A Graph Grammar for Modeling RNA Folding

A. L. Mamuye, E. Merelli and L. Tesei

GaM 2016 Eindhoven, Netherlands

April 3, 2016

Outline

Graph for Modelling Complex Biological Systems

Complex biological systems are made of a number of components that interact with each other in a nonlinear fashion.

It is necessary to understand the interaction between components and pathways.

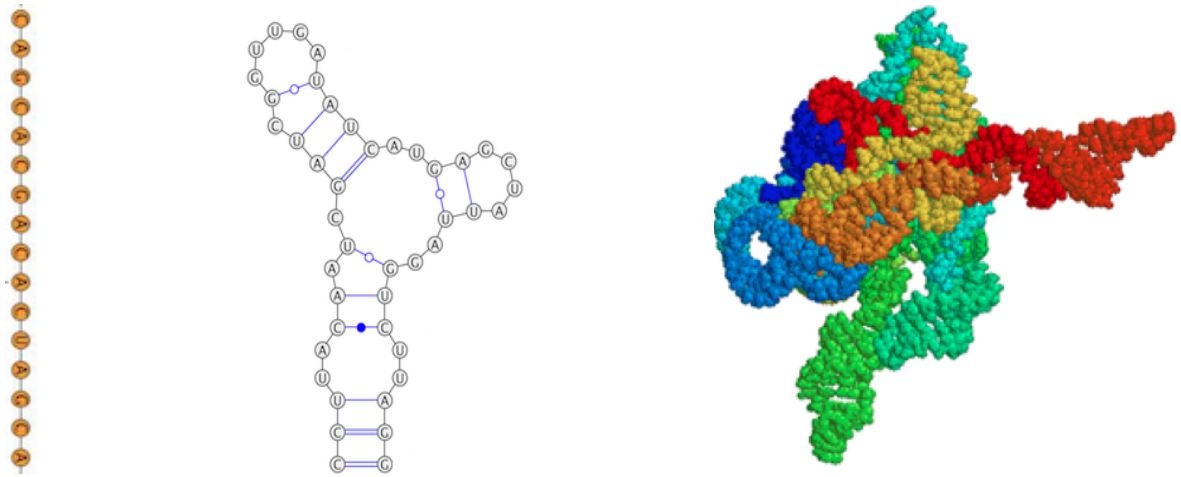
Graph can be applied to modeling complex biological systems.

The elements of a system are represented as vertices of a graph and the interactions between them are represented as edges.

Graph algorithm can be used to analyze, simulate and visualize the system.

RNA: as a Graph

RNA naturally exhibits auto-regulative mechanism that continuously triggers sequences of foldings until stable folding is attained.

