$ resp. $j$ indicate the existence of a base pair $(i, j)$ between $i$ and $j$. For instance, by means of chemical probing, the secondary structure of the 49 nucleotide RNA selenocysteine insertion sequence `fruA` was determined by A. Böck (Ludwig-Maximilians Universität München, personal communication) to be as follows:

```
CCUCGAGGGGAACCCGAAAGGGACCCGAGAGG
((((..(((...(((....))).))).))..))))
```

Here, the first position 1, occupied by `C`, is base-paired with the last position 32, occupied by `G`, and the hairpin loop is closed by base pair $(15, 20)$ with `C` at position 15 and `G` at position 20, etc. Well-balanced parenthesis expressions can be represented as a particular kind of planar graph (technically known as an *outerplanar graph*), giving rise to a more familiar representation, as in the clover-leaf structure for tRNA. In contrast, the gag/pro ribosomal frameshift site of Bovine Leukemia Virus with EMBL accession number AF033818 has a pseudoknot and requires two kinds of parenthesis to depict the structure

```
AAAAAACUAAUAGAGGGGGGACUUAGCGCCCCCCAAACCGUAACCCC
..............[[[[[[.....(((]]]]]]....)))......
```

[3]Given an RNA sequence $a_1, \ldots, a_n$ of length $n$, a base pair $(i, j)$ is an ordered pair of indices $1 \leq i < j \leq n$, such that the nucleotides $a_i, a_j$ form a hydrogen bonded Watson-Crick or GU wobble pair.

For further examples of pseudoknotted structures, consult the *PseudoBase* database [57].

Nussinov and Jacobson [46] pioneered the use of dynamic programming to determine the *minimum free energy* (MFE) structure for a given RNA nucleotide sequence. In the simplest cast, the Nussinov-Jacobson energy model stipulates a contribution of $-1$ for each base pair of a secondary structure, in which case the MFE structure simply maximizes the number of base pairs. Dynamic programming is used to compute intermediate terms in the energy array

$$E_{i,j} = \min \begin{cases} \min\{E_{i,k-1} + E_{k+1,j} + e_{k,j} : i \leq k < j\} \\ E_{i,j-1} \end{cases}$$

where $e_{k,j} = -1$ if positions $k, j$ can form a Watson-Crick or GU base pair. Zuker's algorithm [65] is a highly sophisticated extension of the Nussinov-Jacobson algorithm, based on the *nearest neighbor* energy model. In contrast to the Nussinov-Jacobson model, which ascribes a negative, stabilizing energy contribution for each base pair, the *nearest neighbor* or *Turner* energy model involves negative, stabilizing free energy terms for *base pairs stacking* (hence exterior base pairs contribute no free energy), along with positive, destabilizing free energy terms for hairpin loops, bulges, internal loops and multiloops. Except for an affine approximation for multiloops, these energy terms have been inferred by *optical melting* experiments, pioneered by Tinoco [27], involving ultraviolet absorption measurements in a spectrophotometer [40, 61]. For instance, from melting experiments at 37 degrees Celsius Turner's rules assign stacking free energy of $-2.24$ kcal/mol to $\begin{matrix} 5'\text{-AC-}3' \\ 3'\text{-UG-}5' \end{matrix}$ and of $-3.26$ kcal/mol to $\begin{matrix} 5'\text{-CC-}3' \\ 3'\text{-GG-}5' \end{matrix}$. See [37] for a recent review on minimum free energy structure prediction.

Many groups, including our own lab, have worked on aspects of RNA secondary structure prediction and alignment. We have already mentioned the pioneering work and subsequent refinements of M. Zuker [65, 64]. A panoply of important RNA secondary structure tools including an implementation of Zuker's algorithm is found in the Vienna RNA Package, developed by I.L. Hofacker, P. Stadler, and co-workers [28, 24]. In [41] McCaskill lifted Zuker's MFE structure prediction algorithm to compute the Boltzmann partition function. McCaskill's algorithm was subsequently used by Ding and Lawrence to develop a clever algorithm to sample secondary structures from

the low energy ensemble, thereby allowing simple frequency counts of occurrence of particular features (unpaired regions, hairpins, etc.) in the sampled structures. Since hybridization targets of small interfering RNA appear to be those single-stranded regions of targeted messenger RNAs, the resulting web server `Sfold` has applications for the design of small interfering RNAs for gene silencing [15, 14].

