Chapter 1

MODELING RNA FOLDING

Ivo L. Hofacker
Institute for Theoretical Chemistry and Structural Biology, University of Vienna
ivo@tbi.univie.ac.at

Peter F. Stadler
Bioinformatics, Department of Computer Science, University of Leipzig
peter.stadler@bioinf.uni-leipzig.de

Abstract In recent years it has become evident that functional RNAs in living organisms are not just curious remnants from a primoridal RNA world but an ubiquitous phenomenon complementing protein enzyme based activity. Functional RNAs, just like proteins, depend in many cases upon their well-defined and evolutionarily conserved three-dimensional structure. In contrast to protein folds, however, RNA molecules have a biophysically important coarse-grained representation: their secondary structure. At this level of resolution at least, RNA structures can be efficiently predicted given only the sequence information. As a consequence, computational studies of RNA routinely incorporate structural information explicitly. RNA secondary structure prediction has proven useful in diverse fields ranging from theoretical models of sequence evolution and biopolymer folding, to genome analysis and even the design biotechnologically or pharmaceutically useful molecules.

Keywords: Functional RNA, Non-Coding RNA, RNA Secondary Structure Prediction, Conserved RNA Structures

1. Introduction

It is not hard to argue that **RNomics**, i.e., the understanding of functional RNAs (both ncRNA genes and functional motifs in protein-coding RNAs) and their interactions at a genomic level, is of utmost practical and theoretical importance in modern life sciences: The comprehensive understanding of the biology of a cell obviously requires the knowledge of the identity of *all* encoded

RNAs, the molecules with which they interact, and the molecular structures of these complexes (Doudna, 2000).

Structural genomics, the systematic determination of all macro-molecular structures represented in a genome, until very recently has been focused almost exclusively on proteins. Although it is commonplace to speak of *"genes and their encoded protein products"*, thousands of human genes produce transcripts that exert their function without ever producing proteins. The list of functional non-coding RNAs (ncRNAs) includes key players in the biochemistry of the cell. Many of them have characteristic secondary structures that are highly conserved in evolution. Databases (referenced below) collect the most important classes:

tRNA. Transfer RNAs are the adapters that translate the triplet nucleic acid code of RNA into the amino acid sequence of proteins. (Sprinzl et al., 1998)

rRNA. Ribosomal RNAs are central to the translational machinery. Recent results strongly indicate that peptide bond formation is catalyzed by rRNA. (Gutell et al., 2000; Szymanski et al., 2000; Van de Peer et al., 2000; Maidak et al., 2001; Wuyts et al., 2001)

snoRNA. Small nucleolar RNAs are required for rRNA processing and base modification in the eukaryotic nucleolus. (Samarsky and Fournier, 1999; Omer et al., 2000)

snRNA. Small nuclear RNAs are critical components of spliceosomes, the large ribonucleoprotein complexes that splice introns out of pre-mRNAs in the nucleus. (Zwieb, 1996)

tmRNA. The bacterial tmRNA (also known as 10Sa RNA or SsrA) was named for its dual tRNA-like and mRNA-like nature. tmRNA engages in a translation process, adding in trans a C-terminal peptide tag to the unfinished protein at a stalled ribosome. The tmRNA-directed tag targets the unfinished protein for proteolysis. (Zwieb and Wower, 2000; Williams, 2002)

RNase P. Ribonuclease P is responsible for the 5'-maturation of tRNA precursors. Ribonuclease P is a ribonucleoprotein, and in bacteria (and some Archaea) the RNA subunit alone is catalytically active in vitro, i.e. it is a ribozyme. (Brown, 1999). RNase MRP, which shares structural similarities with RNase P RNA, cleaves at a specific site in the precursor-rRNA transcript to initiate processing of the 5S rRNA

telRNA. Telomerase RNA. Telomeres are the specialized DNA protein structures at the ends of eukaryotic chromosomes. Telomerase is a ribonucleoprotein reverse transcriptase that synthesizes telomeric DNA. (Blackburn, 1999)

SRPRNA. The signal recognition particle is a universally conserved ribonucleoprotein. It is involved in the co-translational targeting of proteins to membranes. (Gorodkin et al., 2001)

miRNA. Micro-RNAs (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001) regulate gene expression by regulating mRNA expression by a mechanism closely linked to RNA interference by small double stranded RNAs, see e.g. (Bosher and Labouesse, 2000; Matzke et al., 2001). They are cleaved from their precursors, the small temporal RNAs (**stRNAs**), by the enzyme *Dicer*.

In addition, there is a diverse list of ncRNAs with sometimes enigmatic function. We give just a few examples, see also the `Rfam` database (Griffiths-Jones et al., 2003): The 17kb *Xist* RNA of humans and the smaller roX RNAs of *Drosophila* play a key role in dosage compensation and X chromosome inactivation (Avner and Heard, 2001; Franke and Baker, 2000). Several large ncRNAs are expressed from imprinted regions. Many of these are cis-antisense RNAs that overlap coding genes on the other genomic strand, see e.g. (Erdmann et al., 2001). An RNA (meiRNA) regulates the onset of meiosis in fission yeast (Ohno and Mattaj, 1999). Human vaults are intracellular ribonucleoprotein particles believed to be involved in multidrug resistance. The complex contains several small untranslated RNA molecules (van Zon et al., 2001). No precise function is known at present for the human H19 transcript, the hrsω transcript induced by heat shock in *Drosophila*, or the *E.coli* 6S RNA, see e.g. (Erdmann et al., 1999).

Even though the sequence of the human DNA is known by now, the contents of about half of it remains unknown. The diversity of sequences, sizes, structures, and functions of the known ncRNAs strongly suggests that we have seen only a small fraction of the functional RNAs. Most of the ncRNAs are small, they do not have translated ORFs, and they are not polyadenylated. Unlike protein coding genes, ncRNA gene sequences do not seem to exhibit a strong common statistical signal, hence a reliable general purpose computational genefinder for noncoding RNA genes has been elusive. It is quite likely therefore that a large class of genes has gone relatively undetected so far because they do not make proteins (Eddy, 2001).

Another level of RNA function is presented by functional motifs within protein-coding RNAs. A few of the best-understood examples of structurally conserved RNA motifs in viral RNAs

◦ An *IRES* (internal ribosomal entry site) region is used instead of a CAP to initialize translation by Picornaviridae, some Flaviviridae including Heptatitis C virus, and a small number of mRNAs, see e.g. (Rueckert, 1996; Huez et al., 1998; Pesole et al., 2001).

◦ The TAR hairpin structure in HIV and related Retroviruses is the target for viral transactivation.

◦ The RRE structure of Retroviruses serves as binding site for the Rev protein and is essential for the viral replication. The RRE is a characteristic five-fingered structural motif, see e.g. (Dayton et al., 1992).

◦ The CRE hairpin (Witwer et al., 2001) in Picornaviridae is vital for replication.

