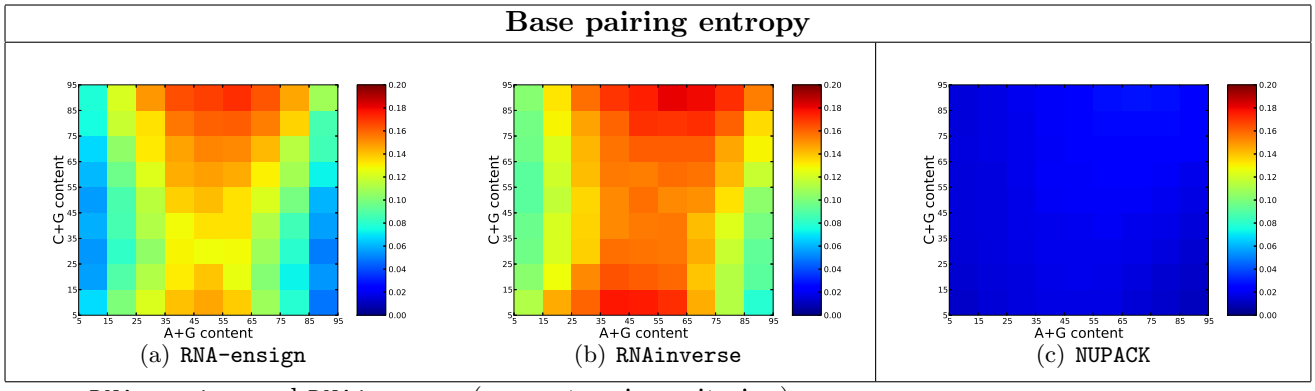
A global sampling approach to designing and reengineering of RNA secondary structures (Supplementary Material)

Alex Levin, Mieszko Lis, Yann Ponty, Charles W. O'Donnell,
Srinivas Devadas, Bonnie Berger and Jérôme Waldispühl.

1 Impact of seed composition on base pairing entropy

This table completes the data showed in Figure 2. We study the influence of the `A+G` content and `C+G` content of the seed on the base pairing entropy of the designed sequence.

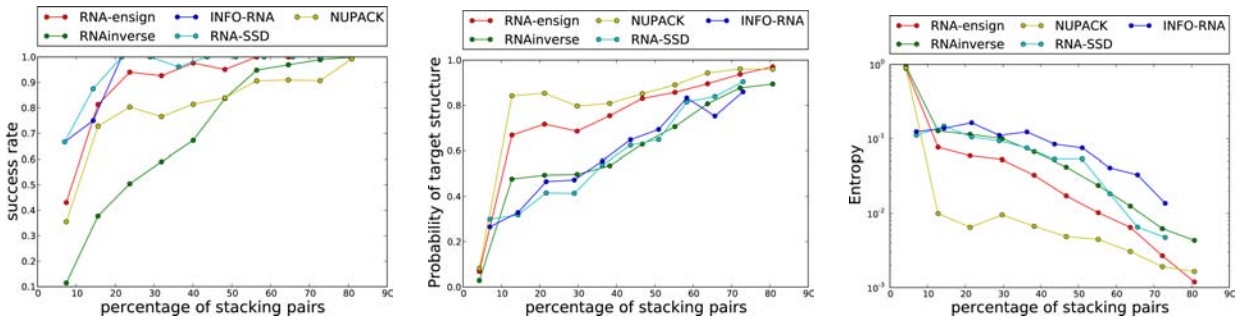


`RNA-ensign` and `RNAinverse` (same stopping criterion)

Figure 1: Evaluation of the influence of the nucleotide composition of the seeds on the base pairing entropy of `RNA-ensign` (first column), `RNAinverse` (second column) and `NUPACK` (third column). The x and y axis represent respectively the `A+G` content and `C+G` content of the sequences.

2 Impact of target structure

We provide an alternate benchmark to complete the analysis of the impact of target structures on the performance of `RNA-ensign`, `RNAinverse`, `NUPACK`, `RNA-SSD` and `INFO-RNA`. We estimate the impact of the percentage of stacking pairs on the success rate, probability of the target structure on the solution and the base pairing entropy. Instead of reporting the average performance over all runs, here we report for `RNA-ensign`, `RNAinverse` and `NUPACK` the best solution found among all seeds with all possible combination of `A+G` content and `C+G` content. Since `RNA-SSD` and `INFO-RNA` do not allow us to specify a seed, we ran them 100 times and reported their best result.



(a) Success rate (b) Target probability (c) Base pair entropy

Figure 2: Evaluation of the influence of target structure. We report the best results over all seeds for `RNA-ensign`, `RNAinverse` and `NUPACK`, and the best results over 100 runs of `RNA-SSD` and `INFO-RNA`. The x-axis represents the percentage of stacks in the target structure. On the left (a), we show how this parameter impacts the success rates of the programs. In the middle figure (b), we depict the probability of the target structure for the designed sequence. On the right (c), we show the influence on the base pairing entropy.

3 Base pairing entropy of mutants from the low-energy ensemble

This section complete the Fig. 1 of the main document. It displays the base pairing entropies of the mutants sampled from the low energy ensemble and the uniform distribution. It shows that mutants sampled from the low-energy ensemble tend have lower base pair entropies and thus less "competing" structures in the energy landscape.

The structures randomly selected from `RNA STRAND` database are:
1) "`..(((((((((...(((((((((...)))..)).))))..)))))))))`"
2) "`(((((((....((.(((.......))).)).....(((((.......)))))))))))....`"

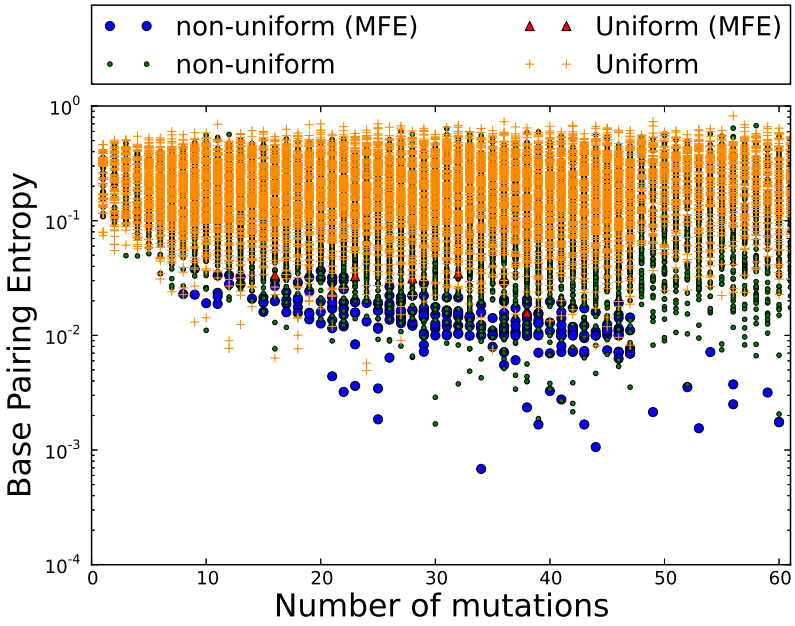


Figure 3: A scatter plot of the base pairing entropies on samples versus number of mutations from the seed. The "non-uniform" sequences (black and white circles) are generated from the low-energy ensemble, while the "uniform" sequences (triangles and crosses) are generated uniformly at random from all k-mutants consistent with the structures. The sequences satisfying the MFE criterion are indicated with a black circle (non-uniform) and a triangle (uniform). In both cases, we sampled 100 k-mutants for each k.