Many algorithms concern thermodynamic equilibrium minimum free energy structure (`mfold` [65, 64], `RNAfold` [28], `RNAstructure` [39]), the low energy ensemble of structures (`Sfold` [14], `RNAsubopt` [60]), pseudoknot prediction [50, 16, 49], hybridization of two molecules [13, 36, 6], multiple sequence/structure alignment (`Foldalign` [22], `Dynalign` [38]), etc. Analysis of the landscape of all secondary structures is provided by `RNAshapes` of Giegerich and co-workers [54, 58], while *parametric* aspects of the landscape of RNA secondary structures are studied by Clote and co-workers in [21, 11, 10, 59].

While all the previously cited work broadly concerns free energy minimization using dynamic programming, in a different direction, *stochastic context free grammars* have been used [19, 51, 33].

## 1.2 RNA tertiary structure prediction

Tertiary contacts disappear much earlier than stacked base pairs when temperature is raised [3], hence it is commonly believed that RNA secondary structure serves as a stable scaffold for tertiary structure formation. Early pioneering work on RNA 3-dimensional structure prediction and modeling was done by R. Cedergren, F. Major and co-workers, who in [34] applied *constraint programming*[4] methods followed by all-atom refinements. F. Major's program, `Mc-Sym` is an efficient and sophisticated tool for modeling loop regions and motifs of RNA. Other RNA tertiary structure algorithms have been developed; for instance, Hubbard and Hearst [25] designed an early approach to predict the tertiary fold of tRNA, and Ogata et al. [43] as well as Yamaguchi and Del Carpio [62] have implemented genetic algorithms for RNA tertiary structure prediction, all of limited success.

While currently there is limited work on RNA 3-dimensional structure prediction, apart from important contributions of F. Major and co-workers

[4]Constraint programming is an algorithmic design method, where all possible configurations are explored in a "branch-and-cut" tree, which is suitably pruned when realized constraints entail inconsistency.

[34], there is a trend to categorize and better understand particular motifs, such as the A-motif [45] found in the large ribosomal subunit from *H. marismortui* [2], the sarcin/ricin loop [55], etc. Efforts to classify RNA motifs involving non-canonical hydrogen bonding (i.e. not Watson-Crick or GU bonds) have led Leontis and Westhof to introduce a notation which distinguishes between *cis* resp. *trans* base pairs, depending on whether the ribose sugar is on the same resp. opposite side of a median line between the hydrogen bonds, and designates as well the participation of the shallow groove (sugar), Hogsteen or Watson-Crick edge. See [30] for an overview of the Leontis-Westhof notation and see [63, 31] for additional RNA motif descriptions. Algorithms for the automatic classification of canonical and non-canonical base pairing in RNA from X-ray structures given in PDB files (PDB, Protein Data Bank[7]) have been developed by [63, 29]. The article "FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures" by Sarver et al. in the current issue of *J Mol Biol.* describes a 3-dimensional RNA motif detection tool; faster but less accurate methods are given in [18, 20]. Despite such fundamental and important work, there is large gap in the reported success of tertiary structure prediction methods for proteins when compared with the situation for RNA. Perhaps the introduction of a CASP-style[5] could stimulate research of RNA secondary and tertiary structure prediction algorithms, and idea discussed by Eric Westhof in the RNA Benasque Workshop in July 2006.

This concludes our brief introduction to RNA. We now give an overview of the contributions to this special issue.

# 2 Reviews

In "Computational methods in noncoding RNA research", the authors, Machado-Lima, del Portillo and Durham, provide an overview of computational methods to predict the secondary structure of a given RNA nucleotide sequence, and how such methods can be used to detect *noncoding* RNA genes. In "Multiple pattern matchine: A Markov chain approach", Lladser, Betterton and Knight present a very detailed, self-contained review of automata the-

[5]The Critical Assessment for the Structure of Protein (CASP) competition is a biannual blind test where experimentally determined protein structures are temporarily withheld from publication to allow an unbiased competition of protein structure prediction methods [53].

ory, generating functions and transfer matrix methods in their application to recognition of biological sequence data.