Genes in eukaryotes are often interrupted by intervening sequences, *introns*, that must be removed during gene expression. Similarly, rRNAs are produced from a pre-rRNA that contains so-called internal and external transcribed spacers. These contain regions with characteristic secondary structures (Denduangboripant and Cronk, 2001). RNA splicing is the process by which these parts are precisely removed from the pre-mRNA and the flanking, functional exons are joined together (Green, 1991). Regulated mechanisms of alternative splicing allow multiple different proteins to be translated from a single RNA transcript. Mutations can affect splicing of certain introns, leading to abnormal conditions. For example a form of thalassemia, a blood disorder, is due to a mutation causing splicing failure of an intron in a globin transcript, which then becomes untranslatable, see e.g. (Stoss et al., 2000). The splicing of most nuclear genes is performed by the spliceosome; however, in many cases the splicing reaction is self-contained; that is, the intron — with the help of associated proteins — splices itself out of the precursor RNA, see e.g. (Mattick, 1994) for a review.

A textbook example of a functional RNA secondary structure is the *Rho*-independent termination in *E.coli*. The newly synthesized mRNA forms a hairpin in the 3'NTR that interacts with the RNA polymerase causing a change in conformation and the subsequent dissociation of the Enzyme-DNA-RNA complex. For a computational analysis of the *Rho*-independent transcription terminators we refer to (d'Aubenton Carafa et al., 1990).

Only part of the mature RNA is translated into a protein. At the beginning of the mRNA, just behind the cap, is a non-coding sequence, the so-called leader sequence (10-200nt) that may be followed by another non-coding sequence of up to 600nt. An increasing number of functional features in the untranslated regions of of eukaryotic mRNA have been reported in recent years (Pesole et al., 2001; Jacobs et al., 2002).

An extreme example are the Early Noduline genes. Enod40, which is expressed in the nodule primordium developing in the root cortex of leguminous plants after infection by symbiotic bacteria (Sousa et al., 2001), codes for an RNA of about 700nt that gives rise to two short peptides, 13 and 27 amino acids, respectively. The RNA structure itself exhibits significant conservation of secondary structure motifs (Hofacker et al., 2002), and might take part in localization of mRNA translation (Oleynikov and Singer, 1998), as in the case of the bicoid gene *bcd* of *Drosophila* (Macdonald, 1990).

2. RNA Secondary Structures and Their Prediction

As with all biomolecules, the function of RNAs is intimately connected to their structure. It does not come as a surprise therefore, that most the the classes of functional RNAs listed in the introduction have, like the well-known *clover-*

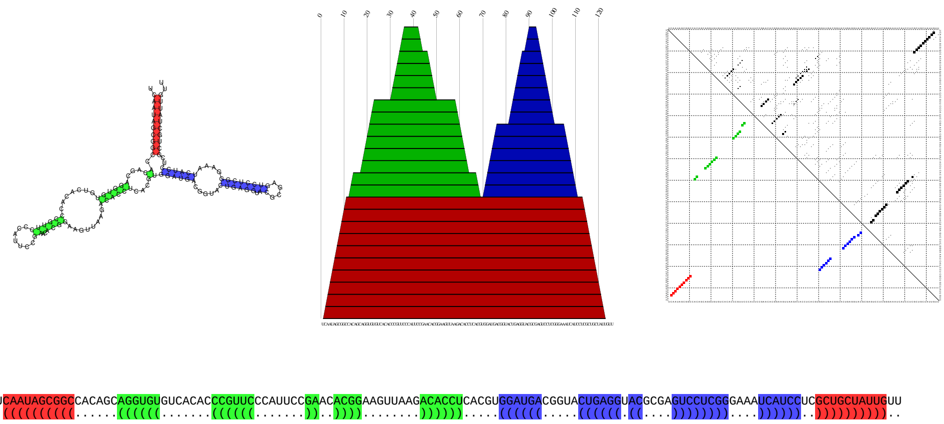

Figure 1.1. RNA secondary structure of an 5S ribosomal RNA. Secondary structure graph (left), mountain representations (middle), dot plot (right), and bracket notation (bottom).

leaf structure of tRNAs, distinctive structural characteristics. While successful predictions of RNA tertiary structure remain exceptional feats, RNA secondary structures can be predicted with reasonable accuracy, and have proven to be a biologically useful description.

A secondary structure of a given RNA sequence is the list of (Watson-Crick and wobble) base pairs satisfying two constraints: (i) each nucleotide takes part in at most one base pair, and (ii) base pairs do not cross, i.e., there are no knots or pseudo-knots. While pseudo-knots are important in many natural RNAs (Westhof and Jaeger, 1992), they can be considered part of the tertiary structure for our purposes. Secondary structure can be represented in various equivalent ways, see Fig. 1.1.

The restriction to knot-free structures is necessary for efficient computation by means dynamic programming algorithms (Nussinov et al., 1978; Waterman, 1978; Zuker and Stiegler, 1981; Zuker and Sankoff, 1984; Zuker, 1989; McCaskill, 1990; Schmitz and Steger, 1992; Hofacker et al., 1994; Wuchty et al., 1999; Hofacker et al., 2002). The memory and CPU requirements of these algorithms scale with sequence length n as $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$, respectively, making structure prediction feasible even for large RNAs of about 10000 nucleotides, such as the genomes of RNA viruses (Hofacker et al., 1996; Huynen et al., 1996a; Witwer et al., 2001). There are two implementations of various variants of these dynamic programming algorithms: the `mfold` package by Michal Zuker, and the the `Vienna RNA Package` by the present authors and their collaborations. The latter is freely available from `http://www.tbi.univie.ac.at/`.

These thermodynamic folding algorithms are based on an energy model that considers additive contributions from stacked base pairs and various types of loops, see e.g. (Walter et al., 1994; Mathews et al., 1999). Two widely used methods for determining nucleic acid thermodynamics are absorbance melting curves and microcalorimetry, see (SantaLucia Jr. and Turner, 1997) for a review.

Recently, algorithms have been described that are able to deal with certain classes of pseudo-knotted structures, however at considerable computational cost (Rivas and Eddy, 1999; Akutsu, 2001; Lyngsø and Pedersen, 2000; Giegerich and Reeder, 2003). Alternatively, heuristics such as genetic algorithms can be used (Lee and Han, 2002). A common problem of all these approaches is the still very limited information about the energetics of pseudoknots (Gultyaev et al., 1999; Isambert and Siggia, 2000).

3. Neutral Networks in Sequence Space

A more detailed analysis of functional classes of RNAs shows that their structures are very well conserved while at the same time there may be little similarity at the sequence level, indicating that the structure has actual importance for the function of the molecule. In order to understand the evolution of functional RNAs one therefore has to understand the relation between sequence (genotype) and structure (phenotype). Although qualitatively there is ample evidence for neutrality in natural evolution as well as in experiments under controlled conditions in the lab, very little is known about regularities in general genotype-phenotype relations. In the RNA case, however, the phenotype can be approximated by the minimum free energy structure of RNA, see e.g. (Schuster, 2001) for a recent review. This simplifying assumption is met indeed by RNA evolution experiments *in vitro* (Biebricher and Gardiner, 1997) as well as by the design of RNA molecules through artificial selection (Wilson and Szostak, 1999).

