

# The Effect of RNA Secondary Structures on RNA-Ligand Binding and the Modifier RNA Mechanism: A Quantitative Model

Jörg Hackermüller<sup>a,b,\*</sup>, Nicole-Claudia Meisner<sup>c,\*</sup>,  
Manfred Auer<sup>c</sup>, Markus Jaritz<sup>a</sup> and Peter F. Stadler<sup>d,b</sup>

<sup>a</sup>*Novartis Institutes for Biomedical Research Vienna,  
Informatics and Knowledge Management at NIBR, Insilico Sciences,  
Brunnerstrasse 59, A-1235 Vienna, Austria*

<sup>b</sup>*Department of Theoretical Chemistry and Structural Biology,  
University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*

<sup>c</sup>*Novartis Institutes for Biomedical Research Vienna,  
Discovery Technologies, Innovative Screening Technologies,  
Brunnerstrasse 59, A-1235 Vienna, Austria*

<sup>d</sup>*Bioinformatics Group, Department of Computer Science, and  
Interdisciplinary Center for Bioinformatics, University of Leipzig,  
Kreuzstraße 7b, D-04103 Leipzig, Germany*

\* These authors contributed equally.

---

## Abstract

RNA-ligand binding often depends crucially on the local RNA secondary structure at the binding site. We develop here a model that quantitatively predicts the effect of RNA secondary structure on effective RNA-ligand binding activities based on equilibrium thermodynamics and the explicit computations of partition functions for the RNA structures. A statistical test for the impact of a particular structural feature on the binding affinities follows directly from this approach. The formalism is extended to describing the effects of hybridizing small “modifier RNAs” to a target RNA molecule outside its ligand binding site. We illustrate the applicability of our approach by quantitatively describing the interaction of the mRNA stabilizing protein *HuR* with AU-rich elements [Meisner *et al.* (2004), *Chem. Biochem.* in press]. We discuss our model and recent experimental findings demonstrating the effectivity of modifier RNAs *in vitro* in the context of the current research activities in the field of non-coding RNAs. We speculate that modifier RNAs might also exist in nature; if so, they present an additional regulatory layer for fine-tuning gene expression that could evolve rapidly, leaving no obvious traces in the genomic DNA sequences.

*Key words:* RNA secondary structure, non-coding RNA, mRNA stability regulation, AU-rich elements, gene expression.

---

## 1 Introduction

It has become evident in the last decade that many, if not the majority, of genes are regulated post-transcriptionally [29]. A plethora of different mechanisms modify the primary transcripts on their way to the protein (or non-coding RNA) that they eventually code for, see e.g. [22, 55] for recent reviews. Most of these control mechanisms involve specific RNA-protein interactions that depend crucially on the recognition of sequence and/or structural features of the RNA [28]. Examples include the regulation of viral life-cycles [6, 9, 42, 60], pre-mRNA processing [39], nuclear RNA export [15], and the control of RNA degradation [7] and stabilization [47].

In recent years an increasing number of functional features has been reported in the untranslated regions of eukaryotic mRNA [38, 49, 57]. Well-known motifs include internal ribosomal entry sites (IRES) in viral as well as cellular mRNAs, see e.g. [61, 35, 57], the *Rho*-independent termination signal (see [16] for a detailed computational analysis), the iron responsive element (IRE) [30], and the AU-rich elements (ARE) [4, 47]. RNA secondary structure motifs are necessary in all these examples to enable the ligand to recognize (degenerate) sequence motifs. In some cases it is known that RNA secondary structure motifs without sequence constraints are recognized by regulatory proteins, see e.g. [72, 48].

Alterations of the RNA secondary structure thus open an interesting perspective for biotechnology. The effects of mRNA secondary structure modification on bacterial translation, for instance, are studied in [56]. Inhibition of ribozymes by means of oligonucleotide directed RNA misfolding has been demonstrated e.g. for group I introns [11] and RNase P [12]. Oligomeric nucleic acid analogs were recently used to specifically inhibit IRES-dependent translation in hepatitis C virus [54], presumably by interfering with the IRES structure.

Complementary to such regulatory protein-RNA interactions there is a rapidly increasing number of different classes of non-coding RNAs that actively take part in mRNA processing and expression regulation [36, 20, 31]: RNaseP, MRP RNA, spliceosomal RNA, signal recognition particle RNAs, and microRNAs all perform their function as part of RNA-protein complexes. In addition, there is a number of RNA-protein complexes, such as vaults [71] and Ro particles [70] that have been known for decades, but whose function has remained enigmatic so far.

Despite the fact that the experimental findings summarized above (and many others not cited here) clearly indicate a pivotal role of RNA structure — and hence of the thermodynamics of RNA folding — in RNA ligand interactions,

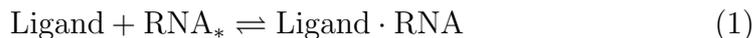
this topic has not yet been investigated systematically. In this contribution we derive a quantitative model for the effect of RNA secondary structure on RNA-ligand binding. We then use this theory to devise a statistical test for the involvement of specific RNA secondary structure features in RNA-ligand binding. In the subsequent section the theoretical framework is expanded to modeling the modification of the RNA secondary structure by means of hybridization of small oligonucleotides outside the ligand binding motif. A recent study shows that such a mechanism is feasible *in vitro*: Isaacs *et al.* [37] demonstrate that translation of mRNAs that are not translatable because their ribosome binding site is inaccessible due to stable secondary structures can be activated by means of small artificial “transactivating RNAs”.

Finally, we propose that modifier RNAs might also occur naturally and briefly discuss the impact of such an additional regulatory mechanism on our understanding of gene regulation and its evolution.

We will illustrate the biophysical model throughout this contribution by comparing computational predictions with recent experimental results for the *HuR*/mRNA system [47]. *HuR* (or *ELAVL1*) is a key regulator at the post transcriptional regulatory level of eukaryotic gene expression. It is the only, ubiquitously expressed, positive regulator of AU-rich element (ARE) dependent mRNA stability control which is currently known. *HuR* regulates functionally diverse genes in a highly specific manner and could be involved in the regulation of up to 3000 genes [8]. The *HuR*-mediated activation crucially depends on the structural context of a degenerate binding sequence motif [47].

## 2 Quantitative Model of RNA-Ligand Binding

We consider here a (protein) ligand that binds to an RNA molecule in a simple two-state process with 1:1 stoichiometry. Multi-state processes involving a conformational rearrangement after ligand binding (induced fit, e.g [73]) are also described by this model provided that the free energy changes due to the structural rearrangement after binding are (nearly) independent of the RNA sequence. Furthermore we assume that only those RNA molecules can be bound that present the binding site(s) in a particular spatial conformation. We use the symbol  $\text{RNA}_*$  to denote this sub-population of RNA molecules.



