Component-Based Modelling of RNA Structure Folding

Carsten Maus

University of Rostock
Institute of Computer Science, Modelling and Simulation Group
Albert-Einstein-Str. 21, 18059 Rostock, Germany
carsten.maus@uni-rostock.de
http://www.informatik.uni-rostock.de/%7Ecm234

Abstract. RNA structure is fundamentally important for many biological processes. In the past decades, diverse structure prediction algorithms and tools were developed but due to missing descriptions in clearly defined modelling formalisms it's difficult or even impossible to integrate them into larger system models. We present an RNA secondary structure folding model described in ML-DEVS, a variant of the DEVS formalism, which enables the hierarchical combination with other model components like RNA binding proteins. An example of transcriptional attenuation will be given where model components of RNA polymerase, the folding RNA molecule, and the translating ribosome play together in a composed dynamic model.

Keywords: RNA folding, secondary structure, DEVS, model components, multi-level.

1 Introduction

Single stranded ribonucleic acids (RNA) are able to fold into complex three-dimensional structures like polypeptide chains of proteins do. The structure of RNA molecules is fundamentally important for their function, e.g. the well studied structures of tRNA and the different rRNA variants. But also other transcripts of the DNA, i.e. mostly mRNAs, perform structure formation which has been shown to be essential for many regulatory processes like transcription termination and translation initiation [1,2,3]. The shape of a folded RNA molecule can also define binding domains for proteins or small target molecules which can be found for example within riboswitches [4]. The enormous relevance for many biological key processes led to raised research efforts in identifying various RNA structures over the past decades. Unfortunately the experimental structure identification with NMR and X-ray techniques is difficult, expensive, and highly time-consuming. Therefore, many *in silico* methods for RNA structure prediction were developed which cover different requirements. Diverse comparative methods exist using alignments of similar RNA sequences to predict structures [5,6], but also many single sequence prediction algorithms work very well. Some of

M. Heiner and A.M. Uhrmacher (Eds.): CMSB 2008, LNBI 5307, pp. 44–62, 2008.

them predict the most stable RNA structure in a thermodynamical equilibrium, e.g. [7,8,9], whereas some other simulate the kinetic folding pathway over time [10,11,12,13]. The latter is also in the focus of the presented modelling approach here. Results of RNA structure predictions as well as kinetic folding simulations have reached a high level of accuracy and thus *in silico* folding became a widely used and well established technique in the RNA community. However, none of the existing tools and programs provides a flexible integration into larger system models which is also due to the fact that they are written in proprietary formalisms and do not distinguish between model description and simulation engine. To illuminate the importance of the folding processes and the possibility to integrate them into larger models, lets take a look at a concrete example of gene regulation.

2 Motivation – Modelling of Transcription Attenuation

The tryptophan (Trp) operon within bacterial genomes represents one of the best understood cases of gene regulation and has been subject to various modelling approaches [14,15]. Tryptophan is an amino acid, a building block for the production of proteins. The Trp operon includes five consecutive genes, coding for five proteins. The joint action of the proteins permits the synthesis of Trp through a cascade of enzymatic reactions. This ability is vital since the bacterium may be unable to feed on Trp from the environment. As long as Trp is obtained from the surrounding medium, its costly synthesis is impaired by a threefold control mechanism: repression of transcription initiation, transcriptional attenuation, and inactivation of the cascade of enzymatic reactions actually producing Trp. Each of these are triggered by Trp availability. Transcriptional attenuation follows if transcription starts although Trp was available. This has a small but non-negligible chance. As soon as RNA polymerase (RNAP) has transcribed the operon's leader region into an mRNA molecule, a ribosome can access this. The ribosome starts translating the mRNA content into a growing sequence of amino acids. The speed of the ribosome depends on Trp availability. The ribosome advances quickly as long as Trp is abundant, which prevents RNAP from proceeding into the operon's coding region. The attenuation is caused by the formation of a certain constellation of RNA hairpin loops in presence of a Trp molecule at a distinct segment of the mRNA molecule (figure 1). Attenuation depends on the synchronised advance of both RNAP and ribosome, and their relative positioning with respect to mRNA.

In [15] a model of the tryptophan operon was developed, in which repressors, the operon region, and the mRNA were modelled individually. However, the latter in not much detail. Only the repression of transcription initiation was included in the model. Consequently, the simulation result showed stochastic bursts in the Trp level, caused by the repressor falling off, which started the transcription. Integrating a transcriptional attenuation model would have prevented this non-realistic increase in Trp concentration, and mimicked the threefold regulation of the Trp more realistically. However, the question is how to model the

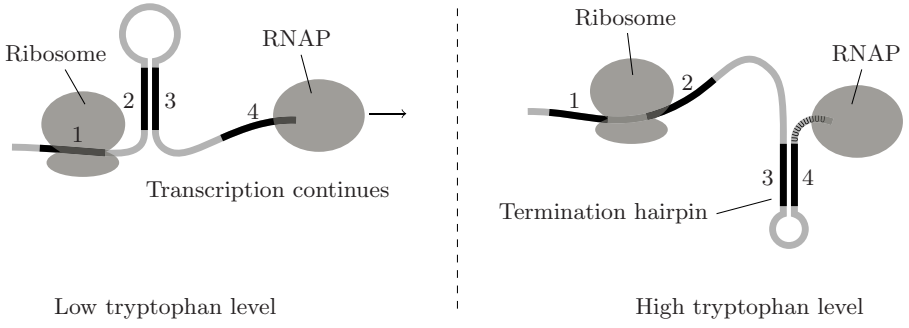


Fig. 1. Attenuation in the Trp leader sequence. With low Trp concentration the ribosome stalls at Trp codons in leader domain 1 and transcription can continue. At high Trp level, the leader domain 1 can be completely translated and thus the ribosome prevents base pairing of domain 2. Leader domains 3 and 4 can form an intrinsic transcription termination structure which causes disruption of the mRNA-RNAP complex.

attenuation. As this regulating process depends largely on structure formation, modelling of RNA folding would be a big step in the right direction for reflecting attenuation dynamics. Additionally, modelling interactions between mRNA and RNAP as well as mRNA and the ribosome are needed because both influence the kinetic folding process and RNA termination structures break up gene transcription. The focus of this paper is the RNA folding process, but at the end we will also give a detailed outlook how the composed model of tryptophan attenuation looks like and how the individual model components act together.