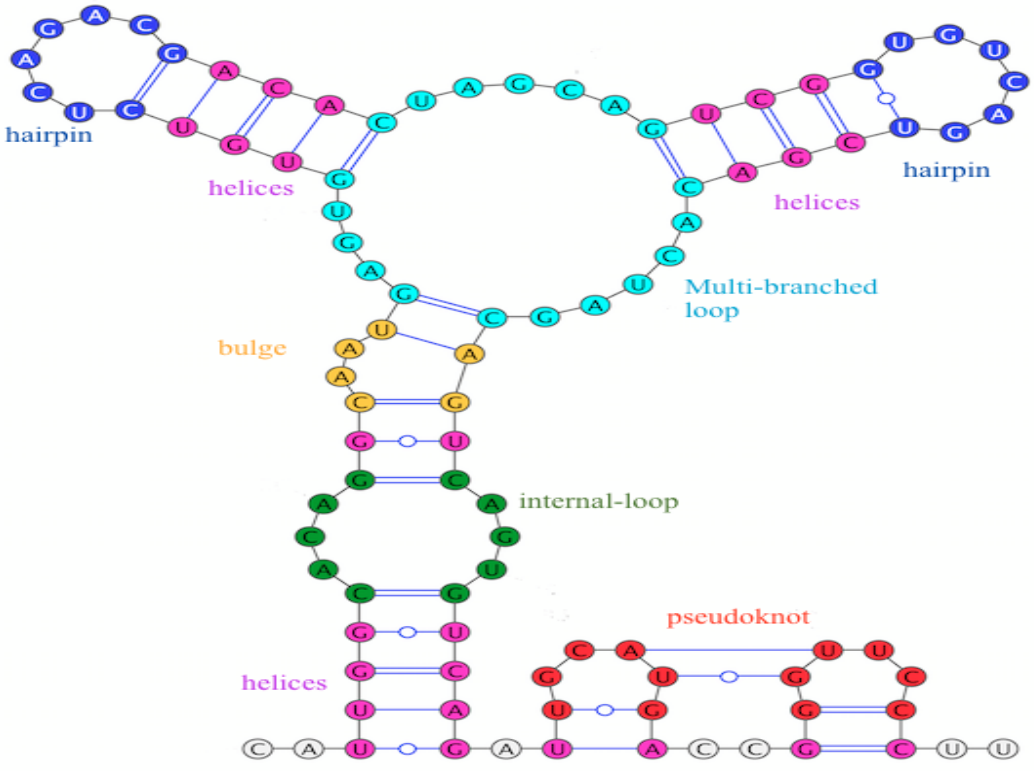


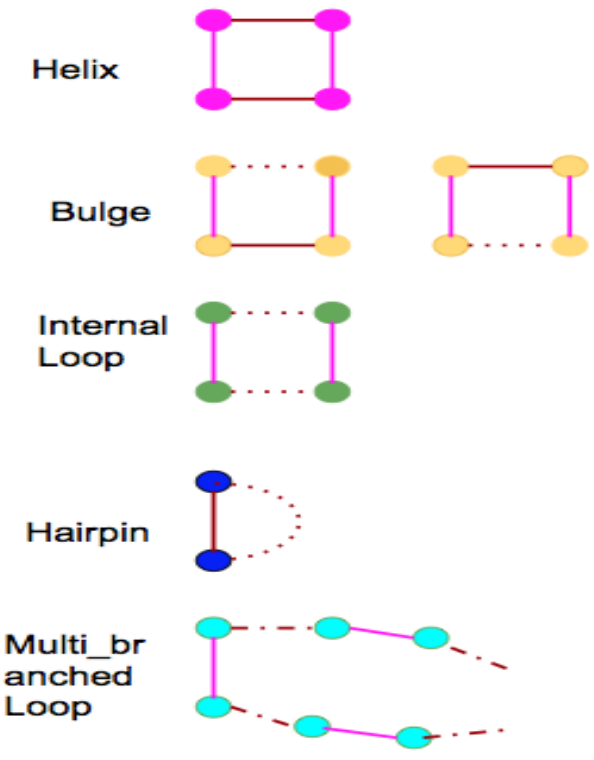
RNA folding can be regarded as a hierarchical process in which secondary structure (SS) forms before tertiary structure.

SS formation is due to nucleotide base pairings, namely Watson-Crick (C-G and A-U) and Wobble (G-U).

RNA SS Sub-Graphs

The structure of RNA can be regarded as a conformation with various local elements (sub-graphs), called structural elements.





SSs can be distinguished as pseudoknot free and pseudonotted.

Why RNA Prediciton

RNA plays a central role in:

- Protein synthesis
- Enzymatic catalysis
- Genome organization

Predicting RNA SSs is important for inferring structure-function relationship of RNA molecules.

Approaches Used

The optimal (minimum free energy) RNA secondary structure can be determined

- Experimental techniques: X-ray, crystallography and NMR
- Computational methods: dynamic programming and comparative sequence analysis

Predicting pseudoknot free structure is inaccurate [Lorenz et. al., 2016] and predicting pseudoknotted structure is NP-complete [Lyngsø et. al., 2000].

Moreover, it is still an open question to what extent the functional structures of natural RNAs are determined by folding kinetic than by the optimal one [Flamm et. al., 2008].

Study Aim

RNA molecule exhibits:

- Dynamical behaviour
- self-adaptability behaviour

We devise a new approach based on:

- Graph transformation
- S[B]-paradigm

Graph Transformation

It has been used in computational biology in different contexts:

- Employed to model the evolution of developmental pathways [Benk et. al., 2004]
- Used to encode RNA tertiary structure motifs [St-Onge et. al., 2007]
- Gene expression was simulated using a general purpose graph rewriting system [Schimmel et. al., 2009]

An RNA primary structure is represented as a graph and its folding evolution as a graph transformation in the folding space.