# 3 RNA structure prediction

In "Boltzmann ensemble features of RNA secondary structures: a comparative analysis of biological RNA sequences and random shuffles", Chan and Ding investigate statistical features of secondary structures in the low energy Boltzmann ensemble sampled by Sfold given the nucleotide sequence of messenger RNA, as well as that of various structural RNA classes (tRNA, rRNA, RNase-P RNA, Introns, SRP RNA, tmRNA, precursor miRNA).

In "Efficient sampling of RNA secondary structures from the Boltzmann ensemble of low-energy" by Ponty, the *boustrophedon* method is applied to improve the run time for sampling RNA secondary structures. It is shown that worst-case time complexity improves from $O(n^2)$ to $O(n \log n)$, where $n$ is RNA sequence length, while average-case time complexity is improved from $O(n\sqrt{n})$ to $O(n \log n)$ for the Nussinov energy model homopolymer.Moreover, a significant speed-up for the Turner energy model is shown with experiments performed on *Drosophilia melanogaster* 5S mRNA and *Staphylococcus aureus*, using Ponty's boustrophedon modification of theZuker-Markham software `UnaFold` [36].

In "Variations on RNA folding and alignment: Lessons from Benasque", Bompfünewer et al.[6] describe two novel algorithms. The first concerns a method to compute the minimum free energy and partition function for secondary structures having no *lonely* (isolated) base pairs. Isolated base pairs generally contribute nonnegative free energy, and by disallowing a secondary structure to contain such base pairs, the Vienna RNA Package algorithm `RNAfold` gains in efficiency. The second contribution concerns a structural alignment algorithm which can be used to detect conserved noncoding transcribed RNA.

In "Prediction of RNA Psedoknots via Graph Tree Decomposition", by Zhao, Malmberg and Cai, the authors describe a novel algorithm for RNA

[6]In the tradition of the collective pseudonym Nicolas Bourbaki, representing an elite group of French mathematicians, Athanasius F. Bompfünewer is a fictitious entity, who however boasts a photo, curriculum vitae, valid address (in Vienna's main cemetery), etc. Bompfünewer derives from the Austrian pronunciation of the French *pompes funèbres*, where the *dirigeant de pompes funèbres* is the *undertaker*, i.e. *mortician*.

secondary structure prediction including (all possible) pseudoknots, by using graph tree decomposition. Tree decompositions were introduced and exploited by N. Robertson and P.D. Seymour (1986) in a series of landmark papers in graph theory; by parametrizing pseudoknots with tree width, Zhao et al. provide an approach to general pseudoknot prediction, known to be an NP-complete problem.

In "Predicting RNA secondary structures with pseudoknots by MCMC sampling", Metzler and Nebel describe a new algorithm, `McQFold`, which computes the base pairing probabilities over all secondary structures including pseudoknots. While many existent RNA pseudoknot structure prediction algorithms depend on a convenient restriction to a particular class pseudoknots, the authors handle arbitrary pseudoknots by means of a Markov chain Monte Carlo algorithm.

# 4 RNA noncoding gene finders

The article, "Prediction of small, noncoding RNAs in bacteria using heterogeneous data" by Tjaden, describes a new machine learning method for prediction of noncoding RNA genes in bacterial genomes. The method uses a General Markov Model, which incorporates sampled state duration times, and trains on three different forms of data: primary nucleotide sequence, conserved secondary structure and gene expression data.

In "PSSMTS: Position specific scoring matrices on tree structures", Morito, Sato and Sakakibara describe a novel algorithm, that extends earlier work of Sakikabara on pair hidden Markov models on tree structures (PHMMTS). In this paper, the authors describe a pairwise structural alignment algorithm for two RNA secondary structures, that incorporates sequence identities by position specific scoring matrices. Such sequence identities, in addition to a consensus secondary structure, indeed form an important feature of some RNA classes, such as snoRNA (small nucleolar RNA) which directs nucleotide modification of ribosomal RNA.

# 5 RNA 3-dimensional structure

In "FR3D: Finding local and composite recurrent structural motifs in RNA 3D structures", Sarver, Zirbel, Stombaugh, Mokdad, and Leontis describe

a novel algorithm for RNA 3-dimensional motif detection, capable of both symbolic and geometic motif search. Given PDB (Protein DataBank) files for a target RNA of length $m$ and a query motif of length $n$, where the query is possibly *composite* (i.e. containing three or more noncontiguous nucleotide segments), FR3D first applies a filter to remove unlikely candidates, and then computes the *geometric discrepancy* between query and candidate structure. Here the discrepancy is a meaure of root mean square RMS sum of positional and angular differences between a reduced atom representation of the query and candidate structures. The authors illustrate the use of their algorithm on the important structural motifs of sarcin/ricin loop, kink turn and GNRA-tetraloop.