There is ample evidence for redundancy in genotype-phenotype maps f in the sense that many genotypes cannot be distinguished by an evolutionarily relevant coarse grained notion of phenotypes which, in turn, give rise to fitness values that cannot be faithfully separated through selection. Regarding the folding algorithms as a map f that assigns a structure $s = f(x)$ to each sequence x we can phrase our question more precisely: We need to know how the set of sequences $f^{-1}(s)$ that folds into a given structure s is embedded in the sequence space (where the genotypes are interpreted as nodes and all Hamming distance one neighbors are connected by an edge). The subgraphs of the sequence space that are defined by the sets $f^{-1}(s)$ are called *neutral networks* (Schuster et al., 1994).

The most important global characterization of neutral networks is its average fraction of neutral neighbor $\bar{\lambda}$, usually called the (degree of) neutrality. Neglecting the influence of the distribution of neutral sequences over sequence space, the degree of neutrality will increase with size of the pre-image. Generic properties of neutral networks (Reidys et al., 1997) are readily derived by means of a random graph model. Theory predicts a phase transition like change in the appearance of neutral networks with increasing degree of neutrality at a critical value:

$$\lambda_{\mathrm{cr}} = 1 - \kappa^{-\frac{1}{\kappa-1}} , \qquad (1.1)$$

where κ is the size of the genetic alphabet. For example, $\kappa = 4$ for the canonical genetic alphabet $\{\mathsf{A}, \mathsf{U}, \mathsf{G}, \mathsf{C}\}$. If $\bar{\lambda} < \lambda_{\mathrm{cr}}$ then the network consists of many isolated parts with one dominating giant component. On the other hand, the network is generically connected if $\bar{\lambda} > \lambda_{\mathrm{cr}}$. The critical value λ_{cr} is the connectivity threshold. This property of neutral networks reminds of percolation phenomena known from different areas of physics, although the high symmetry of sequence space, with all points being equivalent, introduces a difference in the two concepts.

A series of computational studies (Fontana et al., 1993b; Fontana et al., 1993a; Schuster et al., 1994; Huynen et al., 1996b; Grüner et al., 1996a; Grüner et al., 1996b; Fontana and Schuster, 1998a; Fontana and Schuster, 1998b) has in the last decade drawn a rather detailed picture of the genotype-phenotype map of RNA, see also Fig. 1.2.

(i) **More sequences than structures**. For sequence spaces of chain lengths $n \geq 10$ there are orders of magnitude more sequences than structures and hence, the map is many-to-one.

(ii) **Few common and many rare structures**. Relatively few common structures are opposed by a relatively large number of rare structures, some of which are formed by a single sequence only ("relatively" points at the fact that the numbers of both common and rare structures increase exponentially with n, but the exponent for the common structures is smaller than that for the rare ones).

(iii) **Shape space covering**. The distribution of neutral genotypes, these are sequences that fold into the same structure, is approximately random in sequence space. As a result it is possible to define a spherical ball, with a diameter d_{cov} being much smaller than the diameter n of sequence space, which contains on the average for every common structure at least one sequence that folds into it.

(iv) **Existence and connectivity of neutral networks**. Neutral networks, being pre-images of phenotypes or structures in sequence space, of com-

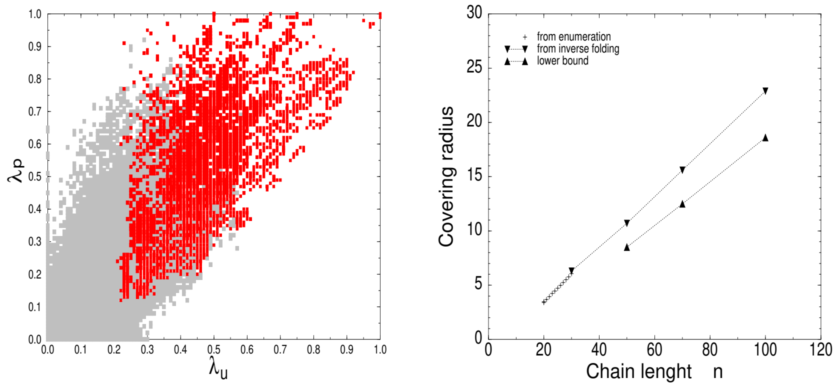

Figure 1.2. Neutral Networks and Shape Space Covering.
(a) Neutral networks in an exhaustive survey of the GC sequence space with length $n = 30$ (Grüner et al., 1996b) are fragmented (light grey) if the fractions λ_u and λ_p of neutral mutations in the unpaired and paired parts of the sequence are below a threshold value. Above the threshold the the neutral networks consist of one to four connected components. The fragmentation of the single connected component into a small number of (barely) separated subsets can be explained by details of energy based folding model, see (Schuster and Stadler, 1998).
(b) The shape space covering radius $d_{\rm cov}$ scales linearly with the chain length n with a slope $\vartheta \approx 1/4$. Data are taken from (Grüner et al., 1996b).

mon structures are connected unless specific and readily recognizable special features of RNA structures require specific non-random distribution in the $\{\mathsf{A}, \mathsf{U}, \mathsf{G}, \mathsf{C}\}$ sequence space, $\mathcal{Q}^{(\mathsf{AUGC})}$ (For structures formed from sequences over a $\{\mathsf{G}, \mathsf{C}\}$ alphabet the connectivity threshold is higher, whereas, at the same time, the mean number of neutral neighbors is smaller).

Shape space covering, item (iii) above, is a consequence of the high susceptibility of RNA secondary structures towards randomly placed point mutations. Computer simulations (Fontana et al., 1993a; Schuster et al., 1994) showed that a small number of point mutations is very likely to cause large changes in the secondary structures: mutations in 10% of the sequence positions already lead almost surely to unrelated structures if the mutated positions are chosen randomly.

The set of nodes of the neutral network $f^{-1}(s)$ is embedded in a compatible set $C(s)$ which includes all sequences that can form the structure s as suboptimal *or* minimum free energy conformation $f^{-1}(S) \subseteq C(s)$. Sequences at the intersection $C(s') \cap C(s'')$ of the compatible sets of two neutral networks in the same sequence space are of actual interest because these sequences can simultaneously carry properties of the different RNA folds. For example, they can exhibit catalytic activities of two different ribozymes at the same time (Schultes and Bartel, 2000). The intersection theorem (Reidys et al., 1997) states that for all pairs of structures s' and s'' the intersection $C(s') \cap C(s'')$ is always non-empty. In other words, for each arbitrarily chosen pair of structures there will be at least one sequence that can form both. If s' and s'' are both common structures, bistable molecules that have equal preference for both structures are easy to design (Flamm et al., 2000; Höbartner and Micura, 2003). A particularly interesting experimental case is described in (Schultes and Bartel, 2000).