3 Principles of RNA Folding

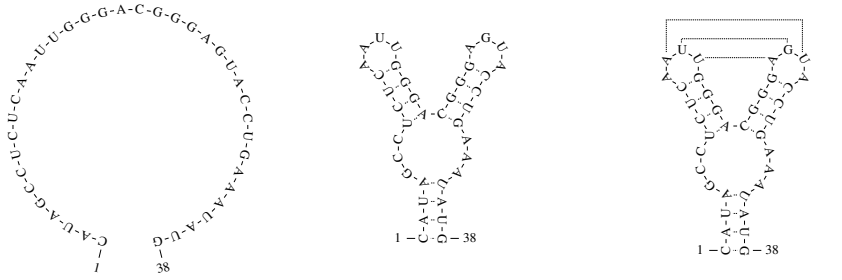
3.1 Thermodynamics

The reason for RNA folding is the molecules' general tendency to reach the most favourable thermodynamical state. Complementary bases of RNA nucleotides can form base pairs by building hydrogen bonds similar to DNA double helices. Adenine (A) and Uracil (U) are complementary bases as well as Cytosine (C) and Guanine (G). In addition, the wobble base pair G-U is also frequently found in RNA structure folding. Each additional hydrogen bond of base pairs affords a small energy contribution to the overall thermodynamic stability, but there is another chemical interaction which is even more important for the RNA structure than just the number and type of base pairs. It's called base stacking and describes the interactions between the aromatic rings of adjacent bases by Van-der-Waals bonds. Base pair stacking is the cause why an uninterrupted long helix is thermodynamically more favourable than a structure of multiple single base pairs or short helices interrupted by loop regions, even if the number and type of base pairs are equal. Since the 1970s, significant progress has been done on identifying thermodynamic parameters of different base pair neighbourhoods

and structural elements like hairpin loops, e.g. [16,17]. This was a precondition to develop RNA structure prediction algorithms based on energy minimisation, i.e. finding the thermodynamical most stable structure.

3.2 Primary, Secondary, and Tertiary Structure

RNA structures are hierarchically organised (see figure 2). The most simple hierarchy level is the primary structure which is nothing else than the linear sequence of nucleotides. Two nucleotides are linked over the 3' and 5' carbon atoms of their ribose sugar parts resulting in a definite strand direction. The secondary structure consists of helices formed by base pairs and intersecting loop regions. Such structural elements are formed rapidly within the first milliseconds of the folding process [18]. Interacting secondary structure elements finally build the overall three-dimensional shape of RNA molecules. Although they are formed by simple base pairs like secondary structures, helices inside loop regions are often seen as tertiary structures. Such pseudoknots and higher order tertiary interactions are, due to their complexity and analog to many other RNA structure prediction methods, not covered by our model. However, it should not retain unstated here that there are some existing tools which can predict pseudoknots quite well, e.g. [10].



Primary structure Secondary structure Tertiary structure

Fig. 2. Different hierarchical levels of RNA folding

4 Modelling Formalism

As already mentioned, typically kinetic RNA folding simulations, as e.g. [10,11,12], are aimed at efficiently and accurately simulating the molecules structure formation in isolation rather than supporting a reuse of RNA folding models and a hierarchical construction of models. For approaching such model composition, we use the modelling formalism ML-DEVS [19], a variant of the DEVS formalism [20]. As DEVS does, it supports a modular-hierarchical modelling and allows to define composition hierarchies. ML-DEVS extends DEVS by supporting variable structures, dynamic ports, and multi-level modelling. The latter is based on two ideas. The first is to equip the coupled model with a state and a behaviour of its own, such that

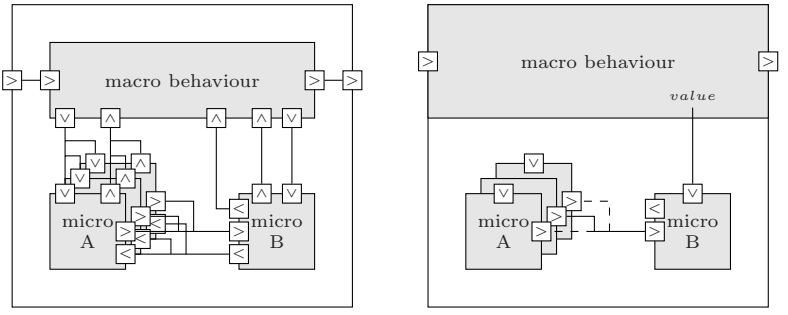


Fig. 3. Comparison of multi-level modelling with DEVS and ML-DEVS. (left) With DEVS the macro level is modelled at the same hierarchical level as the micro level models. (right) With ML-DEVS the macro dynamics are part of the coupled model. Functions for downward and upward causation reduce the number of explicit couplings needed.

the macro level does not appear as a separate unit (an executive) of the coupled model. Please recall that in traditional DEVS coupled models do not have an own state nor a behaviour. Secondly, we have to explicitly define how the macro level affects the micro level and vice versa. Both tasks are closely interrelated. We assume that models are still triggered by the flow of time and the arrival of events. Obviously, one means to propagate information from macro to micro level is to exchange events between models. However, this burdens typically modelling and simulation unnecessarily, e.g. in case the dynamics of a micro model has to take the global state into consideration. Therefore, we adopt the idea of value couplings. Information at macro level is mapped to specific port names at micro level. Each micro model may access information about macro variables by defining input ports with corresponding names. Thus, downward causation (from macro to micro) is supported. In the opposite direction, the macro level needs access to crucial information at the micro level. For this purpose, we equip micro models with the ability to change their ports and to thereby signalise crucial state changes to the outside world. Upward causation is supported, as the macro model has an overview of the number of micro models being in a particular state and to take this into account when updating the state at macro level. Therefore, a form of invariant is defined whose violation initiates a transition at macro level. In the downward direction, the macro level can directly activate its components by sending them events – thereby, it becomes possible to synchronously let several micro models interact which is of particular interest when modelling chemical reactions. These multi-level extensions facilitate modelling, figure 3 depicts the basic idea, see also [21].

5 The RNA Folding Model

The central unit in composed models using RNA structure information is an RNA folding model. Therefore, we first developed a model component which describes the folding kinetics of single stranded RNA molecules. It consists of

a coupled ML-DEVS model representing the whole RNA molecule and several atomic models.