The law of mass action implies that the concentrations  $[\text{RNA}_*]$ ,  $[\text{Ligand}]$ , and  $[\text{Ligand} \cdot \text{RNA}]$  of free accessible RNA, free protein, and complex are related

through the dissociation constant

$$K_d = \frac{[\text{RNA}_*] [\text{Ligand}]}{[\text{Ligand} \cdot \text{RNA}]} \quad (2)$$

An RNA molecule with nucleotide sequence  $s$  may form many different structures. For our purposes it is sufficient to distinguish between secondary structures only. The set  $\Sigma(s)$  consists of all secondary structures (i.e., lists of base pairs)  $\Psi$  satisfying the following conditions: (i) Each nucleotide  $s_i$  takes part in at most one base pair; (ii) base pairs do not cross, i.e.,  $(s_i, s_j) \in \Psi$  and  $(s_k, s_l) \in \Psi$  with  $i < j$ ,  $k < l$ , and  $i < k$  implies either  $j < l$  and  $j > l$ ; and (iii) each pair  $(s_i, s_j) \in \Psi$  is one of the six canonical pairs GC, CG, AU, UA, GU, or UG. For each secondary structure  $\Psi$  of  $s$  one can compute a free energy  $F(\Psi)$  by adding up energy contributions for stacked base pairs, hairpin loops, interior loops, bulges, and multi-branched loops. These energy contributions have been determined experimentally, see [44]. The frequency of a particular secondary structure  $\Psi$  in thermodynamic equilibrium ensemble can therefore be computed as

$$p(\Psi) = \frac{1}{Z} \exp\left(-\frac{F(\Psi)}{RT}\right) \quad (3)$$

where  $Z = \sum_{\Upsilon \in \Sigma(s)} \exp(-F(\Upsilon)/RT)$  is the partition function of the RNA molecule  $s$ .

Writing  $A(s) \subseteq \Sigma(s)$  for the accessible structures of our RNA molecule  $s$  we obtain

$$[\text{RNA}_*] = p_* [\text{RNA}] \quad (4)$$

where  $p_*$  is the fraction of accessible secondary structures:

$$p_* = \sum_{\Psi \in A(s)} p(\Psi) = \frac{1}{Z} \sum_{\Psi \in A(s)} \exp\left(-\frac{F(\Psi)}{RT}\right) = \frac{Z_*}{Z}. \quad (5)$$

We remark that we can of course describe the concentration of accessible RNA in terms of the law of mass action: The equilibrium constant for the refolding



between accessible and inaccessible conformations is given by

$$K_* = \frac{[\text{RNA}_*]}{[\text{RNA}_-]} = \frac{p_*}{1 - p_*} \quad (7)$$

The problem thus reduces to computing the partition functions for the two sets of secondary structures  $\Sigma(s)$  and  $A(s)$ . This can be achieved e.g., by means of dynamic programming [46] as we shall see below. Substituting equ.(4) into equ.(2) yields

$$\frac{[\text{RNA}] [\text{Ligand}]}{[\text{Ligand} \cdot \text{RNA}]} = \frac{K_d}{p_*} =: K_d^{\text{app}} \quad (8)$$

Using conventional methods to measure RNA protein interactions, only the total concentration of unbound RNA,  $[\text{RNA}]$ , can be measured. Hence, only the apparent dissociation constant  $K_d^{\text{app}} = K_d/p_*$  can be determined experimentally. As a consequence, we predict a structure dependence of the measured values of  $K_d^{\text{app}}$ . Under the assumption that the true value of  $K_d$  depends only on the ligand and the sequence-structure motif that binds the ligand, we can predict sequence-dependent variations in RNA-ligand binding affinity by means of a computational analysis of the ensemble of RNA structures.

In the simplest case the sequences under consideration contain a single copy of the binding motif which must be present in a particular secondary structure conformation. Usually, the structural requirements will only be a few local base pairs at the binding site, or, conversely, it might be necessary that all or a part of the binding site remains unpaired. The number of accessible structures will therefore in general be too large to use equ.(5) directly. Instead, a modification of McCaskill’s partition function algorithm [46] can be used to compute partition functions restricted to structures that contain a specified list of base pairs and/or a specified list of unpaired positions. We refer to [33] for a description of the algorithms, which are implemented as part of the **Vienna RNA Package** [34, 32]. The up-to-date collection of energy parameters contains both enthalpies and entropies and thus can be used to compute the ensemble of equilibrium secondary structures for a given temperature[44].

The situation becomes more complicated if the sequence motif is very degenerate and hence a single RNA sequence  $s$  can have more than one potential binding site. We will restrict ourselves here to the case in which RNA-ligand complexes are always of 1:1 stoichiometry even if the protein has multiple binding sites. In this case  $A(s)$  consists of all those secondary structures in which at least one binding site  $B_i$ ,  $i = 1, \dots, M$  is accessible. The **RNAfold** program from the **Vienna RNA Package** can be used to compute the partition function  $Z(\mathcal{X})$  over all secondary structures that satisfy a given structural constraint  $\mathcal{X}$ . The fraction of structures that satisfy  $\mathcal{X}$  in thermodynamic equilibrium is thus  $p(\mathcal{X}) = Z(\mathcal{X})/Z$ . In particular, we can compute the probability  $p(\mathcal{A})$  that a subset  $\mathcal{A} \subseteq \{B_1, B_2, \dots, B_M\}$  of binding sites has correct conformation, *irrespective* of the conformations at all other binding sites. In this notation  $\mathcal{A} = \emptyset$  means that there is no constraint on the structure (and hence  $p(\emptyset) = 1$ ), while  $\mathcal{A} = \{B_1, B_2, \dots, B_M\}$  means that all  $M$  binding sites are accessible si-