At least, the features (i), (ii), and (iv) of the neutral networks of RNA seem to hold for the more complicated protein spaces as well (Babajide et al., 1997; Babajide et al., 2001), see e.g. (Keefe and Szostak, 2001) for experimental data.

The impact of these features on evolutionary dynamics is discussed in detail in (Huynen et al., 1996b; Schuster, 1995): A population explores sequence space in a diffusion-like manner along the neutral network of a viable structure. Along the fringes of the population novel structures are produced by mutation at a constant rate (Huynen, 1996). Fast diffusion together with perpetual innovation makes these landscapes ideal for evolutionary adaptation (Fontana and Schuster, 1998b) and sets the stage for the evolutionary biotechnology of RNA (Schuster, 1995).

4. Conserved RNA Structures

As we have seen, even a small number of randomly placed point mutations very likely leads to a complete disruption of the RNA structure. Secondary structure elements that are consistently present in a group of sequences with less than, say, 95% average pairwise identity are therefore almost certainly the result of stabilizing selection, not a consequence of the high degree of sequence conservation. If selection acts to preserve structure, then this structure must have some function. It is of considerable practical interest therefore to efficiently compute the consensus structure of a collection of such RNA molecules.

A promising approach towards this goal is the combination of the the "phylogenetic" information that is contained in the sequence co-variations and the information on the (local) thermodynamic stability of the molecules. Such methods for predicting RNA conserved and consensus secondary structure fall into two broad groups: those starting from a multiple sequence alignment and algorithms that attempt to solve the alignment problem and the folding problem simultaneously. The main disadvantage of the latter class of methods (Sankoff, 1985; Tabaska and Stormo, 1997; Gorodkin et al., 1997a; Gorodkin et al., 1997b) is their high computational cost, which makes them unsuitable for long sequences such as 16S or 23S RNAs. Most of the alignment based methods start from thermodynamics-based folding and use the analysis of sequence co-variations or mutual information for post-processing, see e.g., (Le and Zuker, 1991; Lück et al., 1996; Lück et al., 1999; Juan and Wilson, 1999; Hofacker et al., 2002). The converse approach is taken in (Han and Kim, 1993), where ambiguities in the phylogenetic analysis are resolved based on thermodynamic considerations.

It is important to clearly distinguish the consensus structure of a set of RNA sequences from the collection of structural features that are conserved among these sequences. Whenever there are reasons to assume that the structure of the whole molecule is conserved one may attempt to compute a consensus structure. On the other hand, consensus structures are unsuitable when a significant part of the molecule has no conserved structures. RNA virus genomes, for instance contain only local structural patterns (such as the IRES in Picorna viruses or the TAR hairpin in HIV). Such features can be identified with a related approach that is implemented in the algorithm `alidot` algorithm (Hofacker et al., 1998; Hofacker and Stadler, 1999). This program ranks base pairs using both the thermodynamic information contained in the base pairing probability matrix and the information on compensatory, consistent, and inconsistent mutations contained in the multiple sequence alignment. The approach is different from other efforts because it does not assume that the sequences have a single common structure. In this sense `alidot` combines structure prediction and motif search (Dandekar and Hentze, 1995). An implementation of this algorithm is available from `http://www.tbi.univie.ac.at/`. This approach to surveying functional structures goes beyond search software such as `RNAmot` (Gautheret et al., 1990) in that it does not require any *a priori* knowledge of the functional structure motifs and it goes beyond searches for regions that are thermodynamically especially stable or well-defined (Jacobson and Zuker, 1993) in that it returns a specific prediction for a structure if *and only if* there is sufficient evidence for structural conservation.

Of course, it is not possible to determine the function of a conserved structure or structural element without additional experimental input. Nevertheless, knowledge about their location can be used to guide, for instance, deletion studies (Mandl et al., 1998). Knowledge of both protein coding regions and functional RNA structures in the viral genome is needed e.g. to rationally design attenuated mutants for vaccine development.

Structure predictions of a set of sequences are conveniently summarized in the form of Hogeweg-style mountain plots (Hogeweg and Hesper, 1984), Fig. 1.3.

The computation of consensus and conserved RNA structures has been used to compile an *Atlas* of potentially functional RNA motifs in RNA virus genomes. Detailed data are available at present for Picornaviridae (Witwer et al., 2001),

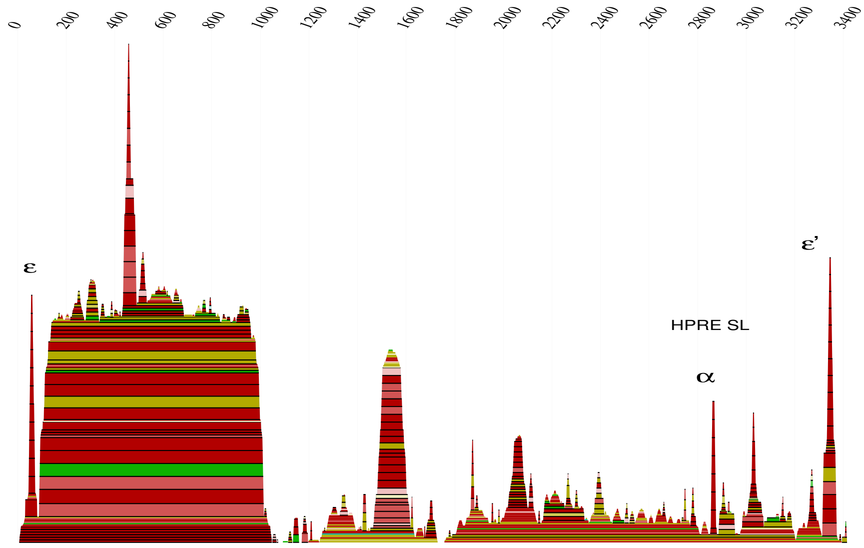

Figure 1.3. Predicted functional RNA structures in the genome of Hepatitis B Virus. The function of the ε, ε', and the α element of the HPRE region, have been determined experimentally. The prediction suggests several new conserved structures with unknown function.
In the "mountain representation" each base pair (i, j) is represented by a bar from i to j. The thickness of the bar indicates its probability or the reliability of the prediction. A color scheme can be used used to indicating sequence covariations. Hue encodes the number of compensatory and consistent mutations, while reduced saturation indicates that a small number of sequences is inconsistent with the structure.

Hepatitis B virus (Stocsits et al., 1999; Kidd-Ljunggren et al., 2000), and Flaviviridae (Thurner et al., 2003).