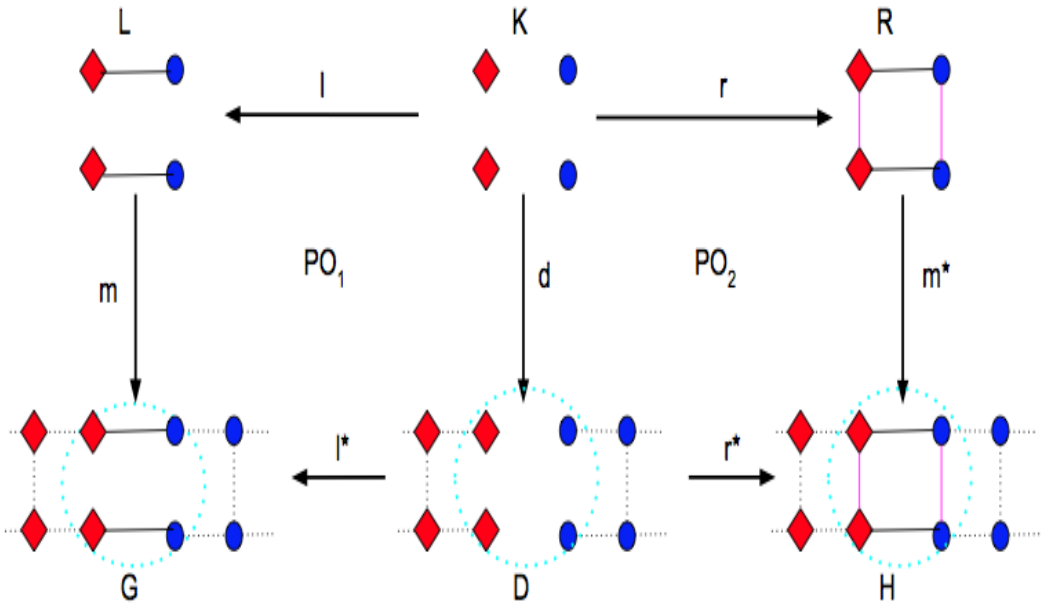
RNA Graph Grammar and Transformation

DPO is used to model RNA folding evolution.

RNA graph grammar:

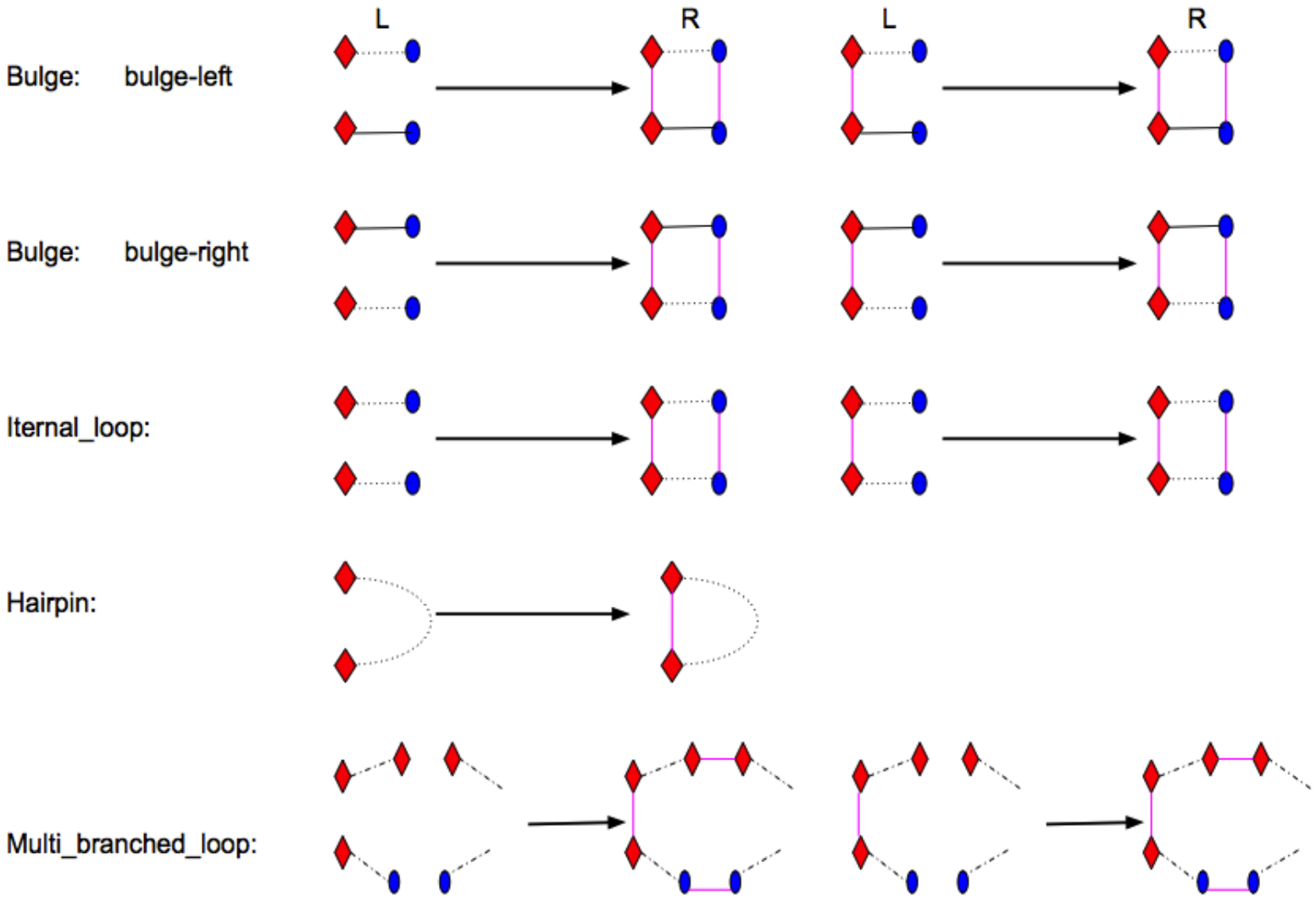
$$G_{RNA} = (\{p : L \leftarrow K \rightarrow R\}_{p \in P}, G_0)$$