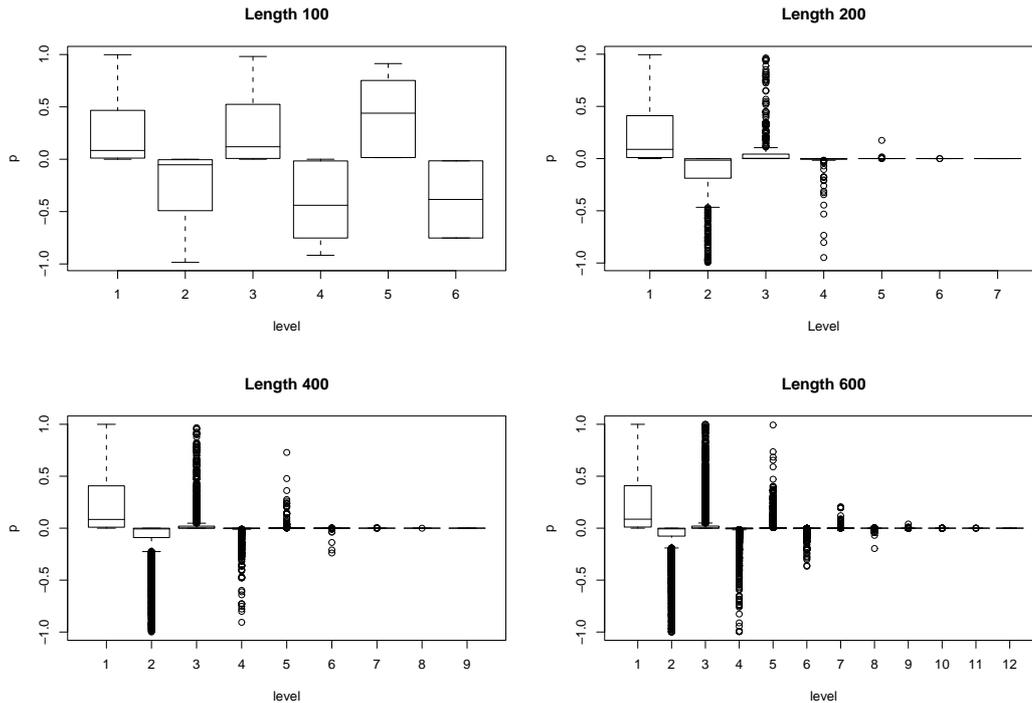


Fig. 1. Convergence of Equ.(9) for the motif NNUUNUUU in single-stranded conformation in random target sequences as a function of sequence length  $n$ . We plot total effect of the  $\ell$ -th order terms in equ.(9) as a function of  $\ell$  in sample of 1000 sequences. Boxes give the range from 1st to 3rd quartile, with median indicated by a line; whiskers indicate the position of the most extreme data point within 1.5 times the interquartile distance, outliers are shown as circles.

multaneously. The probability that all binding sites are inaccessible can be computed immediately using the inclusion-exclusion principle:

$$1 - p_* = \sum_{\ell=0}^M (-1)^\ell \sum_{\substack{\mathcal{A} \\ |\mathcal{A}|=\ell}} p(\mathcal{A}) \quad (9)$$

For large numbers  $M$  of potential binding sites this becomes infeasible since equ.(9) requires  $2^M$  evaluations of a partition function (one for the unconstrained molecule and  $2^M - 1$  for the different combinations of binding sites). As an approximation the expansion can be truncated at order  $\ell_{\max} < M$ . The expansion in general converges quickly for long sequences, while for short sequences we need more or less all the terms, see Fig. 1. In practice, however, one will usually encounter binding motifs that are relatively rare since a ligand can fulfill its regulatory role only if it does not indiscriminately bind everywhere. We remark that  $p_*$  could alternatively be evaluated by using stochastic backtracking to obtain a Boltzmann-weighted sample of secondary structures instead of computing constrained partition functions [69, 18, 19, 33]. The sampling approach is computationally more efficient, but it is less accurate for

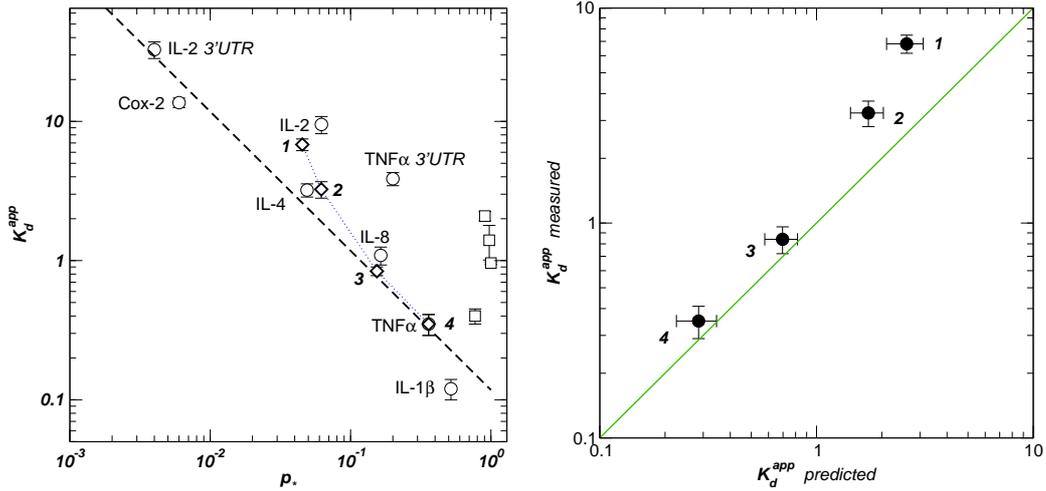


Fig. 2. (a) Apparent dissociation constants for *HuR*-mRNA complexes at  $23.5^\circ\text{C}$  for natural ARE and UTR sequences ( $\circ$ ), four artificial molecules ( $\square$ ), see Table 1 for details, and four designed mutants of the TNF $\alpha$  ARE ( $\diamond$ ). The dashed line marks  $K_d^{\text{app}} = K_d/p_*$  with  $K_d = 0.118$ . The value of  $K_d$  is obtained by non-linear regression with a correlation coefficient of 0.946 and  $\chi^2 = 122.6$ . The artificial repetitive sequences might be more regularly structured than expected from the secondary structure calculation. (b) Comparison of predicted and measured values of  $K_d^{\text{app}}$  for 4 mutants of TNF $\alpha$ . Calculations have been performed for  $T = 23.5^\circ\text{C}$  at which the measurements were performed in ref. [47].

small probabilities  $p_*$  of the accessible structures.

We next demonstrate the validity of our model in the *HuR*/mRNA system. The recognition of mRNAs by the protein *HuR* is dependent on the presentation of the degenerate sequence motif NNUUNUUU in a single stranded conformation [47]. The presentation of the *HuR* binding site in the mRNA secondary structure appears to act analogously to a regulatory on/off switch which specifically controls *HuR* access to mRNAs *in cis* [47]. Apparent dissociation constants  $K_d^{\text{app}}$  were measured for a number of sequences from ARE-containing mRNAs, Tab. 1.

Figure 2 displays the data from Tab. 1 on a double-logarithmic scale. The dashed line is a regression of equ.(8) to the data with  $K_d$  as the only fitting parameter.

### 3 A Statistical Test for the Influence of Secondary Structure

The theory outlined above predicts a dependence of the measured apparent  $K_d^{\text{app}}$  on  $p_*$  if the RNA-ligand binding depends on particular secondary struc-



ture features of the binding site. This relationship can be turned into a statistical test for the influence of secondary structure given a set of binding data of RNA sequences that contain a known *sequence* motif required for ligand recognition. Given a hypothesis  $\Xi$  for a possible structural feature one can compute the probability  $p_*[\Xi]$  that at least one binding sequence motif in the RNA sequence  $s$  satisfies the secondary structure constraints  $\Xi$  as outlined above. The influence of  $\Xi$  on binding manifests itself in a (linear) correlation between  $K_d^{\text{app}}$  and  $1/p_*[\Xi]$ .

Following [14] the null hypothesis of *no correlation* can be rejected if

$$\frac{(k-2)r^2}{1-r^2} \geq t_{(k-2)}(1-2/\alpha)^2 \quad (10)$$

is satisfied. Here  $k$  is the number of sequences,  $t_{(k-2)}(y)$  is Student's  $t$ -distribution with  $k-2$  degrees of freedom,  $\alpha$  is the desired significance level and  $r$  is the empirical correlation coefficient between  $K_d^{\text{app}}$  and  $1/p_*[\Xi]$ .

