1 Evolving RNA populations

1.1 “RNA-World” hypothesis

• RNA is a molecule that performs catalytic, structural, and information coding functions
• Origin of life and “RNA-World” hypothesis: Early replicators relied on RNA to carry genetic information and to catalyze biochemical reactions
• Molecular evolution experiments at CAB: small, functional, synthesised RNA molecules (C. Briones) RNA viruses (E. Lázaro)

1.2 Molecular evolution

• Replication: RNA replicates with relatively high error rates (unlike DNA) → heterogeneous population (quasispecies)
• Mutation acts on the sequence
• Selection acts on the function, i.e., on the structure, NOT on the sequence!

1.3 Evolution algorithm: evolution towards a target structure

• take a population of $N$ RNA molecules of length $n$
• fix a target structure, e.g., hairpin or hammerhead

• perform realistic folding for each molecule sequence (RNAfold, Vienna Package)
• compare each secondary structure with the target structure (→ distance $d$, e.g., base-pair distance)
• sequences with a structure close to the target structure are more likely to replicate (selection $\propto \exp(-\beta d)$, selective pressure $\beta$)
• selection on structures
• mutation is random and acts on the single nt (rate $\rho$)

Numerical details:
• size of molecules $n = 12$nt to $n = 70$nt, usually $n = 37$nt
• population size: $N = 150$ to $N = 6000$, usually $N = 602$
• protocol: vary mutation rate $\rho$ for fixed selective pressure $\beta$
• $25 - 200$ realizations, $10^3 - 10^4$ replication generations

1.4 Temporal development

1.5 First appearance and fixation

• for large mutation rates, target structure is not fixed
• persistence of advantageous mutations vs. generation of diversity

1.6 Consensus structure and structure of consensus sequence

• Consensus structure obtained in a similar way as consensus sequence

1.7 Fixation as a collective phenomenon

1.8 Consensus structure below the error threshold

• Structural robustness (here for hairpin structure)

1.9 Consensus structure above the error threshold

• First lost bonds correspond to thermodynamically unstable pairs

1.10 Summary

• Population of RNA molecules is a well-suited system to study evolutionary processes, relevant for understanding, e.g., origin of life, adaption of viruses in changing environments
• Time of fixation of the target structure is optimized relatively far from the error threshold
• Structural robustness: even for large mutation rates, the population contains signatures of the target structure, as revealed by consensus structure and the structure of the consensus sequence


2 Structural repertoire of a random pool of RNA molecules

• $10^5$ molecules of size $n = 37$, $25\%$ probability for A, C, G, U
• folding by RNAfold (Vienna package)

2.1 Classification of secondary structures

2.2 Complete results for $10^5$ molecules

2.3 Frequency-rank distribution

2.4 Structure families distribution

2.5 Nucleotide composition

2.6 Summary

• We give additional evidence that random sequences fold preferably in simple structures such as stem-loops or hairpins and support the hypothesis that these are building blocks of functional RNA structures
• We observe a strong correlation of the rank and nucleotide composition with the structure family, having implications for the design of RNA structures from pools