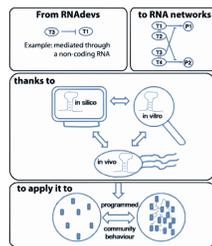


Introduction

As part of the RiboNets project (www.ribonets.eu) we aim to develop a newly designed RNA-based toolbox for cellular computing. This toolbox will be created following a three-step process: (i) rational de-novo design and analysis of RNA-based devices using computational approaches, (ii) selecting best performers in vitro within highly parallel microfluidic reactors and, finally, (iii) integrating and testing them in living cells. The combination of all three layers of analysis, *in silico*, *in vitro* and *in vivo*, is a major point of this project. The comprehensive aim is to build complex regulatory networks and to examine the underlying mechanisms of information transmission and processing within cells. This will allow an unconventional way of fast and energy efficient computing based on a large number of well-characterised RNA-based devices without affecting the host-cell. In this sub-project we will focus on the first step of this approach, the de-novo design of RNA based devices and the computational analysis and selection of potential design solutions.



In silico design of RNA devices

Uncovering the function of an RNA molecule is a time consuming and challenging task. Therefore we aim to solve the inverse problem, namely building a molecule with well described, predefined properties - and therefore functions (Fig 1). The design process starts with generating sequences with specific structural and sequence constraints, followed by an optimisation towards an objective function and a further selection of the best solutions, which show additional desired properties not included in the objective function.

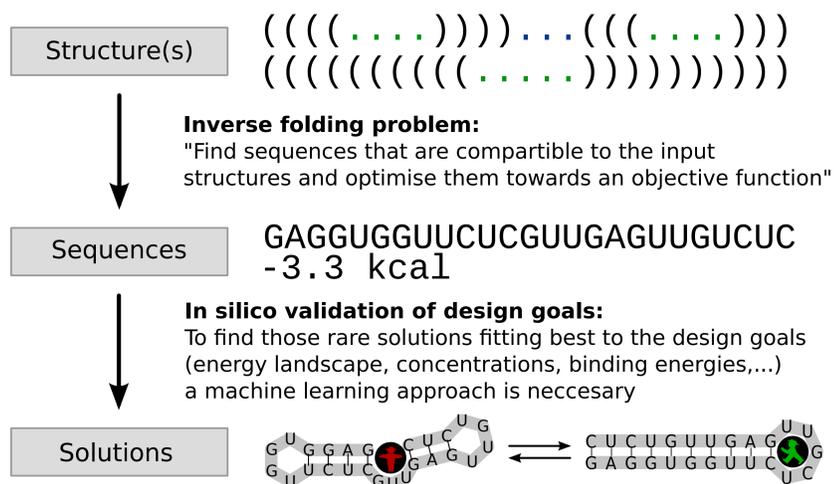


Fig 1: General concept of the RNAdesign pipeline. The input is a well characterised and defined description of the design goals. This includes structural constraints, sequence constraints (SD, RBS,...) as well as energy landscape properties such as binding energies, MFE or even kinetic properties like equilibrium distributions and structural switching rates. As we cannot build a single objective function handling all the goals, we need to analyse and select the best solutions using a machine learning approach.

Sampling in the solution space

The most important feature to gain defined functionality of an RNA sequence is its ability to fold into a specific structure. Therefore we restrict the solution space by introducing positive structural constraints, implemented as a so called Dependency Graph (Fig 2A). Recursive enumeration of possible nucleotides on this graph then forms the basis of our dynamic programming approach (Fig 2B). To traverse through the solution space, stochastic backtracking is used. At this stage sequence constraints are introduced [1].

A prospective goal is to support dynamic structural constraints, meaning that the input can contain elements of variable length or even abstract shapes. Thus, it is necessary to pre-select specific instances from this set of input structures for which to solve the design problem. This are probably those structures with the biggest solution space.

Optimisation towards an objective function

Valid sequences need to be optimised towards a objective function in order to gain specific energy landscape properties. For a multi-stable riboswitch we typically include the energy of the partition function for the desired states, the Gibbs free energy of the ensemble as well as the energy of struct. This assures that our desired states are dominant in the ensemble and have a equal and low MFE separated by a barrier.