In general there is a large number of different secondary structure elements  $\Xi$  that can be realized simultaneously by a set of related sequences [1]. Thus it may not be feasible to find the optimal structure constraint *ab initio*. The test procedure above, however, allows to select or exclude a secondary structure element from a set of candidate elements.

Applying this test to the data in Tab. 1 yields a correlation coefficient of  $r \approx 0.901$  and the inequality (10) with  $k = 12$  is satisfied ( $43.34 > 16.12$ ). Thus  $K_d^{\text{app}}$  and  $1/p_*[\Xi]$  are significantly correlated with more than 99% confidence.

Once the importance of the secondary structure  $\Xi$  has been verified by the above test, one can use a simple least-squares fit to determine  $K_d$  from the  $(K_d^{\text{app}}, 1/p_*[\Xi])$  pairs. This results in a model for predicting apparent dissociation constants for any RNA molecule that contains the binding sequence motif. This model can be used to design RNA sequences with a predefined binding affinity. In [47] the feasibility of this approach is demonstrated by the design of four mutants of the TNF $\alpha$  UTR with predefined values of  $p_*$ , see Figure 2b.

## 4 Modifier RNAs

The thermodynamics of an RNA molecule  $M$  changes when it hybridizes with a short oligonucleotide  $O$ . Since the nucleotides of  $M$  that bind the oligonucleotide  $O$  are no longer available for pairing in the intra-molecular secondary structure, the molecule  $M$  will typically refold. This can have drastic effects

on the secondary structure of a binding sequence motif even if the oligonucleotide  $O$  binds far away from the binding site. Depending on the sequence of the oligonucleotide, the effect can be either an increase or a decrease in the fraction  $p_*$  of accessible secondary structures.

The thermodynamics of RNA-RNA hybridization is well understood [17]. At the time of writing this study, no implementation was available that considers all possible structures within each strand of two hybridized RNA molecules so that we use an approximate model here. An extension of the **Vienna RNA Package** that implements the complete folding model for two interacting RNAs is forthcoming [25]. We thus briefly describe the complete theory here and derive an approximation that can probably be used in most cases of practical interest, including the application to the *HuR*/ARE model system.

The mRNA molecule  $M$  and the oligonucleotide  $O$  together can form five molecular species<sup>1</sup>: the monomers  $M$  and  $O$ , the homodimers  $MM$  and  $OO$  and well as the heterodimer  $MO$  that we are primarily interested in. In thermodynamic equilibrium we have

$$[MM] = K_{MM}[M]^2 \quad [OO] = K_{OO}[O]^2 \quad [MO] = K_{MO}[M][O] \quad (11)$$

with equilibrium constants  $K_{MM}$ ,  $K_{OO}$ , and  $K_{MO}$  that can be computed from partition functions by means of an extension of McCaskill's algorithm, see [17, 25]. For each of the monomer and dimer species one can then compute, by the same approach as in the previous section, the probabilities  $p_*(M)$ ,  $p_*(MM)$ , and  $p_*(MO)$  that the binding motifs(s) are accessible in the three molecular species. We can therefore compute the effective fraction  $p_*$  of mRNAs with accessible binding sites as

$$p_* = p_*(M) \frac{[M]}{[M]_0} + p_*(MM) \frac{[MM]}{[M]_0} + p_*(MO) \frac{[MO]}{[M]_0}, \quad (12)$$

where  $[M]_0 = [M] + 2[MM] + [MO]$  is the total concentration of mRNA that is not bound to the ligand. The concentration  $[M]_0$  is determined by the value of  $K_d$ , the three equilibrium constants  $K_{MM}$ ,  $K_{MO}$ , and  $K_{OO}$ , and the initial concentrations of the mRNA and the oligonucleotide.

Let us now make the following simplifying assumptions:

- (i) The oligonucleotide  $O$  is (nearly) complementary to a unique target site on the mRNA  $M$ . This assumption is inspired by the small interfering RNAs

---

<sup>1</sup> We neglect here multiple binding, i.e., species such as  $MO_2$ . These could be taken into account without conceptual difficulties at the expense of a more complicated set of equations.

- [21] and their relatives, see e.g. [24] and the references therein.
- (ii) Both the oligonucleotide  $O$  and the mRNA  $M$  are not significantly self-complementary.
  - (iii) The oligonucleotide  $O$  is present in excess.

Under these hypotheses we have  $K_{MM}, K_{OO} \ll K_{MO}$ , and  $[M]_0 \ll [O]_0$ , i.e., almost all mRNAs are hybridized with the oligonucleotide  $O$ . This allows us to use the approximation

$$p_* \approx p_*(MO) \frac{[MO]}{[M]_0} \approx p_*(MO). \quad (13)$$

The set of possible secondary structures of the  $MO$  duplex can be approximated by those structures of the mRNA  $M$  in which the target site  $T$  of the oligonucleotide cannot pair with other nucleotides of  $M$ . The energy of such a secondary structure is  $F(\Psi_{M \setminus O}) + F(TO)$  where  $F(\Psi_{M \setminus O})$  is the energy of the secondary structure  $\Psi_{M \setminus O}$  in which the target site for the oligonucleotide is unpaired and  $F(TO)$  is the energy contribution for the hybridization of the oligonucleotide to its target site on  $M$ . While  $F(TO)$  can in principle be computed, we can simply treat it as a constant independent of  $\Psi_{M \setminus O}$  which therefore cancels in the partition function computations. Thus we obtain

$$p_*^{MO}(\mathcal{A}) = Z(\mathcal{A} \cup \mathcal{T}) / Z(\mathcal{T}) \quad (14)$$

directly from the constrained partition functions  $Z(\cdot)$  of the mRNA  $M$  using the additional constraint  $\mathcal{T}$  that the target site  $T$  is unpaired. If a binding site  $B_i \in \mathcal{A}$  and  $T$  overlap, then  $T$  takes precedence, i.e., we assume that  $B_i$  cannot be accessible when the oligonucleotide is bound at this position. We can now use equ.(9) to derive  $p_*$  in the same way as for the mRNA alone.

Equ.(14) describes the effect of a particular oligonucleotide  $O$ . Since  $O$  modifies the RNA-ligand binding we refer to  $O$  as a *modifier RNA*. It has been demonstrated experimentally in the *HuR*-mRNA system that modifier RNAs are functional [47]. Modifier RNAs can be designed by means of the following, generally applicable procedure: We fix a length  $N_0$  of the modifier oligonucleotide  $O$ , say  $N_0 = 20$  inspired by siRNAs and microRNAs, and compute the effect of the oligonucleotide when it binds the mRNA  $M$  starting from sequence position  $k$ . Figure 1 shows an example of a profile  $p_*^{MO}[k]$ . Such modifier-effect profiles can be computed for moderate size mRNAs (e.g. TNF $\alpha$ ) within about a day from equ.(14) using 30 Xeon CPUs. A sampling approach based on stochastic backtracking will be much more efficient provided one is only interested in oligonucleotides leading to large values of  $p_*^{MO}$ .