5.1 Nucleotides

Each nucleotide (nt) of the RNA strand is represented by an instance of the atomic model *Nucleotide* which is either of the type A, C, G, or U meaning its base. They are connected via ports in the same order as the input sequence (primary structure) and have knowledge about their direct neighbours. For example, the nt at sequence position 8 is connected with the nt number 7 on its 5' side and on the 3' location it is connected with the nt at position 9 (see figure 4). State variables hold rudimentary information about the neighbours, to be exact their base type and current binding partners. "Binding partner" means a secondary structure defining base pair and the term is used only in this context here and does not mean the primary backbone connections. If a partner of a nucleotide changes, an output message will be generated and the receiving (neighboured) nucleotides will update their state variables. Holding information about other atomic model states is normally not the case in DEVS models as they are typically seen as black boxes. However, here it is quite useful because of some dependencies concerning base pair stability.

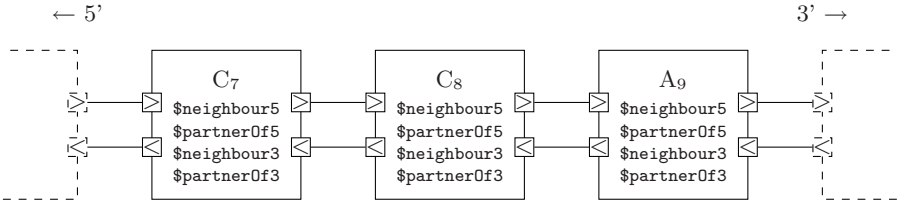


Fig. 4. Adjacent nucleotides are connected via input and output ports. A set of variables stores basic information about their neighbourhood.

Base pairs are modelled by wide range connections of nucleotides via additional interfaces. Whereas the RNA backbone bonds of adjacent nucleotides are fixed after model initialisation, the connections between base pairing nucleotides are dynamically added and removed during simulation (figure 5). Therefore, two different major states (phases) of nucleotides exist: they can be either *unpaired* or *paired*.

As already stated in section 3.1, base pair stability depends on the involved bases and their neighbourhood, especially stacking energies of adjacent base pairs provide crucial contributions for structure stabilisation. In our kinetic folding model, base pair stability is reflected by binding duration, i.e. the *time advance* function of the *paired* phase. Thus, pairing time depends on thermodynamic parameters for nucleic acids base stacking which were taken from [17] and are also be used by MFOLD version 2.3 [7]. This thermodynamic data set not only provides the free energy (Gibbs energy) for a given temperature of 37°C, but also

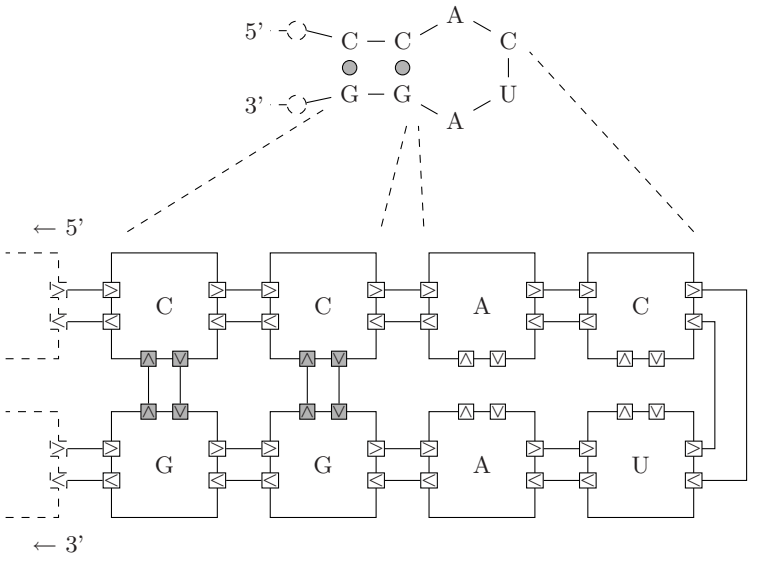


Fig. 5. Small hairpin loop structure. (top) Conventional RNA secondary structure representation. (bottom) The same structure represented by atomic DEVS models. Base pairing is modelled by dynamically coupled nucleotides (grey interfaces) with a distance of at least 5 sequence positions reflecting the minimal hairpin loop size.

the enthalpy change ΔH of various stacking situations. The enthalpy together with the free energy and the absolute temperature allows us to calculate the entropy change ΔS which allows us further to calculate the activation energy ΔE_a for base pair dissociation at any temperature T between 0 and 100°C:

$$\Delta E_a = -(\Delta H - T\Delta S) \; .$$

ΔE_a is directly used as one parameter for base pair opening time, i.e. the duration of a *paired* phase is directly dependent on the activation energy for base pair disruption. To allow RNA structures to escape from local energy minima and refold to more stable structures, the base pair dissociation time will be randomised, which leads to very short bonding times in some cases although the activation energy needed for opening the base pair is quite large.

For base pair closing an arbitrary short random time is assigned with the *unpaired* phase of nucleotides assuming that RNA base pair formation is a very fast and random process. After *unpaired* time has expired, the nucleotide model tries to build a base pair with another nucleotide randomly chosen from within a set of possible pairing partner positions. This set is determined by sterically available RNA strand regions and thus an abstraction of spatial constraints. For example, a hairpin loop smaller than 4 nucleotides is sterically impossible, but also many other nucleotide positions with larger distance can be excluded for secondary structure folding (see figure 6).

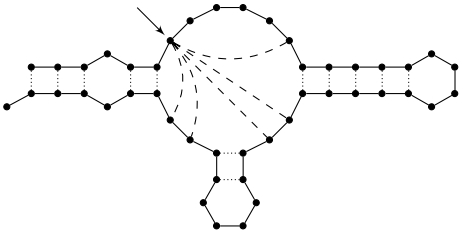


Fig. 6. Possible partners for secondary structure base pairing. The arrow-marked nucleotide is only able to pair with 5 positions within its own loop region (connected via dashed lines). All other nucleotides are excluded from base pairing with the marked position.