p $(Helix - 1)$:



RNA Graph Grammar...

Production rules:



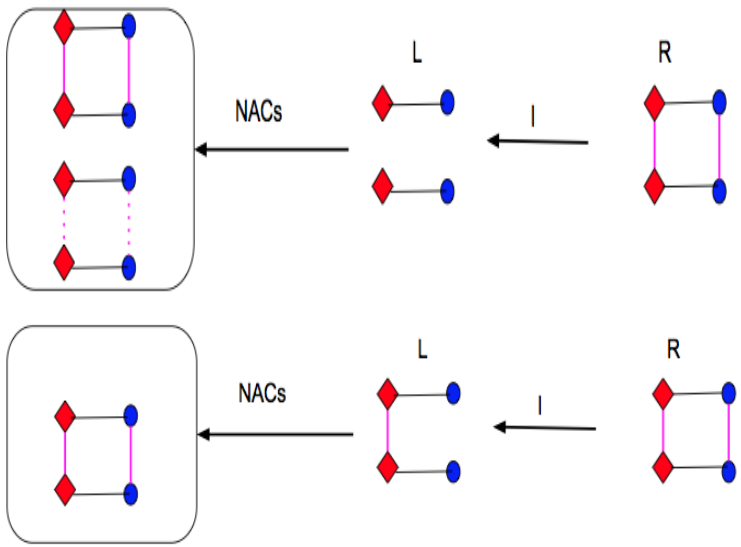
Negative Application Conditions

DPO rules are applicable whenever there are a valid matches for their LHS.

Only G-C and A-U (Watson-Crick) and G-U (wobble).