The modifier-effect profiles allow the specific design of RNA oligonucleotides that modulate the ligand binding affinity by *opening* ( $p_*^{MO} \rightarrow 1$ ) or *closing*

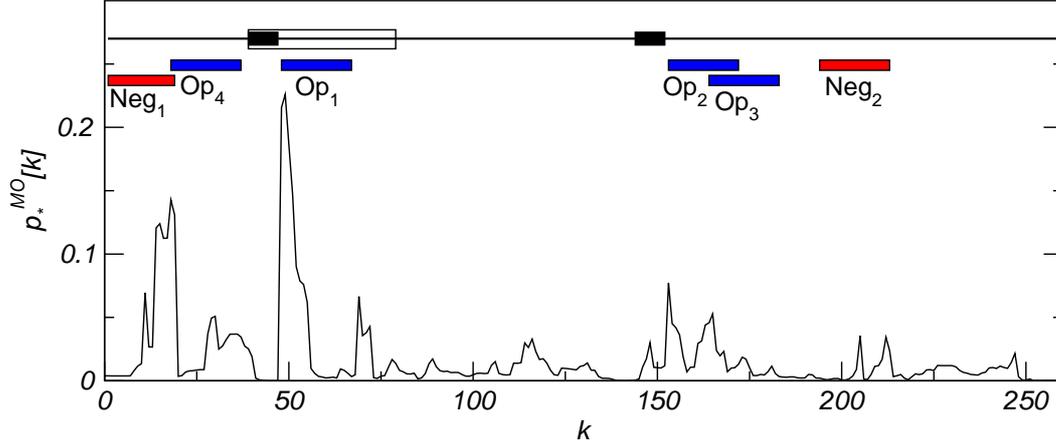


Fig. 3. Modifier profile for the 3'UTR sequence of human *IL2* mRNA for oligonucleotides of length  $N_0 = 20$ . The binding motif for *HuR* is the sequence NNUUNUUU in an open conformation,  $\Xi = ' \dots \dots \dots '$ . The position  $k$  is the start position of the modifier bound region in the target RNA,  $p_*^{MO}[k]$  is the probability of *HuR* accessible sequences in the secondary structure ensemble if a modifier is hybridized to position  $k$ . The ARE is marked as an open box, *HuR* binding motifs are indicated by black filled boxes. Modifiers of significant impact on  $p_*^{MO}$  are restricted to few positions mainly in proximity of the *HuR* binding sites. At several positions, hybridization of an oligonucleotide does not influence the accessibility of *HuR* motifs, which allows to design negative controls. Four openers ( $Op_1, Op_2, Op_3, Op_4$ ) and two negative controls ( $Neg_1, Neg_2$ ), which were studied in more detail in [47], are indicated by blue and red boxes, respectively.

Table 2  
Modifier oligonucleotides for *IL2* mRNA (NM\_000586) from [47].

Name	Position	Sequence	$K_d^{app}$
IL-2	-	no opener	32.77 $\pm$ 4.48
$Op_1$	804-823	AATATAAAATTTAAATATTT	11.80 $\pm$ 1.48
$Op_2$	909-928	TAGAGCCCCTAGGGCTTACA	18.91 $\pm$ 1.91
$Op_3$	920-939	TGAAACCATTTTAGAGCCCC	19.52 $\pm$ 2.20
$Op_4$	774-793	AAGGCCTGATATGTTTTAAG	8.38 $\pm$ 1.18
$Neg_1$	757-775	AGTGGGAAGCACTTAATTAC	32.91 $\pm$ 6.34
$Neg_2$	950-969	CATAATAATAAATATTTTGG	32.77 $\pm$ 3.72

( $p_*^{MO} \rightarrow 0$ ) the binding sites to the ligand, Tab. 1. In the *HuR*-mRNA system one can in fact use small *opener-RNAs* to enhance the expression of the gene by *HuR* dependent stabilization of mRNAs in cell extracts [47].

If we know the binding constant  $K_{MO}$  of the modifier  $O$  to the mRNA  $M$  we

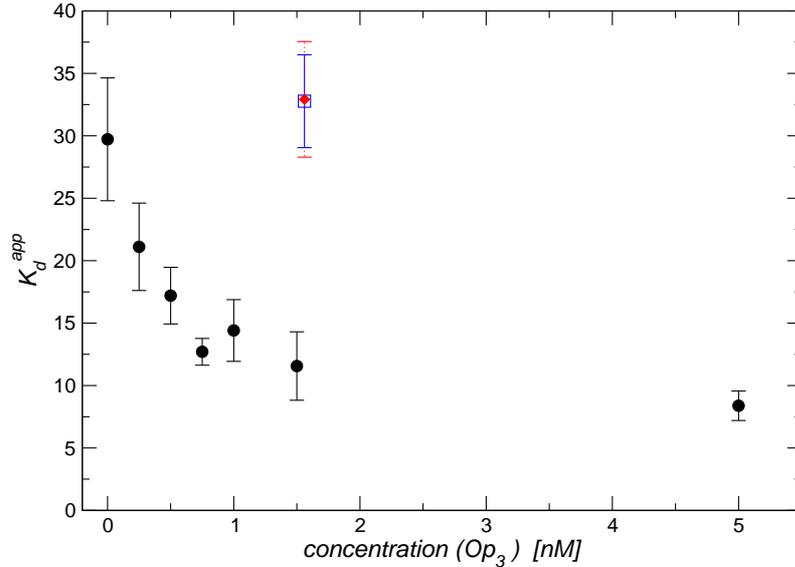


Fig. 4. Effect of a complementary opener of length  $N_0 = 20$  on *in vitro* HuR/RNA affinities. The apparent affinity of recombinant HuR to IL-2 3'UTR was determined in presence and absence of the opener  $\text{Op}_3$  (black circles) and of the negative controls  $\text{Neg}_1$  (red diamond) and  $\text{Neg}_2$  (open blue box) with 1D-FIDA detection [47]. Opener  $\text{Op}_3$  has a concentration dependent effect on the apparent affinity that is consistent with equ.(16).

can calculate the dependence of the apparent dissociation constant

$$K_d^{\text{app}} := \frac{[\text{RNA}] [\text{Ligand}]}{[\text{RNA} \cdot \text{Ligand}]} = \frac{[M] [\text{Ligand}] + [MO] [\text{Ligand}]}{[M \cdot \text{Ligand}] + [MO \cdot \text{Ligand}]} \quad (15)$$

using equ. (8) to substitute equ.(15) for both  $M$  and  $MO$  with their respective fractions  $p_*^M$  and  $p_*^{MO}$ , resp., of accessible structures. We obtain

$$K_d^{\text{app}} = K_d \frac{1 + K_{MO}[O]}{p_*^M + p_*^{MO} K_{MO}[O]} \quad (16)$$

which describes a hyperbolic transition from  $K_d/p_*^M$  to  $K_d/p_*^{MO}$  with increasing concentration  $[O]$  of the modifier oligonucleotide. This behavior is indeed observed for some opener molecules, Figure 4. For other openers, such as  $\text{Op}_1$  from Table 1, we find an increase of  $K_d^{\text{app}}$  for very large opener concentrations which might e.g. be due to binding at multiple sites.