An *unpaired* nucleotide is not able to choose another nt for base pairing by its own. It has no global information about the RNA shape which is strongly needed here. Therefore, an implicit request by adding a new input port will be made to the coupled model that holds such macro knowledge and can therefore choose a valid position with which the requesting nt will try to pair next. For choosing this position at macro level two different model variants exist. The first and more simple one picks a position totally random from within the set of possible partners, whereas the second variant takes the entropy change into account when a pairing partner will be selected. The latter method prefers helix elongation in contrast to introducing new interior loops by single base pair formations and small loops will be more favourable than large ones [12]. Correct RNA folding with the first method is dependent on the base pair stabilities and the random folding nuclei which are the first appearing base pairs of helical regions. This last point is less important for RNA folding with model variant 2 because the chosen binding partners are more deterministic due to loop entropy consideration. A comparison of both approaches with respect to simulation results is given in section 6. Once a *nucleotide* received an input message by the macro model containing the number of a potential pairing partner, it tries to form a base pair with this nt by adding a coupling and sending a request. For a successful pairing, the partners must be of complementary base type and they must be able to pair in principle, e.g. bases can be modified so that they can not bind to others. Figure 7 illustrates the whole state flow for base pair formation and disruption of the *nucleotide* model component.

5.2 Macro Level Model

The role of the macro model and its interactions with the micro level (nucleotides) shows the schematic organisation of the whole RNA folding model in figure 8. Already mentioned in the previous section, high level information about the whole RNA molecule is needed to take sterical restrictions for base pairing into account. Therefore, the coupled model holds the overall RNA secondary structure which will be updated every time the state of a nucleotide changes

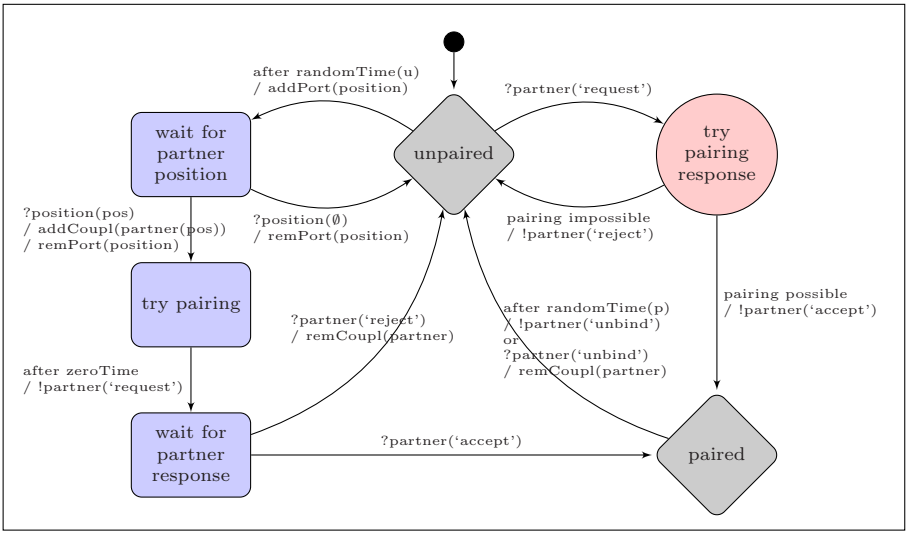


Fig. 7. *Nucleotide* state flow in a Statechart-like representation. Main states are depicted by grey diamonds. States of a nucleotide which takes up the active part of a base pair are shown as rounded rectangles and a state exclusively passed through by a reacting nucleotide is drawn as a circle. Sending and receiving events over ports have prefixed exclamation marks and question marks respectively.

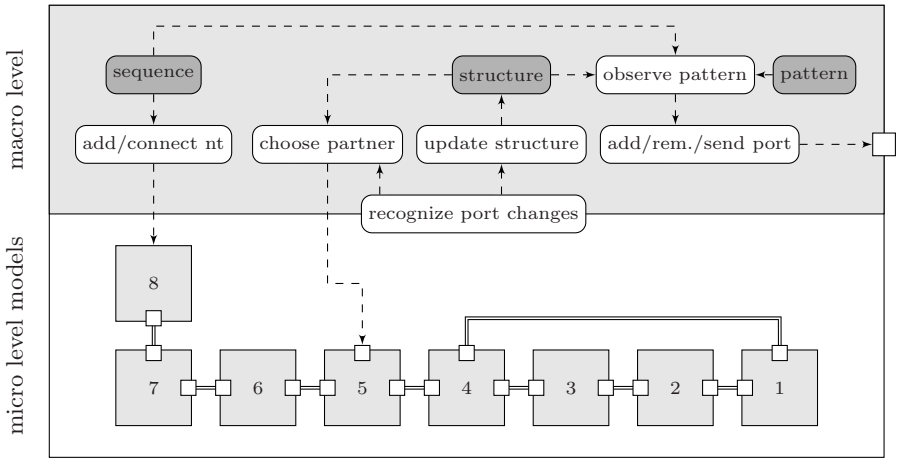


Fig. 8. Schematic overview of the whole RNA folding model. Model components are shaded in light grey and explicit couplings between them are drawn by solid double lines. Dashed lines indicate model-internal information flow and function calls.

from *unpaired* to *paired* and vice versa. This will be triggered by *nucleotide* port adding and removal recognised by the macro level. The same functionality is used to signalise the macro level the wish to try pairing with another nucleotide. The macro model detects a port adding, calculates the sterically possible partner set, chooses a position from within the set, and after all sends this position to the just now added *nucleotide* input port (figure 8, nt 5). The coupled macro model is further responsible for sequence initialisation on the micro level by adding and connecting *nucleotide* models, i.e. it generates the primary RNA structure. Another task of the macro model is to observe the current folding structure for special structural patterns. This could be for example a specific binding domain for a protein or small ligand. Also transcription termination or pausing structures can be of interest for observation. If observed structures are present during folding simulation, the macro level model can signalise this information and thus trigger dynamics to other components by adding new ports to itself (representing docking sites) or send messages over existing ports. A composed example model which uses this capability can be found in section 7.

6 Evaluation of the Folding Model

For evaluating the model's validity we simulated the folding of different RNA molecules with known structure and analysed the results. Three different types of experiments were done:

Native Structure– Correlates the majority of formed structures with the native structure after sufficient long simulation time?

Structure Distribution– Is the equilibrium ratio between minimum free energy (mfe) and suboptimal structures as expected?

Structure Refolding– Are molecules able to refold from suboptimal to more stable structural conformations?

Unfortunately, only few time-resolved folding pathways are experimentally derived and most of them treat pseudoknots [22] and higher order tertiary structure elements [22,23,24] which can not be handled by our folding model and are therefore out of the question for a simulation study. Hence, some comparisons with other *in silico* tools were also made, although we know that one has to be careful with comparing different models for validating a model as it is often unclear how valid the other models are. Because the folding model is highly stochastic, every simulation experiment was executed multiple times. Typically 100 replications were made.