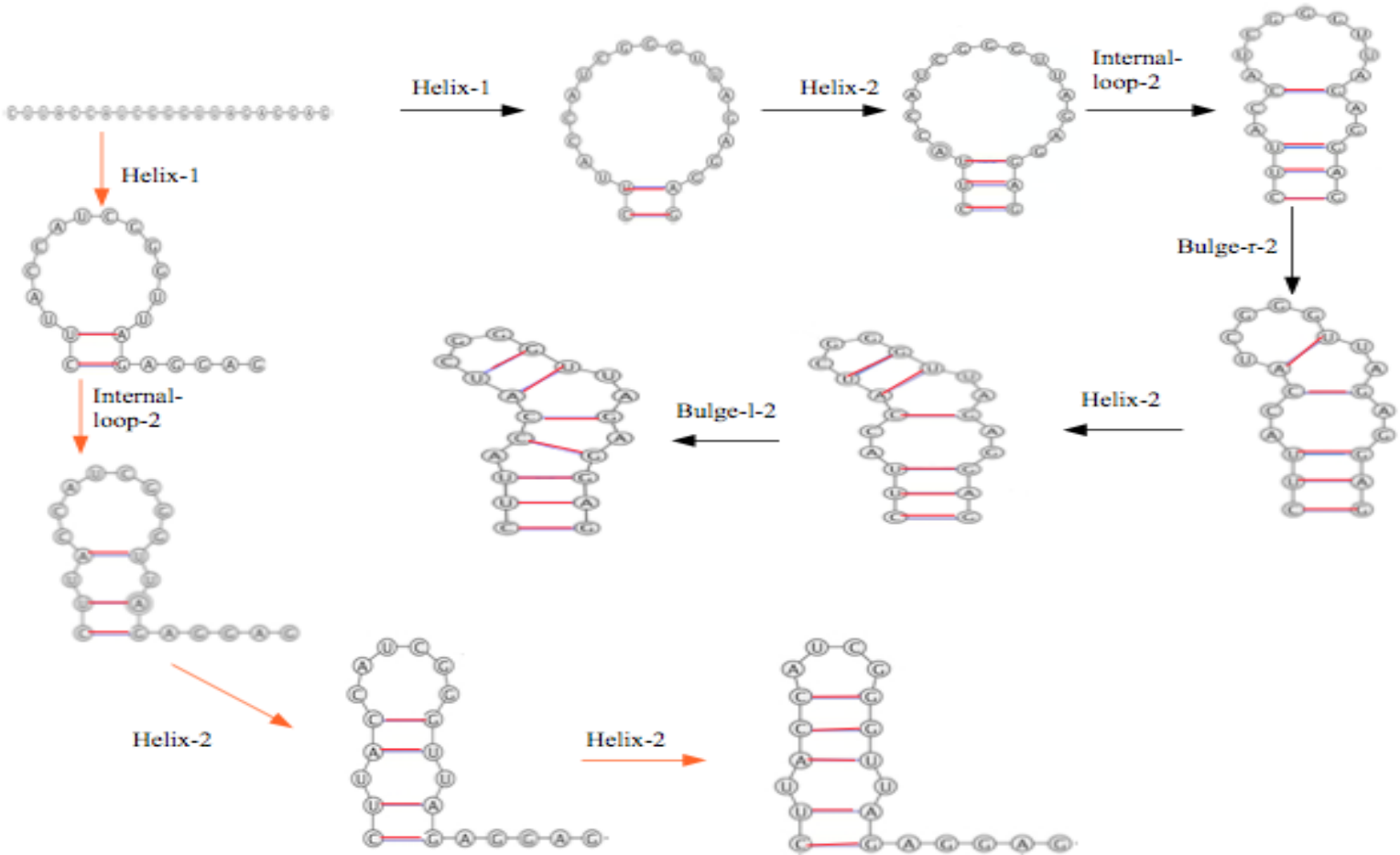
The application of the rewriting rules must take into account a NAC:

- Each nucleotide can form a base pair by interacting with at most one other nucleotide.



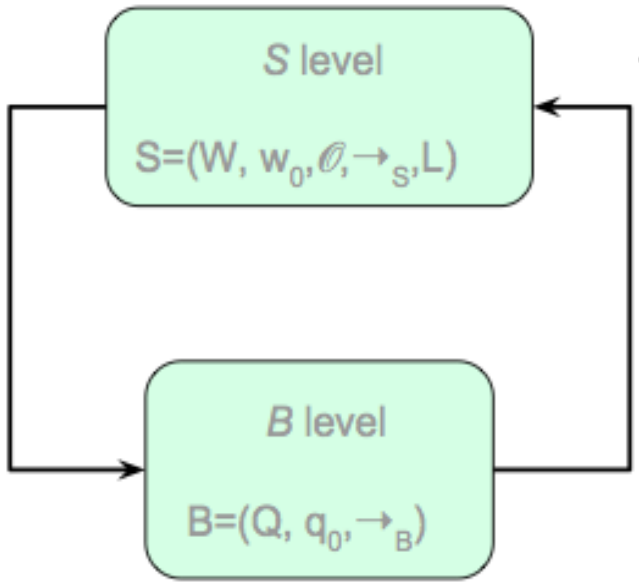
RNA Folding as Graph Transformation: a toy example

Two possible folding transformations of an RNA molecule with 21 nucleotides



S[B] Paradigm: an Overview

A computational model for self-adaptive systems with two coupled levels.