The computation of  $K_{MO}$  requires again a partition function calculation which could in principle be performed using the approach described in [17], the `RNAhybrid` approach [59], or `RNAcofold` [25]. A quantitative comparison of these approaches among each other and in correlation with experimental data

for the *HuR*/ARE system will be reported elsewhere.

## 5 Discussion

We have presented here a model for RNA-Ligand binding that takes into account that interactions between RNAs and their ligands are often dependent on the RNA structure. Since the interaction is local in many cases it can be well described in terms of secondary structures; in this case we can use efficient dynamic programming algorithms to make quantitative predictions. Recently published data for the interaction of ARE-elements of messenger RNAs with the protein *HuR* are used as an example. The framework described in this contribution is much more general, however: it can handle almost arbitrary combinations of sequence and structure requirements that together form a binding motif.

From our model we predict that small RNA “modifiers” can modulate the binding affinities of RNAs and ligands by binding to their target RNA in such a way that it alters the local structure at the ligand binding site. The *HuR*-mRNA interaction again serves as a well-studied *in vitro* example. The regulation of *HuR*-ARE-mediated export and RNA stability *in vivo*, however, remains enigmatic. There is only the single ubiquitously expressed protein *HuR* (and a handful of tissue specific relatives such as the neuronal specific homologue *HuD*) that up-regulates the export and stability of potentially thousands of ARE-carrying mRNAs. It is tempting to speculate that modifying RNA “openers” could be involved in target gene specific regulation of *HuR* activity.

If this is the case, there is no reason to assume that the action of small modifier RNAs should be restricted to the *HuR*/ARE system. This scenario extends David Bartel and Chang-Zheng Chen’s proposal of microRNAs as “micro-managers of gene expression” [5] and follows John Mattick’s argument for a dominating layer of RNA-mediated regulation [45].

There is indeed mounting evidence for a vast variety of regulatory active small RNAs [45, 68]: Some organisms, such as *Leishmania* and related kinetoplastids, have reduced transcriptional regulation of gene expression to a minimum, maybe to the point of having lost any specific polymerase II transcription initiation [13]. Instead, *Leishmania* uses an elaborate cleavage and trans-splicing mechanism based on the action of  $\sim 40$ nt “spliced leader” RNA. *Tetrahymena* appears to use an RNA-based mechanism for directing its genomewide DNA rearrangements [50, 74]. The *E. coli* genome encodes more than 50 small RNA genes at least some of which (e.g. *MicF*, *OxyS*, *DsrA*, *Spot42*, *RhyB*) act by basepairing to activate or repress translation [66]. A large fraction of

the mouse transcriptome consists of non-coding RNAs, many of them anti-sense to known protein-coding transcripts [67]. Similarly, about half of the transcripts from Human chromosomes 21 and 22 are non-coding [10, 40]. The possible roles of anti-sense RNAs are discussed in [52]. Ambros and coworkers [2] reported more than 30 *tiny non-coding RNAs* in a recent survey of *C. elegans*. These “tncRNAs” are slightly shorter than microRNAs, are not processed from hairpin precursors, and are poorly conserved between related species.

*Riboswitches*, i.e., RNAs that drastically change their structure, are important regulatory elements. For instance, the terminator and anti-terminator, two alternative RNA hairpins, regulate gene expression in *E. coli* and *B. subtilis* by attenuation [3, 23, 58]. Riboswitches can provide exact temporal control as in the *hok/sok* system of plasmid R1 which triggers programmed cell death [53, 51]. Riboswitches also play a role in the spliced leader of trypanosomes and nematodes [43]. Artificial RNA switches have been designed as well, see e.g. [65]. For instance, in [64] an RNA is described whose conformation change is triggered by ligand binding using a switching mechanism similar to the one proposed for the ribosomal A site. An RNA controlled allosteric hammerhead ribozyme is presented in [41]. An RNA molecule that has two different ribozyme functions depending on its spatial conformation is described in [62]. A theoretical study shows that potential riboswitches, i.e., RNAs that have very different secondary structures with near-groundstate energy, are relatively frequent and easily accessible in evolution [26].

Riboswitches might be just the extreme cases of a regulatory mechanism that works more generally by modifying the relative concentrations of different RNA structures (or structural classes). The modifier RNA mechanism outlined in this presentation would provide a general and gene specific way to both up- and down-regulate RNA-ligand binding affinities and thus allow a fast and specific fine-tuning of the eventual expression level of a gene product. The mechanism is independent of an elaborate machinery of RNP complexes since the modifiers exert their function by directly binding to their target RNA. This reduces evolutionary constraints on the hypothetical modifier RNAs. Furthermore, mutations in modifier RNAs will often have small quantitative rather than qualitative effects on expression levels because the effect of point mutations on RNA helices is limited to a few kcal/mol. On the other hand, some mutations can lead to drastic changes in the preferred structures in the same way as for isolated RNA molecules [27]. The hypothetical modifier RNAs would therefore not be subject to strong multiple constraints, so that they would rapidly drift along neutral networks in sequence space as described in Ref. [63]. In particular, if we assume that the major source of the hypothetical modifier RNAs are antisense transcripts, they evolve without the need for compensatory mutations to maintain complementarity between the modifier and its target. In another scenario, trans-acting modifiers might avoid

exact complementarity to their target in order to avoid triggering the RNAi pathways; in this case their binding patterns are essentially unconstrained so that compensatory mutations are also not necessary. It is thus entirely plausible that a regulatory level based on modifier RNAs evolves very fast and does not leave phylogenetic footprints or other easy-to-find signals in the genomic DNA.

### *Acknowledgements*

The authors are grateful to Ivo Hofacker, Christoph Flamm and Kurt Gr unberger for stimulating discussions and to Manuel Peitsch, Ren e Amstutz and Jan deVries for continuous support. This work was supported in part by the Austrian *Fonds zur F orderung der Wissenschaftlichen Forschung*, Project No. P15893, the German *DFG* Bioinformatics Initiative BIZ-6/1-2 and the Austrian Academy of Sciences (DOC 11/2000).

### **References**

- [1] I. Abfalter, C. Flamm, and P. F. Stadler. Design of multi-stable nucleic acid sequences. In H.-W. Mewes, V. Heun, D. Frishman, and S. Kramer, editors, *Proceedings of the German Conference on Bioinformatics. GCB 2003*, volume 1, pages 1–7, M unchen, D, 2003. belleville Verlag Michael Farin.
- [2] V. Ambros, R. C. Lee, A. Lavanway, P. T. Williams, and D. Jewell. MicroRNAs and other tiny endogenous RNAs in *C. elegans*. *Current Biology*, 13:807–818, 2003.
- [3] P. Babitzke and C. Yanofsky. Reconstitution of *Bacillus subtilis* Trp attenuation in vitro with TRAP, the Trp RNA-binding attenuation protein. *Proc. Natl. Acad. Sci. USA*, 90:133–137, 1993.
- [4] T. Bakheet, M. Frevel, B. R. Williams, W. Greer, and K. S. Khabar. ARED: human AU-rich element-containing mRNA database reveals an unexpectedly diverse functional repertoire of encoded proteins. *Nucl. Acids Res.*, 29:246–254, 2001.
- [5] D. P. Bartel and C.-Z. Chen. Micromanagers of gene expression: the potentially wide-spread influence of metazoan micrnas. *Nature Genetics*, 5:396–400, 2004.
- [6] G. J. Belsham and N. Sonenberg. RNA-protein interactions in regulation of picornavirus RNA translation. *Microbiol. Rev.*, 60:499–511, 1996.
- [7] A. Bevilacqua, M. C. Ceriani, S. Capaccioli, and A. Nicolin. Post-transcriptional regulation of gene expression by degradation of messenger RNAs. *J. Cell. Physiol.*, 195:356–372, 2003.