6.1 Native Structure

Structural analysis of the cis-acting replication element from Hepatitis C virus revealed a stem hairpin loop conformation where the helix is interrupted by an internal or bulge loop region [25]. Figure 9 shows simulation results of its structure

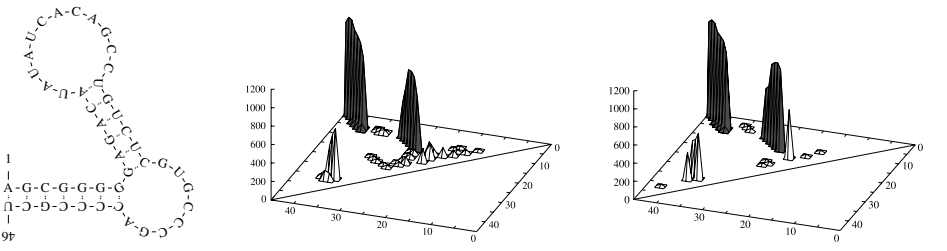


Fig. 9. Folding simulation of the Hepatitis C virus cis-acting replication element. Parameters: time 1000 ms, temperature 310.15 K, 100 replications. (left) Observed secondary structure [25]. (middle) Base pair probability matrix. Simulation with base pair formation model 1. (right) Base pair formation model 2. Peak heights indicate base pair lifetime during simulation and the native helices are shaded.

formation. The three-dimensional base pair lifetime plots indicate correct folding of both helical regions and only few misfolded base pairs. Only small differences can be seen between simulations with the two base pair formation variants described in section 5.1. Without taking entropy into account for pairing, a bit more noise of misfolded base pairs can be observed which is not surprising due to the absolutely random partner choice.

Another well known RNA structure is the cloverleaf secondary structure of tRNAs [26,27] consisting of four helical stems: the amino acid arm, D arm, anticodon arm, and the T arm. Some base modifications and unusual nucleotides exist in tRNA which stabilise its structure formation, e.g. dihydrouridine and pseudouridine. Such special conditions are not considered by our folding model as well as tertiary interactions leading to the final L-shaped form. However, folding simulations result in significant cloverleaf secondary structure formation (figure 10). Although there is much misfolded noise, the four distinct helix peaks are the most stable structural elements introduced during simulation, especially the amino acid arm. No fundamental difference can be seen between base pair formation model 1 and 2.

A third native structure validation experiment treats the Corona virus s2m motif which is a relatively long hairpin structure with some intersecting internal and bulge loops [28,29]. Simulation of the SARS virus s2m RNA folding indicates only for one of the native helix regions a conspicuous occurrence (figure 11). The other helices closer to the hairpin loop show no significant stabilisation. Competing misfolded structural elements can be observed equally frequent or even more often. Base pair formation model 2 provides a slightly better result than the first one, but it is unsatisfying too. A reason for the result can be the multiple internal and bulge loops, which destabilise the stem and thus allow locally more stable structure elements to form.

6.2 Structure Distribution

In [30] a quantitative analysis of different RNA secondary structures by comparative imino proton NMR spectroscopy is described. The results indicate that

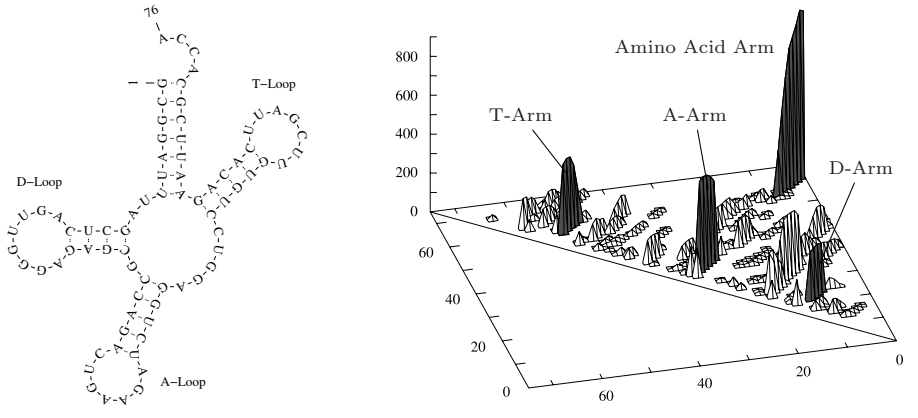


Fig. 10. Simulation of the yeast tRNA^{Phe} folding. Parameters: time 1000 ms, temperature 310.15 K, base pair formation model 2, 100 replications. (left) Native cloverleaf secondary structure. (right) Base pair probability matrix. Peak heights indicate base pair lifetime during simulation and the native helices are shaded.

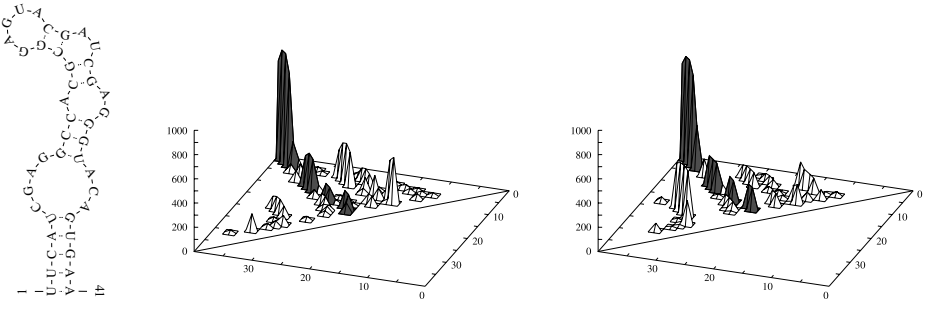


Fig. 11. Folding simulation of the SARS virus s2m motif. Parameters: time 1000 ms, temperature 310.15 K, 100 replications. (left) Known secondary structure [29]. (middle) Base pair probability matrix. Simulation with base pair formation model 1. (right) Base pair formation model 2. Peak heights indicate base pair lifetime during simulation and the native helices are shaded.

a small 34-nt RNA has two equally stable structures in thermodynamic equilibrium, one with 2 short helices and the other with a single hairpin. Folding simulations of the same RNA strand show an equal ratio of the probed structures as well (figure 12). However, both are representing just 20% of all present structures which was not detected in the NMR experiments. Many base pairs were introduced during simulation which are competing with base pairs of the two stable structures and thus reduce their appearance. This can be easily seen in the 3D matrix of figure 12 where some additional peaks show high misfolded base pair lifetimes. Simulating the RNA folding with KINFOLD [12]