5. Discussion

Structural genomics, the systematic determination of all macro-molecular structures represented in a genome, is at present focused almost exclusively on proteins. Over the past two decades it has become clear, however, that a variety of RNA molecules have important, and sometimes essential, biological functions beyond their roles as rRNAs, tRNAs, or mRNAs. Given a handful of related RNA sequences, reliable methods now exist to predict conserved functional RNA structures within these RNAs. Because of their small size and fast evolution the genomes of RNA viruses supply fertile ground for such approaches, and databases of functional viral RNA structures are being built up. These functional RNA motifs in the viral genome are just as essential as the encoded proteins, and thus just as promising targets for the development of drugs and vaccines (Mandl et al., 1998; Ying et al., 1999).

The importance of regulatory functions mediated by RNA has only now found more attention through recent studies on the phenomenon of RNA interference (Cogoni and Macino, 2000; Guru, 2000; Hammond et al., 2001). A recent study (Wang et al., 2002) showed, furthermore, that non-coding RNA motives may act as potent "danger motifs" that trigger an adaptive immune system via innate immune receptors. RNA structure thus receives increased attention in molecular medicine.

A comprehensive understanding of the the biology of a cell will ultimately require the knowledge of all encoded RNAs, the molecules with which they interact, and the molecular structures of these complexes (Doudna, 2000). Various approaches to surveying genomic sequences for putative RNA genes have been devised in last few years.

Structure based searches use the known secondary structure of the major classes of functional RNAs. Programs such as `RNAmot` (Gautheret et al., 1990), `tRNAscan` (Lowe and Eddy, 1997), `HyPa` (Gräf et al., 2001), `RNAMotif` (Macke et al., 2001), `bruce` (Laslett et al., 2002), and many others exploit this avenue. An interesting variant that makes use of evolutionary computation is described by (Fogel et al., 2002). Nevertheless, all these approaches are restricted to searching for new members of the few well-established families. The web-based resource `RNAGENiE` uses a neural network that has been trained on a wide variety of functional RNAs (Carter et al., 2001). It is capable of detecting a wider variety of functional RNAs.

Some noncoding RNAs can be found by searching for likely transcripts that do not contain an open reading frame. A survey of the *Escherichia coli* genome for DNA regions that contain a $\sigma 70$ promotor within a short distance of a *Rho*-independent terminator, for instance, resulted in 144 novel possible ncRNAs (Chen et al., 2002). This approach is limited, however, to functional RNAs that are transcribed in the "usual" manner.

Comparative approaches such as the program `QRNA` (Rivas and Eddy, 2001) can detect novel structural RNA genes in a pair of aligned homologous sequences by deciding whether the substitution pattern fits better with (a) synonymous substitutions, which are expected in protein-coding regions, (b) the compensatory mutations consistent with some base-paired secondary structure, or (c) uncorrelated mutations.

Another approach tries to determine functional RNAs by means of structure prediction. The basic assumption is that functional and hence conserved structures will be thermodynamically more stable (Le et al., 1988; Huynen et al., 1996a). While such procedures are capable of detecting some particularly stable features, a recent study (Rivas and Eddy, 2000) concludes that "although a distinct, stable secondary structure is undoubtedly important in most noncoding RNAs, the stability of most noncoding RNA secondary structures is not sufficiently different from the predicted stability of a random sequence to be useful as a general genefinding approach." Nevertheless, in some special cases such as hyperthermophilic organisms, GC-content (and hence thermodynamic stability) proved sufficient (Klein et al., 2002).

Since most classes of functional RNAs are relatively well conserved while their sequences show little similarities, both comparative procedures and search in single sequences have to rely on structural information. While the prediction of RNA tertiary structures faces much the same problems as protein structure prediction, efficient algorithms exist for handling RNA secondary structure. As we have seen, these methods provide powerful tools for computational studies of RNA structure.

Acknowledgments: This work is supported by the Austrian *Fonds zur Förderung der Wissenschaftlichen Forschung*, Project Nos. P-13545-MAT and P-15893, and the *DFG* Bioinformatics Initiative.

References

Akutsu, Tatsuya (2001). Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discr. Appl. Math.*, 104:45–62.

Avner, P. and Heard, E. (2001). X-chromosome inactivation: counting, choice, and initiation. *Nat. Rev. Genet.*, 2:59–67.

Babajide, Aderonke, Farber, Robert, Hofacker, Ivo L., Inman, Jeff, Lapedes, Alan S., and Stadler, Peter F. (2001). Exploring protein sequence space using knowledge based potentials. *J. Theor. Biol.*, 212:35–46.

Babajide, Aderonke, Hofacker, Ivo L., Sippl, Manfred J., and Stadler, Peter F. (1997). Neutral networks in protein space: A computational study based on knowledge-based potentials of mean force. *Folding & Design*, 2:261–269.

Biebricher, C. K. and Gardiner, W. C. (1997). Molecular evolution of RNA *in vitro*. *Biophys. Chem.*, 66:179–192.

Blackburn, E. (1999). Telomerase. In Gesteland, R, Cech, T, and Atkins, J, editors, *The RNA World*, pages 609–635. Cold Spring Harbor Laboratory Press, New York.

Bosher, Julia M. and Labouesse, Michel (2000). RNA interference: genetic wand and genetic watchdog. *Nature Cell Biol.*, 2:E31–E36.

Brown, J. W. (1999). The ribonuclease P database. *Nucl. Acids Res.*, 27:314–314.

Carter, Richard J., Dubchak, Inna, and Holbrook, Stephen R. (2001). A computational approach ot indentify genes for functional RNAs in genomic sequences. *Nucl. Acids Res.*, 29:3928–3938.

Chen, Shuo, Lesnik, Elena A., Hall, Thomas A., Sampath, Rangarajan, Griffey, Richard H., Ecker, Dave J., and . Blyn, Lawrence B. (2002). A bioinformatics based approach to discover small RNA genes in the *Escherichia coli* genome. *BioSystems*, 65:157–177.

Cogoni, C. and Macino, G. (2000). Post-transcriptional gene silencing across kingdoms. *Genes Dev.*, 10:638–643.

Dandekar, T. and Hentze, M. W. (1995). Finding the hairpin in the haystack: searching for RNA motifs. *Trends. Genet.*, 11:45–50.

d'Aubenton Carafa, Y., Brody, E., and Thermes, C. (1990). Prediction of rho-independent escherichia coli transcription terminators. a statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, 216:835–858.

Dayton, E. T., Konings, D. A. M., Powell, D. M., Shapiro, B. A., Butini, L., Maizel, J. V., and Dayton, A. I. (1992). Extensive sequence-specific information throughout the CAR/RRE, the target sequence of the human immunodeficiency virus type 1 Rev protein. *J. Virol.*, 66:1139–1151.

Denduangboripant, Jessada and Cronk, Quentin C. B. (2001). Evolution and alignment of the hypervariable arm 1 of aeschynanthus (gesneriaceae) ITS2 nuclear ribosomal DNA. *Mol. Phylog. Evol.*, 20:163–172.

Doudna, Jennifer A. (2000). Structural genomics of RNA. *Nature Struct. Biol.*, 7:954–956.