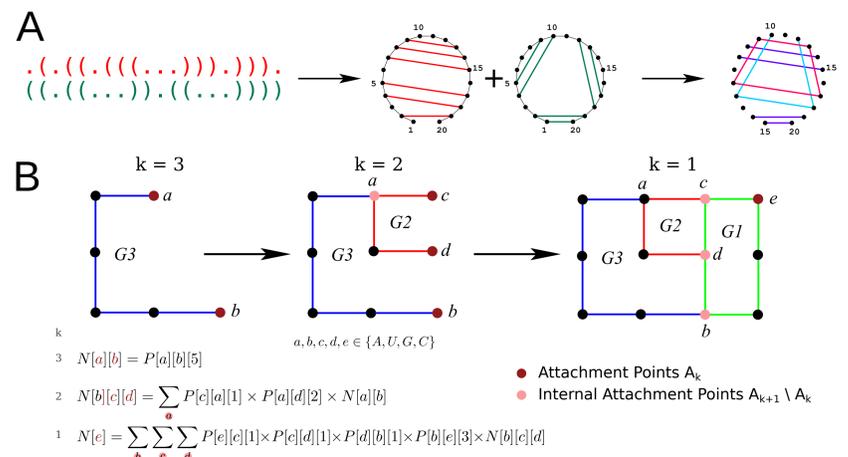


Fig 2: (A) A Dependency Graph Ψ is used to represent all structural constraints. It is derived from the union of Nussinov representations. Ψ is a finite, undirected graph with $|V|$ = sequence-length and $|E|$ = number of base pairs. (B) This recursive enumeration of colourings (= assign base to position) forms the basis of our dynamic programming approach. Stochastic backtracking is used to sample bases for the Attachment Points. Paths in between are coloured with fixed end assignments.

Analysis and selection of best solutions

With the described procedure we can produce big amounts of RNA devices in short time. However, this makes it necessary to in silico analyse the output and select the best solutions. Therefore we analyse many features and collect them in a multi dimensional feature vector. Self organised maps [2] can then be used for clustering solutions with similar features and visualise them as 2D plots (Fig. 3). This makes it possible for humans to process this huge amount of data and select the best designs. As this is work in progress, many interesting data is still missing in our current feature vectors - e.g. tree-edit-distance, dG binding, equilibrium concentrations,...

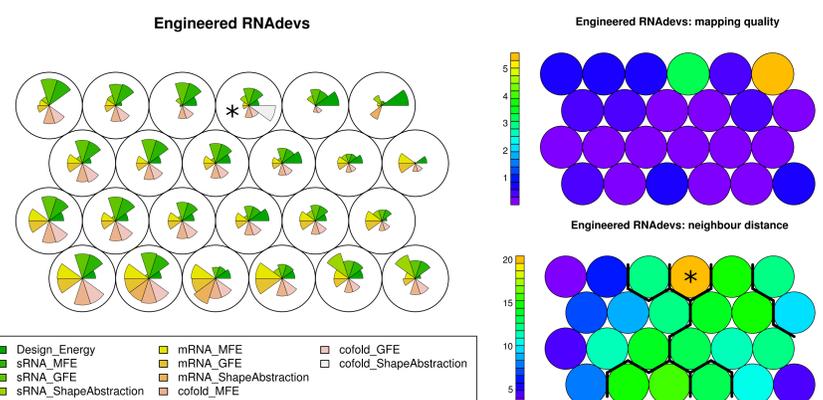


Fig 3: Self organised maps make it possible to cluster solutions according to their multidimensional feature vector and visualise the data in 2D. Here we clustered the feature vector of 50 designs for a sRNA-induced translational OFF switch. Good solutions show small energies and a big shape abstraction level equality between design and solution. Such a cluster is marked with an asterisk.

In vitro validation of desired properties and functions

The selected RNA devices will be sent to the laboratory for in vitro analysis and validation of the design goals and functions. If the desired functionality is not present, we will start an iterative process of design refinement and in vitro analysis. Further we will collect experimental data such as binding energies, switching rates or actual structures, to feed back into the computational models. Methods like PURE [3], SHAPE, fluorescence spectroscopy [4] or microfluidics [5] will help to uncover the actual behaviour of our RNA devices.

Accomplished RNA devices and sensors are then further tested in vivo and finally assembled into complex logic gates which can build up de-novo synthetic regulation networks.

Such networks can, for example, be used to bring synthetic logic to already characterised pathways: e.g. changing quorum sensing behaviour of bacteria or optimise drug production with well defined regulation mechanisms.

References

- Höner zu Siederdisen, C. et al. Computational design of RNAs with complex energy landscapes. *Biopolymers* 99, 1124–1136 (2013).
- Wehrens, R. kohonen: Supervised and unsupervised self-organising maps. (2013).
- Shimizu, Y., Kanamori, T. & Ueda, T. Protein synthesis by pure translation systems. *Methods* 36, 299–304 (2005).
- Bourdoncle, A. et al. Quadruplex-Based Molecular Beacons as Tunable DNA Probes. *J. Am. Chem. Soc.* 128, 11094–11105 (2006).
- Estévez-Torres, A. et al. Fourier transform to analyse reaction-diffusion dynamics in a microsystem. *Lab Chip* 8, 1205–1209 (2008).