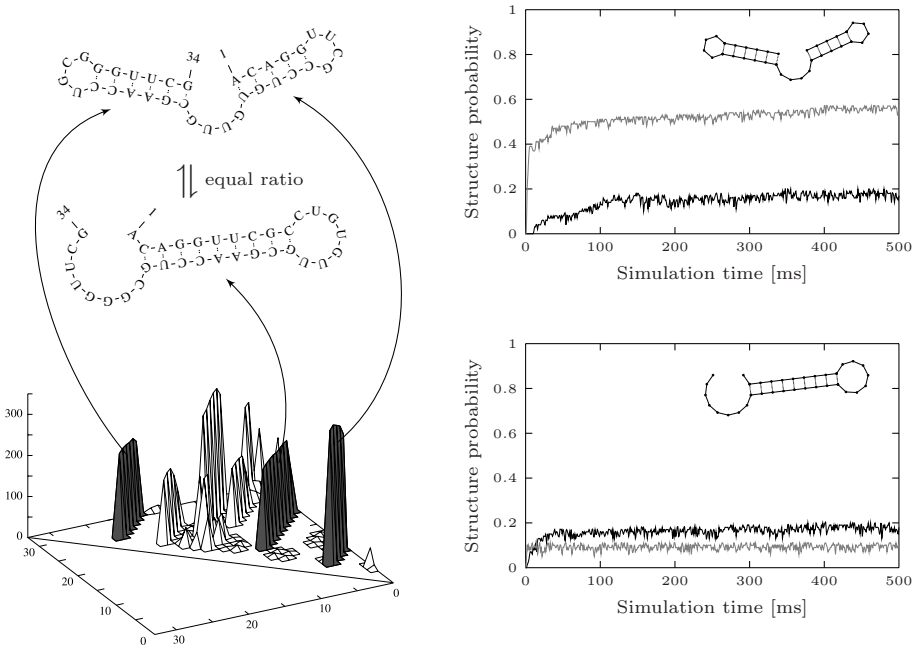


Fig. 12. Folding simulation of a designed small bistable RNA [30]. Parameters: temperature 298.15 K, base pair formation model 2, 100 replications. Grey curves describe mean structure occurrence from 100 KINFOLD simulation runs with 298.15 K.

results in a five times higher amount of the 2-helix conformation than the single hairpin, but their total sum is about 60% of all molecules and thus less misfolded structures can be observed.

6.3 Structure Refolding

Real-time NMR spectroscopy was used by Wenter et al. to determine refolding rates of a short 20-nt RNA molecule [31]. The formation of its most stable helix was temporarily inhibited by a modified guanosine at position 6. After photolytic removal of this modification a structure refolding was observed. To map such forced structure formation to relatively unstable folds at the beginning of an experiment, most RNA folding tools have the capability to initialise simulations with specified structures. We used, much closer to the original wet-lab experiment, a different strategy, i.e. at first G6 was not capable to bind any other base. The time course after removing this prohibition during simulation is shown in figure 13. Wenter et al. detected structure refolding by measuring the imino proton signal intensity of U11 and U17, which show high signal intensity if they are paired with other bases. Accordingly we observed the state of both uracils over simulation time as well. After removal of G6 unpaired locking, a logarithmic decrease of structures with paired U11 and uniform increase of folds

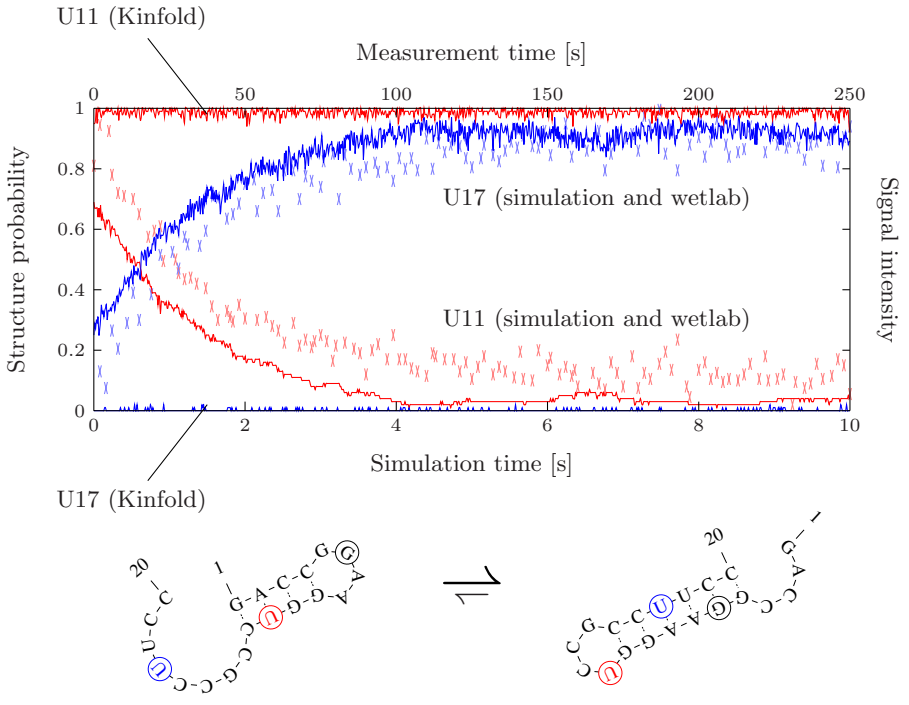


Fig. 13. Refolding of an deliberately misfolded small RNA molecule [31]. Wetlab measurements are drawn by a X. Simulation parameters: time 10 seconds, temperature 288.15 K, base pair formation model 1, 100 replications. Simulations with Kinfold were made with the same temperature and replication number but over a time period of 250 seconds.

with paired U17 can be observed reaching a complete shift of the conformational equilibrium after 4 seconds. A very similar refolding progression was experimentally measured (single spots in figure 13), but with a strong deviating time scale of factor 25. This could be a remaining model parameter inaccuracy or due to special experimental conditions which are not covered by our model, e.g. unusual salt concentrations. However, our model allows a quite realistic refolding from a suboptimal to a more stable RNA structure. Identical *in silico* experiments with KINFOLD [12] by contrast, do not show any significant refolding (figure 13, non-changing curves). The same counts for SEQFOLD [11]. With both approaches the energy barrier seems to be too high to escape from the misfolded suboptimal structure.