Eddy, Sean R. (2001). Non-coding RNA genes and the modern RNA world. *Nature Genetics*, 2:919–929.

Erdmann, V. A., Barciszewska, M. Z., Hochberg, A., de Groot, N., and Barciszewski, J. (2001). Regulatory RNAs. *Cell. Mol. Life Sci.*, 58:960–977.

Erdmann, V. A., Szymanski, M., Hochberg, A., de Groot, N., and Barciszewski, J. (1999). Collection of mRNA-like non-coding rnas. *Nucleic Acids Res.*, 27:192–195.

Flamm, Christoph, Fontana, Walter, Hofacker, Ivo, and Schuster, Peter (2000). RNA folding kinetics at elementary step resolution. *RNA*, 6:325–338.

Fogel, Gary B., Porto, V. William, Weekes, Dana G., Fogel, David B., Griffey, Richard H., McNeil, John A., Lesnik, Elena, Ecker, David J., and Sampath, Rangarajan (2002). Discovery of RNA structural elements using evolutionary computation. *Nucl. Acids Res.*, 30:5310–5317.

Fontana, W., Konings, D. A. M., Stadler, P. F., and Schuster, P. (1993a). Statistics of RNA secondary structures. *Biopolymers*, 33:1389–1404.

Fontana, W. and Schuster, P. (1998a). Continuity in Evolution: On the Nature of Transitions. *Science*, 280:1451–1455.

Fontana, W. and Schuster, P. (1998b). Shaping Space: The Possible and the Attainable in RNA Genotype-Phenotype Mapping. *J. Theor. Biol.*, 194:491–515.

Fontana, Walter, Stadler, Peter F, Bornberg-Bauer, Erich G, Griesmacher, Thomas, Hofacker, Ivo L, Tacker, Manfred, Tarazona, Pedro, Weinberger, Edward D, and Schuster, Peter (1993b). RNA folding landscapes and combinatory landscapes. *Phys. Rev. E*, 47:2083–2099.

Franke, A. and Baker, B. S. (2000). Dosage compensation rox! *Curr. Opin. Cell Biol.*, 12:351–354.

Gautheret, D., Major, F., and Cedergren, R. (1990). Pattern searching/alignment with RNA primary and secondary structures: an effective descriptor for tRNA. *Comput. Appl. Biosci.*, 6:325–331.

Giegerich, Robert and Reeder, Jens (2003). From RNA folding to thermodynamic matching including pseudoknots. Technical Report 2003-03, Universität Bielefeld.

Gorodkin, J., Heyer, L. J., and Stormo, G. D. (1997a). Finding common sequences and structure motifs in a set of RNA molecules. In Gaasterland, T., Karp, P., Karplus, K., Ouzounis, Ch., Sander, Ch., and Valencia, A., editors, *Proceedings of the ISMB-97*, pages 120–123, Menlo Park, CA. AAAI Press.

Gorodkin, J., Heyer, L. J., and Stormo, G. D. (1997b). Finding the most significant common sequence and structure motifs in a set of rna sequences. *Nucleic Acids Res.*, 25:3724–3732.

Gorodkin, J., Knudsen, B., Zwieb, C., and Samuelsson, T. (2001). SRPDB (signal recognition particle database). *Nucleic Acids Res.*, 29:169–170.

Gräf, Stefan, Strothmann, Dirk, Kurtz, Stefan, and Steger, Gerhard (2001). HyPaLib: a database of RNAs and RNA structural elements defined by hybrid patterns. *Nucl. Acids. Res.*, 29:196–198.

Green, M. R. (1991). Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.*, 7:559–599.

Griffiths-Jones, Sam, Bateman, Alex, Marshall, Mhairi, Khanna, Ajay, and Eddy, Sean R. (2003). Rfam: an RNA family database. *Nucl. Acids Res.*, 31:439–441.

Grüner, Walter, Giegerich, Robert, Strothmann, Dirk, Reidys, Christian M., Weber, Jacqueline, Hofacker, Ivo L., Stadler, Peter F., and Schuster, Peter (1996a). Analysis of RNA sequence structure maps by exhaustive enumeration. I. Neutral networks. *Monath. Chem.*, 127:355–374.

Grüner, Walter, Giegerich, Robert, Strothmann, Dirk, Reidys, Christian M., Weber, Jacqueline, Hofacker, Ivo L., Stadler, Peter F., and Schuster, Peter (1996b). Analysis of RNA sequence structure maps by exhaustive enumeration. II. Structures of neutral networks and shape space covering. *Monath. Chem.*, 127:375–389.

Gultyaev, A. P., van Batenburg, F. H. D., and Pleij, C. W. A. (1999). An approximation of loop free energy values of RNA H-pseudoknots. *RNA*, 5:609–617.

Guru, T. (2000). A silence that speaks volumes. *Nature*, 404:804–808.

Gutell, Robin R., Cannone, J. J., Shang, Z., Du, Y., and Serra, M. J. (2000). A story: Unpaired adenosine bases in ribosomal rna. *J. Mol. Biol.*, 304:335–354.

Hammond, S. M., Caudy, A. A., and Hannon, G. J. (2001). Post-transcriptional gene silencing by double-stranded RNA. *Nature Rev. Gen.*, 2:110–119.

Han, Kyungsook and Kim, Hong-Jin (1993). Prediction of common folding structures of homologous RNAs. *Nucl. Acids Res.*, 21:1251–1257.

Höbartner, Claudia and Micura, Ronald (2003). Bistable secondary structures of small rnas and their structural probing by comparative imino proton nmr spectroscopy. *J. Mol. Biol.*, 325:421–431.

Hofacker, Ivo L., Fekete, Martin, Flamm, Christoph, Huynen, Martijn A., Rauscher, Susanne, Stolorz, Paul E., and Stadler, Peter F. (1998). Automatic detection of conserved RNA structure elements in complete RNA virus genomes. *Nucl. Acids Res.*, 26:3825–3836.

Hofacker, Ivo L., Fekete, Martin, and Stadler, Peter F. (2002). Secondary structure prediction for aligned RNA sequences. *J. Mol. Biol.*, 319:1059–1066.

Hofacker, Ivo L., Fontana, Walter, Stadler, Peter F., Bonhoeffer, Sebastian, Tacker, Manfred, and Schuster, Peter (1994). Fast folding and comparison of RNA secondary structures. *Monatsh. Chemie*, 125:167–188.

Hofacker, Ivo L., Huynen, Martijn A., Stadler, Peter F., and Stolorz, Paul E. (1996). Knowledge discovery in RNA sequence families of HIV using scalable computers. In Simoudis, Evangelos, Han, Jiawei, and Fayyad, Usama, editors, *Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining, Portland, OR*, pages 20–25, Menlo Park, CA. AAAI Press.

Hofacker, Ivo L. and Stadler, Peter F. (1999). Automatic detection of conserved base pairing patterns in RNA virus genomes. *Comp. & Chem.*, 23:401–414.