In "RNABC: Forward kinematics to reduce all-atom steric clashes in RNA backbone", the authors, Wang, Kapral, Murray, Richardson, Richardson and Snoeyink, describe new software which attempts to resolve serious steric clashes that result from RNA crystal structures when hydrogen atoms are considered. This software is added to a suite of other useful software developed by the Richardson Lab for working with RNA 3-dimensional structures.

# 6 Acknowledgements

I would like to thank all the authors and anonymous referees, without whom this special issue would not exist. Many of the authors attended the second *RNA Benasque Workshop*, organized by Elena Rivas and Eric Westhof in July 2006. This meeting provided an ideal forum for collaboration and exchange of ideas, and the current special issue exists in part due to that meeting. I am especially indebted to Professors Odo Diekmann and Alan Hastings, joint Editors-in-Chief of the *Journal of Mathematical Biology*, and to Dr. Catriona Byrne of Springer-Verlag, all of whom supported the idea of publishing supplementary issues of *J Math Biol.*, dedicated to certain currently important research areas of mathematical and computational biology. The current issue on RNA is only the first of several planned special issues of *J Math Biol.*.

The work of this author was supported in part by the National Science Foundation under Grant No. DBI-0543506. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

# References


[1] RNA: Really new advances. *Economist*, June 14 2007.

[2] N. Ban, P. Nissen, J. Hansen, P.B. Moore, and T.A. Steitz. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, 289:905, 2000.

[3] A.R. Banerjee, J.A. Jaeger, and D.H. Turner. Thermal unfolding of a group I ribozyme: The low-temperature transition is primarily disruption of tertiary structure. *Biochemistry*, 32:153–163, 1993.

[4] J.E. Barrick, K.A. Corbino, W.C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J.K. Wickiser, and R.R. Breaker. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. USA*, 101(17):6421–6426, 2004.

[5] M. Bekaert, L. Bidou, A. Denise, G. Duchateau-Nguyen, J. Forest, C. Froidevaux, I. Hatin, J. Rousset, and M. Termier. Towards a computational model for −1 eukaryotic frameshifting sites. *Bioinformatics*, 19:327–335, 2003.

[6] S.H. Bernhart, H. Tafer, U. Mückstein, C. Flamm, P.F. Stadler, and I.L. Hofacker. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol*, 1(1), 2006.

[7] F.C. Bernstein, T.F. Koetzle, G.J.B. Williams, E.F. Meyer Jr., M.D. Brice, J.R. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi. The Protein Data Bank: a computer-based archival file for macromolecular structures sequence-structure patterns across diverse proteins. *J. Mol. Biol.*, 112:535–542, 1977.

[8] A. Böck, K. Forschhammer, J. Heider, and C. Baron. Selenoprotein synthesis: An expansion of the genetic code. *Trends Biochem. Sci.*, 16:463–467, 1991.

[9] M. T. Cheah, A. Wachter, N. Sudarsan, and R. R. Breaker. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature*, 447(7143):497–500, May 2007.

[10] P. Clote. An efficient algorithm to compute the landscape of locally optimal RNA secondary structures with respect to the Nussinov-Jacobson energy model. *J. Computational Biology*, 1:83–101, 2005.

[11] P. Clote, J. Waldispuhl, B. Behzadi, and J.-M. Steyaert. Energy landscape of k-point mutants of an RNA molecule. *Bioinformatics*, 21(22):4140–4147, 2005.

[12] S. Commans and A. Böck. Selenocysteine inserting tRNAs: an overview. *FEMS Microbiology Reviews*, 23:333–351, 1999.

[13] R.A. Dimitrov and M. Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.*, 87:215–226, 2004.

[14] Y. Ding. Sfold. `http://sfold.wadsworth.org/index.pl`

[15] Y. Ding and C. E. Lawrence. A statistical sampling algorithm for RNA secondary structure prediction. *Nucleic Acids Res.*, 31:7280–7301, 2003.