6.4 Assessment of Experiments

Local optima can more easily be overcome in comparison to other traditional pure macro methods (see figure 13). We assume that even "stable" base pairs

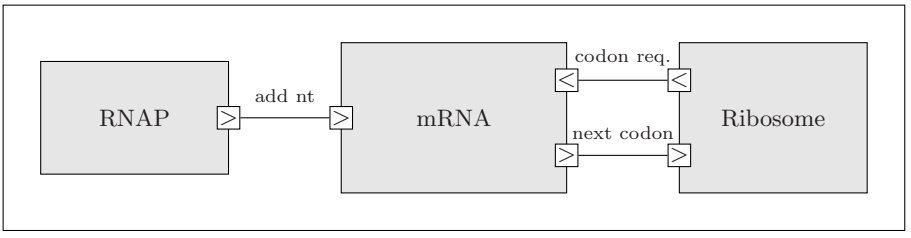


Fig. 14. Connecting a folding mRNA model with RNA polymerase and ribosome to reflect transcription attenuation

might be subject to changes, and let the nucleotides "searching" for a stable structure at micro level. This proved beneficial and emphasised the role of the micro level. However, the simulation revealed the importance of macro constraints for the folding process, and the implications of a lack of those. Macro constraints that have been considered are for example the relative positioning of the nucleotides, particularly within spatial structures like hairpin or internal loops. The interplay between macro and micro level allowed us to reproduce many of the expected structure elements, e.g. figures 9 and 10, although macro constraints have been significantly relaxed. These simplifications lead to "wrongly" formed structures and maybe could have been prevented by integrating terminal base stacking for pairing stability as well as less abstract base pair closing rules as macro constraints. A comparison of the two implemented base pair formation methods indicate only few differences. Without taking entropy into account the noise of unstable single base pairs and short helices increases, but not dramatically. The same stable structures are formed based on both rules.

7 Composed Attenuation Model

Having a working folding model we are now able to combine it with other model components that are influenced by or are influencing the RNA structure formation and come back to the motivation, the attenuation of tryptophan synthesis. At least two further models are needed to reflect transcription attenuation: the RNA polymerase and the ribosome (figure 14).

7.1 RNA Polymerase

RNA molecules are products of the transcription process which is the fundamental step in gene expression. Once the RNA polymerase enzyme complex (RNAP) has successfully bound to DNA (transcription initiation), it transcribes the template sequence into an RNA strand by sequentially adding nucleotides to the 3' end and thus elongates the molecule. To reflect this synthesising process, in the RNA model, new *nucleotide* models and their backbone connections are added dynamically during simulation. This process is triggered by the RNAP model

component which interacts with RNA. This dynamic RNA elongation allows the simulation of sequential folding, where early synthesised parts of the RNA molecule can already fold whereas other parts still have to be added. Please note that this is not a unique feature of the model presented here, as kinetic folding tools typically realise sequential folding by just adding a new nt after a certain time delay. However, a component design allows to combine the RNAP with further models (e.g. the DNA template), or to model it in more detail (e.g. diverse RNAP subunits), and to exchange model components on demand. The pattern observation function of the RNA folding model, which is realised at macro level, allows us to look for an intrinsic transcription termination structure [2] during simulation. If such structure is formed, the folding model removes its elongation input port meaning the release of the RNA from the polymerase enzyme. At this time point the elongation stops, but structure folding and interactions with other components proceed.

7.2 Ribosome

The ribosome enzyme complex translates RNA sequences into protein determining amino acid sequences (peptides). Translation starts at the ribosome binding site of mRNA molecules which is reflected by a pair of input and output ports of the RNA models. The translation begins after connecting it with a ribosome model. The current ribosome position with respect to the RNA sequence is known by the RNA model. A triplet of three RNA bases (codon) encodes for one amino acid. The ribosome requests for the next codon 3' of its current RNA location when peptide chain elongation has proceeded. This is the case when the correct amino acid of the last codon entered the enzyme. The speed of the translation process depends strongly on the availability of needed amino acids. If an amino acid type is not sufficiently available, the ribosome stalls at the corresponding codon and thus pauses translation. A ribosome is quite big and thus 35-36 nucleotides are covered by its shape [32]. Therefore, a region upstream and downstream of the ribosome location is not able to form base pairs. As the RNA model knows the ribosome location, this is handled by the RNA macro level model which sends corresponding events to its nucleotide micro model components. The same counts for the helicase activity of the ribosome [32]. For sequence translation, the macro level model will disrupt a base paired structure element when it is reached by the enzyme.

Whether those additional models are realised as atomic models, or coupled models depends on the objective of the simulation study. Referring to the operon model presented in [15], the RNAP, the mRNA, and the ribosome would replace the simplistic mRNA model, to integrate the attenuation process into the model.

8 Conclusion

We presented a component-based model of RNA folding processes. Unlike traditional approaches which focus on the results of the folding process, e.g. stable

structures in thermodynamical equilibrium, our objective has been different. The idea was to develop an approach that allows to integrate the folding processes into larger models and to take the dynamics into account, that has shown to be crucial in many regulation processes. Therefore, the formalism ML-DEVS was used. At macro level, certain constraints referring to space and individual locations were introduced, whereas at micro level, the nucleotides were responsible for a successful base pairing and for the stability of the structure. A model component for the nucleotides and one model component for the entire RNA molecule have been defined. The simulation results have been compared to wetlab experiments. Therefore, the model components can be parametrised for different RNA sequences (base types) as well as environmental conditions (e.g temperature). The evaluation revealed an overall acceptable performance, and in addition, insights into the role of micro level dynamics and macro level constraints. The integration of the RNA folding model into a model of transcription attenuation has been sketched. Next steps will be to realise this integration and to execute simulation experiments to analyse the impact of this more detailed regulation model on the synthesis of tryptophan.

Acknowledgments. Many thanks to Adelinde M. Uhrmacher for her helpful comments and advice on this work. I will also thank Roland Ewald and Jan Himmelspach for their instructions for using JAMES II. The research has been funded by the German Research Foundation (DFG).