Hogeweg, Pauline and Hesper, B. (1984). Energy directed folding of RNA sequences. *Nucl. Acids Res.*, 12:67–74.

Huez, Isabelle, Créancier, Laurent, Audigier, Sylvie, Gensac, Marie-Claire, Prats, Anne-Catherine, and Prats, Hervé (1998). Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell. Biol.*, 18:6178–6190.

Huynen, M. A., Perelson, A. S., Viera, W. A., and Stadler, P. F. (1996a). Base pairing probabilities in a complete HIV-1 RNA. *J. Comp. Biol.*, 3:253–274.

Huynen, M. A., Stadler, P. F., and Fontana, W. (1996b). Smoothness within ruggedness: The role of neutrality in adaptation. *Proc. Natl. Acad. Sci. USA*, 93:397–401.

Huynen, Martijn A. (1996). Exploring phenotype space through neutral evolution. *J. Mol. Evol.*, 43:165–169.

Isambert, Hervé and Siggia, Eric D. (2000). Modeling RNA folding paths with pseudoknots: Application to hepatitis delta virus ribozyme. *Proc. Natl. Acad. Sci. USA*, 97:6515–6520.

Jacobs, G. H., Rackham, O., Stockwell, P. A., W, Tate, and M., Brown C. (2002). Transterm: a database of mRNAs and translational control elements. *Nucl. Acids Res.*, 30:310–311.

Jacobson, A. B. and Zuker, M. (1993). Structural analysis by energy dot plot of large mRNA. *J. Mol. Biol.*, 233:261–269.

Juan, Veronica and Wilson, Charles (1999). RNA secondary structure prediction based on free energy and phylogenetic analysis. *J. Mol. Biol.*, 289:935–947.

Keefe, A. D. and Szostak, J. W. (2001). Functional proteins from a random-sequence library. *Nature*, 410:715–718.

Kidd-Ljunggren, K., Zuker, M., Hofacker, Ivo L., and Kidd, A. H. (2000). The hepatitis B virus pregenome: prediction of RNA structure and implications for the emergence of deletions. *Intervirology*, 43:154–64.

Klein, Robert J., Misulovin, Ziva, and Eddy, Sean R. (2002). Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proc. Natl. Acad. Sci. USA*, 99:7542–7547.

Lagos-Quintana, Mariana, Rauhut, Reinhard, Lendeckel, Winfried, and Tuschl, Thomas (2001). Identification of novel genes coding for small expressed RNAs. *Science*, 294:853–857.

Laslett, Dean, Canback, Bjorn, and Andersson, Siv (2002). `BRUCE`: a program for the detection of transfer-messenger RNA genes in nucleotide sequences. *Nucl. Acids Res.*, 30:3449–3453.

Lau, N. C., Lim, L. P, Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *caenorhabditis elegans*. *Science*, 294:858–862.

Le, S-Y., Chen, J-H., Currey, K.M., and Maizel, J.V. (1988). A program for predicting significant RNA secondary structures. *CABIOS*, 4:153–159.

Le, S. Y. and Zuker, M. (1991). Predicting common foldings of homologous rnas. *J. Biomol. Struct. Dyn.*, 8:1027–1044.

Lee, Dongkyu and Han, Kyungsook (2002). Prediction of rna pseudoknots — comparative study of genetic algorithms. *Genome Informatics*, 13:414–415.

Lee, R. C. and Ambros, V. (2001). An extensive class of small RNAs in *caenorhabditis elegans*. *Science*, 294:862–864.

Lowe, T. M. and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucl. Acids Res.*, 25:955–964.

Lück, R., Gräf, S., and Steger, G. (1999). ConStruct: A tool for thermodynamic controlled prediction of conserved secondary structure. *Nucl. Acids Res.*, 27:4208–4217.

Lück, R., Steger, G., and Riesner, D. (1996). Thermodynamic prediction of conserved secondary structure: Application to the RRE element of HIV, the tRNA-like element of CMV, and the mRNA of prion protein. *J. Mol. Biol.*, 258:813–826.

Lyngsø, Rune B. and Pedersen, Christian N. S. (2000). Rna pseudoknot prediction in energy-based models. *J. Comp. Biol.*, 7:409–427.

Macdonald, P. M. (1990). Bicoid mRNA localization signal: Phylogenetic conservation of function and RNA secondary structure. *Development*, 110:161–171.

Macke, T. J., Ecker, D. J., Gutell, R. R., Gautheret, D., Case, D. A., and Sampath, R. (2001). RNAMotif, an RNA secondary structure definition and search algorithm. *Nucl. Acids Res.*, 29:4724–4735.

Maidak, Bonnie L., Cole, James R., Lilburn, Timothy G., Parker Jr., Charles T., Saxman, Paul R., Farris, Ryan J., Garrity, George M., Olsen, Gary J., Schmidt, Thomas M., and Tiedje, James M. (2001). The RDP-II (ribosomal database project). *Nucl. Acids Res.*, 29:173–174.

Mandl, Christian W., Holzmann, Heidemarie, Meixner, Tamara, Rauscher, Susanne, Stadler, Peter F., Allison, Steven L., and Heinz, Franz X. (1998). Spontaneous and engineered deletions in the 3'-noncoding region of tick-borne encephalitis virus: Construction of highly attenuated mutants of flavivirus. *J. Virology*, 72:2132–2140.

Mathews, D. H., Sabina, J., Zucker, M., and Turner, H. (1999). Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940.

Mattick, J. S. (1994). Introns: evolution and function. *Curr. Opin. Genet. Dev.*, 4:823–831.

Matzke, Marhori, Matzke, Antonius J. M., and Kooter, Jan M. (2001). RNA: Guiding gene silencing. *Science*, 293:1080–1083.

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

Nussinov, Ruth, Piecznik, George, Griggs, Jerrold R., and Kleitman, Daniel J. (1978). Algorithms for loop matching. *SIAM J. Appl. Math.*, 35(1):68–82.

Ohno, M and Mattaj, I W (1999). Meiosis: MeiRNA hits the spot. *Curr. Biol.*, 28:R66–R69.

Oleynikov, Yuri and Singer, Robert H. (1998). RNA localization: different zipcodes, same postman? *Trends Cell Biol.*, 8:381–383.

Omer, A. D., Lowe, T. M., Russel, A. G., Ebhardt, H., Eddy, S. R., and Dennis, P. (2000). Homologs of small nucleolar RNAs in Archaea. *Science*, 288:517–522.

Pesole, Graziano, Mignone, Flavio, Gissi, Carmela, Grillo, Giorgio, Licciulli, Flavio, and Sabino, Liuni (2001). Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 276:73–81.

Reidys, Christian, Stadler, Peter F., and Schuster, Peter (1997). Generic properties of combinatory maps. Neutral networks of RNA secondary structure. *Bull. Math. Biol.*, 59:339–397.