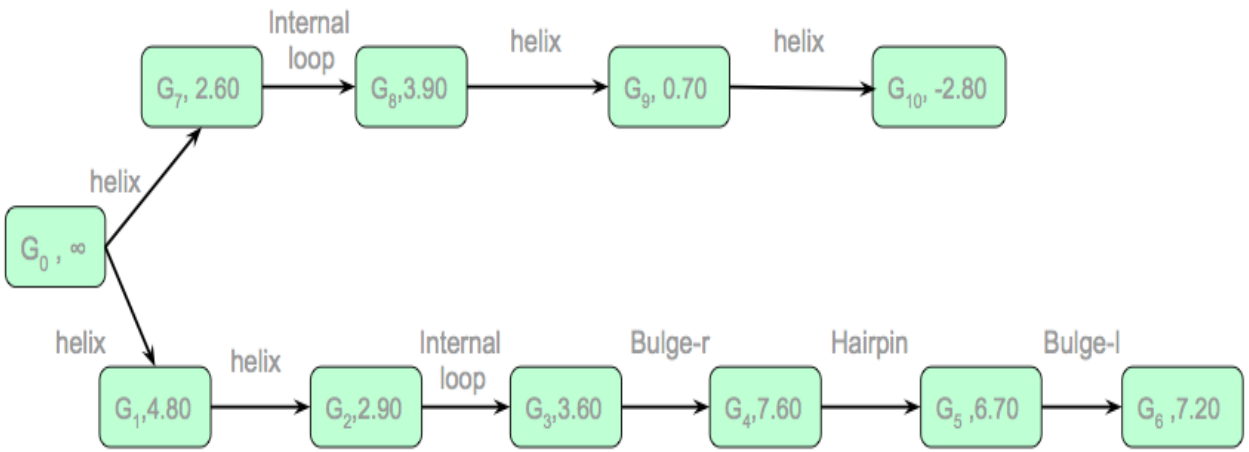


- B (behavioural level): describes the admissible dynamics of the system
- S (structural level): accounts for global and stable features of the system, also regulates B.

RNA Folding Process as a Self-adaptive System

RNA folding is presented as a graph rewriting for which the dynamics is defined by a set of rewriting rules.

A Labelled Transition System (LTS) is constructed with graphs as states and rule application as transitions:



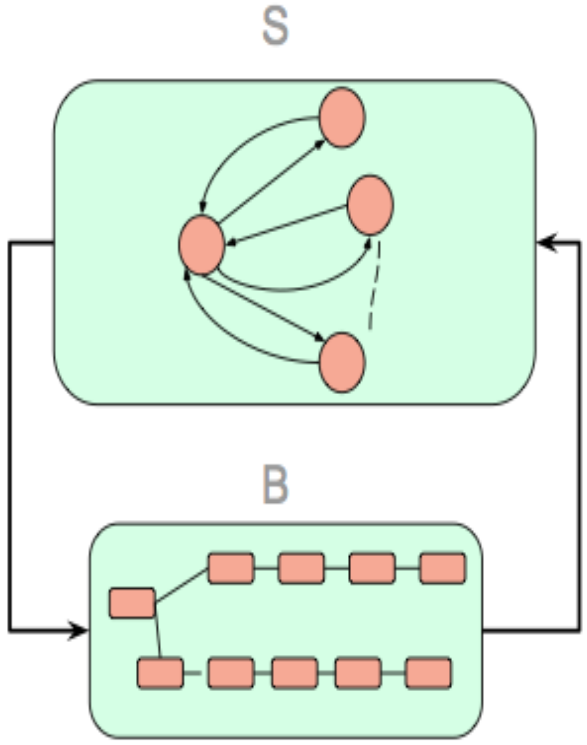
At each state in the LTS the value of the free energy is called the observable value.

RNA Folding Process as a Self-adaptive System

We use the LTS defined above as the B level of an S[B] model.

For the S level we use the state machine.

Each state of S level is associated with a constraint over the observable variables (free energy) of the B level.