- [8] C. M. Brennan and J. A. Steitz. HuR and mRNA stability. *Cell Mol. Life Sci.*, 58:266–277, 2001.
- [9] A. Carlsson and S. Schwartz. Inhibitory activity of the human papillomavirus type 1 AU-rich element correlates inversely with the levels of the *ELAV*-like *HuR* protein in the cell cytoplasm. *Arch. Virol.*, 145:491–503, 2000.
- [10] S. Cawley, S. Bekiranov, H. H. Ng, P. Kapranov, E. A. Sekinger, D. Kampa, A. Piccolboni, V. Sementchenko, J. Cheng, A. J. Williams, R. Wheeler, B. Wong, J. Drenkow, M. Yamanaka, S. Patel, S. Brubaker, H. Tammana, G. Helt, K. Struhl, and T. R. Gingeras. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell*, 116:499–509, 2004.
- [11] J. L. Childs, M. D. Disney, and D. H. Turner. Oligonucleotide directed misfolding of RNA inhibits *Candida albicans* group I intron splicing. *Proc. Natl. Acad. Sci. USA*, 99:11091–11096, 2002.
- [12] J. L. Childs, A. W. Poole, and D. H. Turner. Inhibition of *Escherichia coli* RNase P by oligonucleotide directed misfolding of RNA. *RNA*, 9:1437–1445, 2003.
- [13] C. E. Clayton. Life without transcriptional control? From fly to man and back again. *EMBO J.*, 21:1881–1888, 2002.
- [14] W. Cooper, I. Dunsmore, J. Gower, and *et al.* Inferences about the correlation coefficient. In W. Ledermann, editor, *Handbook of applicable mathematics: Statistics*, page 721. Wiley Interscience, London, 1984.
- [15] B. R. Cullen. Nuclear RNA export. *J. Cell. Sci.*, 116:587–597, 2003.
- [16] Y. d’Aubenton Carafa, E. Brody, and C. Thermes. Prediction of rho-independent escherichia coli transcription terminators. a statistical analysis of their RNA stem-loop structures. *J. Mol. Biol.*, 216:835–858, 1990.
- [17] R. A. Dimitrov and M. Zuker. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys. J.*, 87:215–226, 2004.
- [18] Y. Ding and C. E. Lawrence. Statistical prediction of singlestranded regions in RNA secondary structure and application to predicting effective antisense target sites and beyond. *Nucl. Acids Res.*, 29:1034–1046, 2001.
- [19] Y. Ding and C. E. Lawrence. A statistical sampling algorithm for RNA secondary structure prediction. *Nucl. Acids Res.*, 31:7180–7301, 2003.
- [20] S. Eddy. Non-coding RNA genes and the modern RNA world. *Nature Genetics*, 2:919–929, 2001.
- [21] W. Elbashir, S. Lendeckel and T. Tuschl. RNA interference is mediated by 21- and 22-nucleotide RNAs. *Genes Dev.*, 15:188–200, 2001.
- [22] D. Elliot. Pathways of post-transcriptional gene regulation in mammalian germ cell development. *Cytogenet. Genome Res.*, 103:210–216, 2003.
- [23] G. Fayat, F. J. Mayaux, C. Sacerdot, M. Fromant, M. Springer, M. Grunberg-Manago, and S. Blanquet. *Escherichia coli* phenylalanyl-tRNA synthetase operon region. Evidence for an attenuation mechanism. Identification of the gene for the ribosomal protein L20. *J. Mol. Biol.*,

- 171:239–261, 1983.
- [24] E. J. Finnegan and M. A. Matzke. The small RNA world. *J. Cell Sci.*, 116:4689–4693, 2003.
- [25] C. Flamm, S. Bernhart, I. L. Hofacker, and P. F. Stadler. A full model for RNA-RNA cofolding. in preparation.
- [26] C. Flamm, I. L. Hofacker, S. Maurer-Stroh, P. F. Stadler, and M. Zehl. Design of multi-stable RNA molecules. *RNA*, 7:254–265, 2000.
- [27] W. Fontana, P. F. Stadler, E. G. Bornberg-Bauer, T. Griesmacher, I. L. Hofacker, M. Tacker, P. Tarazona, E. D. Weinberger, and P. Schuster. RNA folding landscapes and combinatorial landscapes. *Phys. Rev. E*, 47:2083–2099, 1993.
- [28] K. B. Hall. RNA-protein interactions. *Curr. Opin. Struct. Biol.*, 12:283–288, 2002.
- [29] S. R. Haynes. Posttranscriptional regulation and RNA binding proteins in development. *J. Biomed. Sci.*, 2:293–301, 1995.
- [30] M. W. Hentze and L. C. Kuhn. Molecular control of vertebrate iron metabolism: mRNA-based regulatory circuits operated by iron, nitric oxide, and oxidative stress. *Proc. Natl. Acad. Sci. USA*, 93:8175–8182, 1996.
- [31] O. Hobert. Common logic of transcription factor and microRNA action. *Trends Biochem. Sci.*, 194, 2004. in press, doi 10.1016/j.tibs.2004.07.0001.
- [32] I. L. Hofacker. Vienna RNA secondary structure server. *Nucl. Acids Res.*, 31:3429–3431, 2003.
- [33] I. L. Hofacker, C. Flamm, and P. F. Stadler. Computational chemistry with RNA secondary structures. *Kemija u industriji*, 53:315–322, 2004.
- [34] I. L. Hofacker, W. Fontana, P. F. Stadler, S. Bonhoeffer, M. Tacker, and P. Schuster. Fast folding and comparison of RNA secondary structures. *Monatsh. Chemie*, 125:167–188, 1994.
- [35] I. Huez, L. Créancier, S. Audigier, M. Gensac, A. Prats, and H. Prats. Two independent internal ribosome entry sites are involved in translation initiation of vascular endothelial growth factor mRNA. *Mol. Cell. Biol.*, 18:6178–6190, 1998.
- [36] A. Hüttenhofer, M. Kiefmann, S. Neier-Ewert, J. O’Brien, H. Lehrach, J. Bachellerie, and J. Brosius. Rnomics: an experimental approach that identifies 201 candidates for novel, small, non-messenger RNAs in mouse. *EMBO J.*, 20:2943–2953, 2001.
- [37] F. J. Isaacs, D. J. Dwyer, C. Ding, D. Pervouchine, C. R. Cantor, and J. J. Collins. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat. Biotech.*, 22:841–847, 2004.
- [38] G. H. Jacobs, O. Rackham, P. A. Stockwell, W. Tate, and C. M. Brown. Transterm: a database of mRNAs and translational control elements. *Nucl. Acids Res.*, 30:310–311, 2002.
- [39] M. S. Jurica and M. J. Moore. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell.*, 12:5–14, 2003.
- [40] D. Kampa, J. Cheng, P. Kapranov, M. Yamanaka, S. Brubaker, S. Cawley, J. Drenkow, A. Piccolboni, S. Bekiranov, G. Helt, H. Tammana, and