[16] R.M. Dirks and N.A. Pierce. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J Comput Chem, 24(13):1664-1677, 2003*, 24(13):1664–1677, 2003.

[17] J.A. Doudna and T.R. Cech. The chemical repertoire of natural ribozymes. *Nature*, 418(6894):222–228, 2002.

[18] O. Dror, R. Nussinov, and H. J. Wolfson. The ARTS web server for aligning RNA tertiary structures. *Nucleic. Acids. Res.*, 34(Web):W412–W415, July 2006.

[19] R. Durbin, S. Eddy, A. Krogh, and G. Mitchison. *Biological Sequence Analysis: Probabalistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998.

[20] F. Ferre, Y. Ponty, W. A. Lorenz, and P. Clote. DIAL: a web server for the pairwise alignment of two RNA three-dimensional structures using nucleotide, dihedral angle and base-pairing similarities. *Nucleic Acids Res.*, 0(O):O, June 2007.

[21] E. Freyhult, V. Moulton, and P. Clote. Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics*, 2007. doi: 10.1093/bioinformatics/btm314.

[22] J.H. Havgaard, R. Lyngsø, G. Stormo, and J. Gorodkin. Pairwise local structural alignment of RNA sequences with sequence similarity less than 40%. *Bioinformatics*, 21(9):1815–1824, 2005.

[23] J. Heider, C. Baron, and A. Böck. Coding from a distance dissection of the mrna elements required for the incorporation of selenocysteine into protein. *EMBO J.*, 11:3759–3766, 1992.

[24] I.L. Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Res.*, 31:3429–3431, 2003.

[25] J.M. Hubbard and J.E. Hearst. Predicting the three-dimensional folding of transfer RNA with a computer modeling protocol. *Biochemistry*, 30:5458–5465, 1991.

[26] A. Hüttenhofer and A.Böck. RNA structures involved in selenoprotein synthesis. In *RNA structure and function*, pages 603–639. Cold Spring Harbor Laboratory Press, 1998.

[27] Jr. I. Tinoco and M. Schmitz. Thermodynamics of formation of secondary structure in nucleic acids. In E. Di Cera, editor, *Thermodynamics in Biology*, pages 131–176. Oxford University Press, 2000.

[28] I.L.Hofacker, W. Fontana, P.F. Stadler, L.S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsch. Chem.*, 125:167–188, 1994.

[29] S. Lemieux and F. Major. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.*, 30(19):4250–4263, 2002.

[30] N. Leontis and E. Westhof. Tools for the automatic identification and classification of RNA base pairs. *Nucl. Acids Res.*, 31(13):3450–3460, 2003.

[31] N.B. Leontis, J. Stombaugh, and E. Westhof. Motif prediction in ribosomal RNAs: Lessons and prospects for automated motif prediction in homologous RNA molecules. *Biochimie*, 84:961–973, 2002.

[32] L.P. Lim, M.E. Glasner, S. Yekta, C.B. Burge, and D.P. Bartel. Vertebrate microRNA genes. *Science*, 299(5612):1540, 2003.

[33] T. Lowe and S. Eddy. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleid Acids Research*, 25(5):955–964, 1997.

[34] F. Major, M. Turcotte, D. Gautheret, G. Lapalme, E. Fillion, and R. Cedergren. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science*, 1225–1260:253(5025), 1991.

[35] M. Mandal, M. Lee, J.E. Barrick, Z. Weinberg, G.M. Emilsson, W.L. Ruzzo, and R.R. Breaker. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science*, 306(5694):275–279, 2004.

[36] N.R. Markham and M. Zuker. DINAMelt web server for nucleic acid melting prediction. *Nucleic Acids Res.*, 33:W577–W581, 2005.

[37] D. H. Mathews and D. H. Turner. Prediction of RNA secondary structure by free energy minimization. *Curr. Opin. Struct. Biol.*, 16:270–278, 2006.

[38] D.H. Mathews and D.H. Turner. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, 317:191–203, 2002.

[39] D.H. Mathews, D.H. Turner, and M. Zuker. Secondary structure prediction. In S. Beaucage, D.E. Bergstrom, G.D. Glick, and R.A. Jones, editors, *Current Protocols in Nucleic Acid Chemistry*, pages 11.2.1–11.2.10. John Wiley & Sons, New York, 2000.