RNA Folding Process as a Self-adaptive System

The S-level impose constraint on the states of B, i.e., q:

Constriant

$$\phi_0 = \exists q_{\text{next}} \in \text{next}(q) : \mathcal{O}(q_{\text{next}}) \leq \mathcal{O}(q) \wedge (\forall q' \in \text{next}(q) : \mathcal{O}(q_{\text{next}}) \leq \mathcal{O}(q'))$$

Where:

- q is the current state of the B level
- $\mathcal{O}(q)$ is the observable value associated to q
- $\text{next}(q)$ a function giving the successors states of q in B

How the Adaption Works

The adaption is triggered whenever q cannot evolve because there is no next B state that satisfies the current S constraints.

During adaption, the S[B] system attempts to evolve towards a new B state that satisfies a new S state, chosen among the successors of the current one.

Adaption terminates successfully when B ends up in a state that fulfils the new global situation (free energy) represented by one of the admissible S states.

Issue of Scalability

An RNA sequence of n nucleotides has $\approx \quad 1.8^n$ possible SSs [Zuker et. al., 1984].

Attempting to explore the complete folding space of RNA SS is computationally expensive.

Approaches used to handle scalability issues:

- Dynamic programming: explores the RNA SS space to find the lowest free energy structure without explicitly generating all possible structures.
- A stochastic simulation, described as a continuous time Markov process, used to handle the RNA folding kinetics.

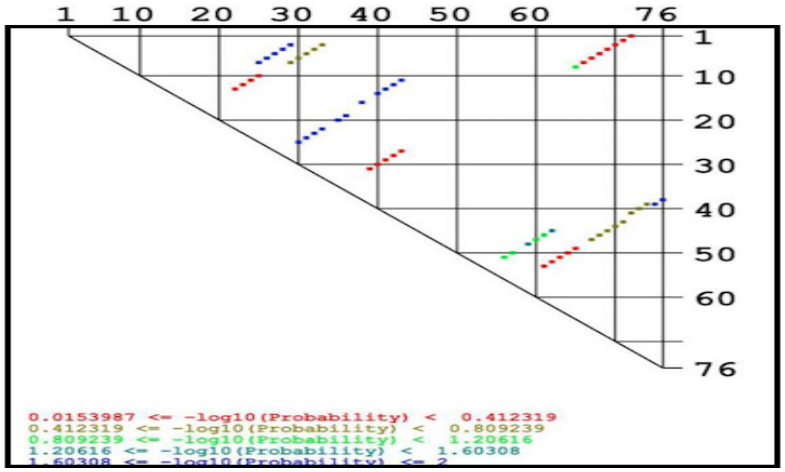
Issue of Scalability

Possible scenarios:

- Graph Transformation:
  - Stochastic graph transformation: a rule with lower energy would be more likely to be applied than a rule with a higher energy
  - Use rules with more probable base pairs.
- State space Exploration:
  - guided greed search algorithm: guided depth first search.

Future Work

- Exploit the partition function of RNA folding space to identify the replacement graph with the more probable base-pairs.



- Check whether parallelism, concurrency and bisimulation theorems are still valid in case of RNA folding.

Acknowledgement

Our study is financed by the European project: TOPology DRIven Methods for complex systems (TOPDRIM) - FP7-ICT/ICT-2011.9.7.
http://www.topdrim.eu

The goal of this project is to provide methods for describing the dynamics of multi-level complex systems based on topological data analysis.

Reference

1. C.Flamm and I.L.Hofacker(2008):Beyond Energy Minimization:Approaches to the Kinetic Folding of RNA. Mon. Che.-Chemical Mon. 139(4):447-457.
2. R. Lorenz et al. (2016): SHAPE Directed RNA Folding. Bioinformatics 32.1 (2016): 145-147. PMC. Web. 28 Mar. 2016.
3. R. Lyngsø and N. Pedersen. Pseudoknot in RNA Secondary Structures. Journal of Computational Biology, 409-427.
4. M. Benk et. al. Graph Grammars as Models for the Evolution of Developmental Pathways. In: The Logic of Artificial Life: Abstracting and Synthesizing the Principles of Living Systems; Proc. of the 6^{th} German Workshop on Artificial Life, April 14-16, 2004, Bamberg, Germany
5. J. Schimmel et. al. (2009). Gene Expression with General Purpose Graph Rewriting Systems. Elec. Com. of the EASST 18.
6. K. St-Onge et. al. (2007). Modeling RNA Tertiary Structure Motifs by Graph-Grammars. Nucleic acids research 35(5):1726-1736.

Thank You!