- T. R. Gingeras. Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res.*, 14:331–342, 2004.
- [41] Y. Komatsu. Regulation of ribozyme activity with short oligonucleotides. *Biol. Pharma. Bull.*, 27:457–462, 2004.
- [42] M. M. Lai. RNA-protein interactions in the regulation of coronavirus RNA replication and transcription. *Biol. Chem.*, 378:477–481, 1997.
- [43] K. A. LeCuyer and D. M. Crothers. Kinetics of an RNA conformational switch. *Proc. Natl. Acad. Sci. USA*, 91:3373–3377, 1994.
- [44] D. Mathews, J. Sabina, M. Zuker, and H. Turner. Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Biol.*, 288:911–940, 1999.
- [45] J. S. Mattick. RNA regulation: a new genetics? *Nature Genetics*, 5:316–323, 2004.
- [46] J. S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.
- [47] N.-C. Meisner, J. Hackermüller, V. Uhl, A. Aszódi, M. Jaritz, and M. Auer. mRNA openers and closers: A methodology to modulate AU-rich element controlled mRNA stability by a molecular switch in mRNA conformation. *Chem. Biochem.*, 2004. in press.
- [48] A. C. Messias and M. Sattler. Structural basis of single-stranded RNA recognition. *Acc. Chem. Res.*, 37:279–287, 2004.
- [49] F. Mignone, C. Gissi, S. Liuni, and G. Pesole. Untranslated regions of mRNAs. *Genome Biology*, 3:reviews0004.1–0004.10, 2002.
- [50] K. Mochizuki, N. A. Fine, T. Fujisawa, and M. A. Gorovsky. Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell*, 110:689–699, 2002.
- [51] J. Møller-Jensen, T. Franch, and K. Gerdes. Temporal translation control by metastable RNA structure. *J. Biol. Chem.*, 276:35707–35713, 2001.
- [52] C. Morey and P. Avner. Employment opportunities for non-coding RNAs. *FEBS Letters*, 567:27–34, 2004.
- [53] J. H. A. Nagel, A. P. Gulyaev, K. Gerdes, and C. W. A. Pleij. Metastable structures and refolding kinetics in *hok* mRNA of plasmid R1. *RNA*, 5:1408–1419, 1999.
- [54] C. J. Nulf and D. Corey. Intracellular inhibition of hepatitis C virus (HCV) internal ribosomal entry site (IRES)-dependent translation by peptide nucleic acids (PNAs) and locked nucleic acids (LNAs). *Nucl. Acids Res.*, 32:3792–3798, 2004.
- [55] H. B. Osborne. An insight into the post-transcriptional control of gene expression in cell function. *Biol. Cell.*, 95:125–127, 2003.
- [56] M. Paulus, M. Haslbeck, and M. Watzele. RNA stem-loop enhanced expression of previously non-expressible genes. *Nucl. Acids Res.*, 32:9/e78, 2004. doi 10.1093/nar/gnh076.

- [57] G. Pesole, F. Mignone, C. Gissi, G. Grillo, F. Licciulli, and L. Sabino. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*, 276:73–81, 2001.
- [58] H. Putzer, N. Gendron, and M. Grunberg-Manago. Co-ordinate expression of the two threonyl-tRNA synthetase genes in *Bacillus subtilis*: Control by transcriptional antitermination involving a conserved regulatory sequence. *EMBO J.*, 11:3117–3127, 1992.
- [59] M. Rehmsmeier, P. Steffen, M. Höchsmann, and R. Giegerich. Fast and effective prediction of microRNA/target duplexes. *RNA*, 2004. in press.
- [60] C. A. Rosen. Regulation of HIV gene expression by RNA-protein interactions. *Trends Genet.*, 7:9–14, 1991.
- [61] R. Rueckert. Picornaviridae: The viruses and their replication. In N. Fields, D. Knipe, and P. Howley, editors, *Virology*, volume 1, pages 609–654. Lippincott-Raven Publishers, Philadelphia, New York, third edition, 1996.
- [62] E. A. Schultes and D. P. Bartel. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. *Science*, 289:448–452, 2000.
- [63] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker. From sequences to shapes and back: A case study in RNA secondary structures. *Proc. Roy. Soc. Lond. B*, 255:279–284, 1994.
- [64] G. A. Soukup and R. R. Breaker. Engineering precision RNA molecular switches. *Proc. Natl. Acad. Sci. USA*, 96:3584–3589, 1999.
- [65] G. A. Soukup and R. R. Breaker. Allosteric nucleic acid catalysts. *Curr. Opin. Struct. Biol.*, 10:318–325, 2000.
- [66] G. Storz, J. A. Opdyke, and A. Zhang. Controlling mRNA stability and translation with small noncoding RNAs. *Cur. Op. Microbiol.*, 7:140–144, 2004.
- [67] M. Suzuki and Y. Hayashizaki. Mouse-centric comparative transcriptomics of protein coding and non-coding RNAs. *BioEssays*, 26:833–843, 2004.
- [68] M. Szymański, M. Z. Barciszewska, M. Żywicki, and J. Barciszewski. Noncoding RNA transcripts. *J. Appl. Genet.*, 44:1–19, 2003.
- [69] M. Tacker, P. F. Stadler, E. G. Bornberg-Bauer, I. L. Hofacker, and P. Schuster. Algorithm independent properties of RNA structure prediction. *Eur. Biophys. J.*, 25:115–130, 1996.
- [70] S. W. Teunissen, M. J. Kruithof, A. D. Farris, J. B. Harley, W. J. Venrooij, and G. J. Pruijn. Conserved features of Y RNAs: a comparison of experimentally derived secondary structures. *Nucl. Acids Res.*, 28:610–619, 2000.
- [71] A. van Zon, M. Mossink, M. Schoester, G. Scheffer, R. Scheper, P. Sonneveld, and E. Wiemer. Multiple human vault RNAs. Expression and association with the vault complex. *J. Biol. Chem.*, 276:37715–37721, 2001.
- [72] G. Varani. RNA-protein intermolecular recognition. *Acc. Chem. Res.*,

- 30:189–195, 1997.
- [73] J. R. Williamson. Induced fit in RNA-protein recognition. *Nat. Struct. Biol.*, 7:834–837, 2000.
- [74] M. C. Yao, P. Fuller, and X. Xi. Programmed DNA deletion as an RNA-guided system of genome defense. *Science*, 300:1517–1518, 2003.