[40] D.H. Matthews, J. Sabina, M. Zuker, and D.H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.

[41] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.

[42] S. Moon, Y. Byun, H.-J. Kim, S. Jeong, and K. Han. Predicting genes expressed via $-1$ and $+1$ frameshifts. *Nucleic Acids Res.*, 32(16):4884–4892, 2004.

[43] H.Ogata nad Y. Akiyuna and M. Kanehisa. A genetic algorithm based molecular modeling technique for RNA stem-loop structures. *Nucleic Acids Research*, 23(3):419–426, 1995.

[44] P Nissen, J. Hansen, N. Ban, P.B. Moore, and T.A. Steitz. The structural basis of ribosome activity in peptide bond synthesis. *Science*, 289(5481):920–923, 2000.

[45] P. Nissen, J.A. Ippolito, N. Ban, P.B. Moore, and T.A. Steitz. RNA tertiary interactions in the large ribosomal subunit: The A-minor motifautomated identification of RNA. *Proc. Natl. Acad. Sci. USA*, 98(9):4899, 2001.

[46] R. Nussinov and A. B. Jacobson. Fast algorithm for predicting the secondary structure of single stranded RNA. *Proceedings of the National Academy of Sciences, USA*, 77(11):6309–6313, 1980.

[47] A.D. Omer, T.M. Lowe, A.G. Russell, H. Ebhardt, S.R. Eddy, and P.P. Dennis. Homologues of small nucleolar RNAs in Archaea. *Science*, 288:517–522, 2000.

[48] ENCODE Project Consortium. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146):799–816, 2007.

[49] J. Reeder and R. Giegerich. Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC. Bioinformatics*, 5:104, August 2004.

[50] E. Rivas and S.R. Eddy. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, 285:2053–2068, 1999.

[51] E. Rivas and S.R. Eddy. Noncoding RNA gene detection using comparative sequence analysis. *Bioinformatics*, 2(8), 2001.

[52] A. Serganov, Y.R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A.T. Phan, C. Hobartner, R. Micura, R.R. Breaker, and D.J. Patel. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, 11(12):1729–1741, 2004.

[53] S. Soro and A. Tramontano. The prediction of protein function at CASP6. *Proteins.*, 61:201–213, 2005.

[54] P. Steffen, B. Voss, M. Rehmsmeier, J. Reeder, and R. Giegerich. Rnashapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, 22(4):500–503, 2006.

[55] A.A. Szweczak and P.B. Moore. The sarcin/ricin loop, a modular RNA. *J. Mol. Biol.*, 247:81–98, 1995.

[56] S. Uliel, X. H. Liang, R. Unger, and S. Michaeli. Small nucleolar RNAs that guide modification in trypanosomatids: repertoire, targets, genome organisation, and unique functions. *Int. J. Parasitol.*, 34(4):445–454, March 2004.

[57] F.H.D. van Batenburg, A.P. Gultyaev, C.W.A. Pleij, J. Ng, and J. Oliehoek. PseudoBase: A database with RNA pseudoknots. *Nucl. Acids Res.*, 28(1):201–204, 2000.

[58] B. Voss, R. Giegerich, and M. Rehmsmeier. Complete probabilistic analysis of RNA shapes. *BMC Biol.*, 4(1):Epub ahead of print, 2006. 15 Feb 1006.

[59] J. Waldispühl and P. Clote. Computing the partition function and sampling for saturated secondary structures of RNA, with respect to the Turner energy model. *J. Computational Biology*, 14(2):190–215, 2007.

[60] S. Wuchty, W. Fontana, I.L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49:145–164, 1999.

[61] T. Xia, Jr. J. SantaLucia, M.E. Burkard, R. Kierzek, S.J. Schroeder, X. Jiao, C. Cox, and D.H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37:14719–35, 1999.

[62] K Yamaguchi and C.A. Del Carpio. A genetic programming based system for the prediction of secondary and tertiary structures of RNA. *Genome Informatics*, 9:382–383, 1998.

[63] H. Yand, F. Jossinet, N. Leontis, L. Chen, J. Westbrook, H. Berman, and E. Westhof. Tools for the automatic identification and classification of RNA base pairs. *Nucleic Acids Res.*, 31(13):3450–3460, 2003.

[64] M. Zuker. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, 31(13):3406–3415, 2003.

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