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

Rivas, Elena and Eddy, Sean R. (2000). Secondary structure alone is generally not statistically significant for the detection of noncoding RNAs. *Bioinformatics*, 16:583–605.

Rivas, Elena and Eddy, Sean R. (2001). Noncoding RNA gene detection using comparative sequence analysis. *BMC Bioinformatics*, 2(8):19 pages.

Rueckert, R. R. (1996). Picornaviridae: The viruses and their replication. In Fields, N.R., Knipe, D.M., and Howley, P.M., editors, *Virology*, volume 1, pages 609–654. Lippincott-Raven Publishers, Philadelphia, New York, third edition.

Samarsky, D. A. and Fournier, M. J. (1999). A comprehensive database for the small nucleolar RNAs from saccharomyces cerevisiae. *Nucleic Acids Res.*, 27:161–164.

Sankoff, D. (1985). Simultaneous solution of the RNA folding, alignment, and proto-sequence problems. *SIAM J. Appl. Math.*, 45:810–825.

SantaLucia Jr., John and Turner, Douglas H. (1997). Measuring the thermodynamics of rna secondary structure formation. *Biopolymers*, 44:309–319.

Schmitz, M. and Steger, G. (1992). Base-pair probability profiles of RNA secondary structures. *Comput. Appl. Biosci.*, 8:389–399.

Schultes, E. A. and Bartel, D. P. (2000). One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452.

Schuster, P. (1995). How to search for RNA structures. Theoretical concepts in evolutionary biotechnology. *Journal of Biotechnology*, 41:239–257.

Schuster, Peter (2001). Evolution *in silico* and *in vitro*: The RNA model. *Biol. Chem.*, 382:1301–1314.

Schuster, Peter, Fontana, Walter, Stadler, Peter F, and Hofacker, Ivo L (1994). From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284.

Schuster, Peter and Stadler, Peter F. (1998). Sequence redundancy in biopolymers: A study on RNA and protein structures. In Myers, Gerald, editor, *Viral Regulatory Structures*, volume XXVIII of *Santa Fe Institute Studies in the Sciences of Complexity*, pages 163–186. Addison-Wesley, Reading MA.

Sousa, Carolina, Johansson, Christina, Charon, Celine, Manyani, Hamid, Sautter, Christof, Kondorosi, Adam, and Crespi, Martin (2001). Translational and structural requirements of the early nodulin gene enod40, a short-open reading frame-containing RNA, for elicitation of a cell-specific growth response in the alfalfa root cortex. *Mol. Cell. Biol.*, 21:354–366.

Sprinzl, M., Horn, C., Brown, M., Ioudovitch, A., and Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucl. Acids Res.*, 26:148–153.

Stocsits, Roman, Hofacker, Ivo L., and Stadler, Peter F. (1999). Conserved secondary structures in hepatitis B virus RNA. In *Computer Science in Biology*, pages 73–79, Bielefeld, D. Univ. Bielefeld. Proceedings of the GCB'99, Hannover, D.

Stoss, Oliver, Stoilov, Peter, Daoud, Rosette, Hartmann, Annette M., Olbrich, Manuela, and Stefan, Stamm (2000). Misregulation of pre-mRNA splicing that causes human diseases. Concepts and therapeutic strategies. *Gene Therapy Mol. Biol.*, 5:9–30.

Szymanski, Maciej, Barciszewska, Miroslawa Z., Barciszewski, Jan, and Erdmann, Volker A. (2000). 5S ribosomal rna database Y2K. *Nucl. Acids Res.*, 28:166–167.

Tabaska, J. E. and Stormo, G. D. (1997). Automated alignment of RNA sequences to pseudoknotted structures. In Gaasterland, T., Karp, P., Karplus, K., Ouzounis, Ch., Sander, Ch., and Valencia, A., editors, *Proceedings of the ISMB-97*, pages 311–318, Menlo Park, CA. AAAI Press.

Thurner, Caroline, Witwer, Christine, Hofacker, Ivo, and Stadler, Peter F. (2003). Conserved RNA secondary structures in flaviviridae genomes. submitted.

Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., and De Wachter, R. (2000). The european small subunit ribosomal rna database. *Nucl. Acids Res.*, 28:175–176.

van Zon, A., Mossink, M. H., Schoester, M., Scheffer, G. L., Scheper, R. J., Sonneveld, P., and Wiemer, E. A. (2001). Multiple human vault RNAs. Expression and association with the vault complex. *J. Biol. Chem.*, 276:37715–37721.

Walter, Amy E., Turner, Douglas H., Kim, James, Lyttle, Matthew H., Müller, Peter, Mathews, David H., and Zuker, Michael (1994). Co-axial stacking of helixes enhances binding of oligoribonucleotides and improves predicions of rna folding. *Proc. Natl. Acad. Sci. USA*, 91:9218–9222.

Wang, Lilin, Smith, Dan, Bot, Simona, Dellamary, Luis, Bloom, Amy, and Bot, Adrian (2002). Noncoding RNA danger motifs bridge innate and adaptive immunity and are potent adjuvants for vaccination. *J. Clin. Invest.*, 110:1175–1184.

Waterman, M. S. (1978). Secondary structure of single - stranded nucleic acids. *Studies on foundations and combinatorics, Advances in mathematics supplementary studies, Academic Press N.Y.*, 1:167 – 212.

Westhof, Eric and Jaeger, Luc (1992). RNA pseudoknots. *Current Opinion Struct. Biol.*, 2:327–333.

Williams, Kelly P. (2002). The tmRNA website: invasion by an intron. *Nucl. Acids Res.*, 30:179–182.

Wilson, David S. and Szostak, Jack W. (1999). *In Vitro* selection of fuctional nucleic acids. *Annu. Rev. Biochem.*, 68:611–647.

Witwer, Christina, Rauscher, Susanne, Hofacker, Ivo L., and Stadler, Peter F. (2001). Conserved RNA secondary structures in picornaviridae genomes. *Nucl. Acids Res.*, 29:5079–5089.

Wuchty, S., Fontana, W., Hofacker, I. L., and Schuster, P. (1999). Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49:145–165.

Wuyts, J., De Rijk, P., Van de Peer, Y., Winkelmans, T., and De Wachter, R. (2001). The european large subunit ribosomal RNA database. *Nucl. Acids Res.*, 29:175–177.

Ying, Han, Zaks, Tal Z., Wang, Rong-Fu, Irvine, Kari R., Kammula, Udai S., Marincola, Francesco M., Leitner, Wolfgang W., and Restifo, Nicholas P. (1999). Cancer therapy using a self-replicating RNA vaccine. *Nature Med.*, 5:823–827.

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

Zuker, M. and Sankoff, D. (1984). RNA secondary structures and their prediction. *Bull. Math. Biol.*, 46:591–621.

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

Zwieb, C (1996). The uRNA database. *Nucleic Acids Res.*, 24:76–79.

Zwieb, C and Wower, J (2000). tmRDB (tmRNA database). *Nucleic Acids Res.*, 28:169–170.