References

1. Kaberdin, V.R., Blasi, U.: Translation initiation and the fate of bacterial mRNAs. FEMS Microbiol. Rev. 30(6), 967–979 (2006)
2. Gusarov, I., Nudler, E.: The Mechanism of Intrinsic Transcription Termination. Mol. Cell 3(4), 495–504 (1999)
3. Yanofsky, C.: Transcription attenuation: once viewed as a novel regulatory strategy. J. Bacteriol. 182(1), 1–8 (2000)
4. Nahvi, A., Sudarsan, N., Ebert, M.S., Zou, X., Brown, K.L., Breaker, R.R.: Genetic Control by a Metabolite Binding mRNA. Chem. Biol. 9(9), 1043–1049 (2002)
5. Torarinsson, E., Havgaard, J.H., Gorodkin, J.: Multiple structural alignment and clustering of RNA sequences. Bioinformatics 23(8), 926–932 (2007)
6. Hofacker, I.L., Fekete, M., Stadler, P.F.: Secondary Structure Prediction for Aligned RNA Sequences. J. Mol. Biol. 319(5), 1059–1066 (2002)
7. Zuker, M.: Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31(13), 3406–3415 (2003)
8. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, L.S., Tacker, M., Schuster, P.: Fast folding and comparison of RNA secondary structures. Monatsh. Chem./Chemical Monthly 125(2), 167–188 (1994)
9. Rivas, E., Eddy, S.R.: A dynamic programming algorithm for RNA structure prediction including pseudoknots. J. Mol. Biol. 285(5), 2053–2068 (1999)
10. Xayaphoummine, A., Bucher, T., Thalmann, F., Isambert, H.: Prediction and Statistics of Pseudoknots in RNA Structures Using Exactly Clustered Stochastic Simulations. Proc. Natl. Acad. Sci. U.S.A. 100(26), 15310–15315 (2003)

11. Schmitz, M., Steger, G.: Description of RNA Folding by "Simulated Annealing". J. Mol. Biol. 255(1), 254–266 (1996)
12. Flamm, C., Fontana, W., Hofacker, I.L., Schuster, P.: RNA folding at elementary step resolution. RNA 6(3), 325–338 (2000)
13. Flamm, C., Hofacker, I.L.: Beyond energy minimization: approaches to the kinetic folding of RNA. Monatsh. Chem./Chemical Monthly 139(4), 447–457 (2008)
14. Santillán, M., Mackey, M.C.: Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data. Proc. Natl. Acad. Sci. U.S.A 98(4), 1364–1369 (2001)
15. Degenring, D., Lemcke, J., Röhl, M., Uhrmacher, A.M.: A Variable Structure Model – the Tryptophan Operon. In: Proc. of the 3rd International Workshop on Computational Methods in Systems Biology, Edinburgh, Scotland, April 3-5 (2005)
16. Serra, M.J., Lyttle, M.H., Axenson, T.J., Schadt, C.A., Turner, D.H.: RNA hairpin loop stability depends on closing base pair. Nucleic Acids Res. 21(16), 3845–3849 (1993)
17. Walter, A.E., Turner, D.H., Kim, J., Lyttle, M.H., Müller, P., Mathews, D.H., Zuker, M.: Coaxial Stacking of Helixes Enhances Binding of Oligoribonucleotides and Improves Predictions of RNA Folding. Proc. Natl. Acad. Sci. U.S.A. 91(20), 9218–9222 (1994)
18. Russell, R., Millett, I.S., Tate, M.W., Kwok, L.W., Nakatani, B., Gruner, S.M., Mochrie, S.G.J., Pande, V., Doniach, S., Herschlag, D., Pollack, L.: Rapid compaction during RNA folding. Proc. Natl. Acad. Sci. U.S.A. 99(7), 4266–4271 (2002)
19. Uhrmacher, A.M., Ewald, R., John, M., Maus, C., Jeschke, M., Biermann, S.: Combining Micro and Macro-Modeling in DEVS for Computational Biology. In: WSC 2007: Proceedings of the 39th conference on Winter simulation, pp. 871–880. IEEE Press, Los Alamitos (2007)
20. Zeigler, B.P., Praehofer, H., Kim, T.G.: Theory of Modeling and Simulation. Academic Press, London (2000)
21. Uhrmacher, A.M., Himmelspach, J., Jeschke, M., John, M., Leye, S., Maus, C., Röhl, M., Ewald, R.: One Modelling Formalism & Simulator Is Not Enough! A Perspective for Computational Biology Based on James II. In: Fisher, J. (ed.) FMSB 2008. LNCS (LNBI), vol. 5054, pp. 123–138. Springer, Heidelberg (2008)
22. Bokinsky, G., Zhuang, X.: Single-molecule RNA folding. Acc. Chem. Res. 38(7), 566–573 (2005)
23. Sclavi, B., Sullivan, M., Chance, M.R., Brenowitz, M., Woodson, S.A.: RNA Folding at Millisecond Intervals by Synchrotron Hydroxyl Radical Footprinting. Science 279(5358), 1940–1943 (1998)
24. Zhuang, X., Bartley, L.E., Babcock, H.P., Russell, R., Ha, T., Herschlag, D., Chu, S.: A Single-Molecule Study of RNA Catalysis and Folding (2000)
25. You, S., Stump, D.D., Branch, A.D., Rice, C.M.: A cis-Acting Replication Element in the Sequence Encoding the NS5B RNA-Dependent RNA Polymerase Is Required for Hepatitis C Virus RNA Replication. J. Virol. 78(3), 1352–1366 (2004)
26. Rich, A., RajBhandary, U.L.: Transfer RNA: Molecular Structure, Sequence, and Properties. Annu. Rev. Biochem. 45(1), 805–860 (1976)
27. Clark, B.F.C.: The crystal structure of tRNA. J. Biosci. 31(4), 453–457 (2006)
28. Jonassen, C.M., Jonassen, T.O., Grinde, B.: A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J. Gen. Virol. 79(4), 715–718 (1998)

29. Robertson, M.P., Igel, H., Baertsch, R., Haussler, D., Ares, M.J., Scott, W.G.: The structure of a rigorously conserved RNA element within the SARS virus genome. PLoS Biol. 3(1), 86–94 (2005)
30. Höbartner, C., Micura, R.: Bistable Secondary Structures of Small RNAs and Their Structural Probing by Comparative Imino Proton NMR Spectroscopy. J. Mol. Biol. 325(3), 421–431 (2003)
31. Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H., Pitsch, S.: Kinetics of Photoinduced RNA Refolding by Real-Time NMR Spectroscopy. Angew. Chem. Int. Ed. Engl. 44(17), 2600–2603 (2005)
32. Takyar, S., Hickerson, R.P., Noller, H.F.: mRNA Helicase Activity of the Ribosome. Cell 120(1), 49–58 (2005)