New alphabet–dependent morphological transition in a random RNA alignment

O.V. Valba[a,b], M.V Tamm[c] and S.K Nechaev[b,d]

[a] *Moscow Institute of Physics and Technology, 141700, Dolgoprudny, Russia,*
[b] *LPTMS, Université Paris Sud, 91405 Orsay Cedex, France,*
[c] *Physics Department, Moscow State University, 119992, Moscow, Russia, E–mail: tamm@polly.phys.msu.ru,*
[d] *P.N. Lebedev Physical Institute of the Russian Academy of Sciences, 119991, Moscow, Russia*

We study the fraction f of nucleotides involved in the formation of a cactus–like secondary structure of random heteropolymer RNA–like molecules as a function of the number c of different nucleotide species. We show that in the low–temperature limit there exist two distinct morphologies of the secondary structures depending on c. For small values of c, $c \le c_{\rm cr}$ $f(c)$ converges to unity as the chain length n goes to infinity, signaling the formation of a virtually "perfect" gapless secondary structure; while for $c > c_{\rm cr}$ $f(c) < 1$, which means that a non-perfect structure with gaps is formed. The transition between two morphologies is thus of the type of solvability–unsolvability type, wll studied in computer science, with solvability in this case being interpreted as the existence of a perfect structure.

The strict upper and lower bounds $2 \le c_{\rm cr} \le 4$ are proven, and a possible generalization of the model to the non-integer values of c is discussed, as well as the numerical evidence for the generalized model presented.

Also, we argue in favor of a possible relevance of the transition discovered from the evolutionary point of view. Namely, we point out that RNAs with $c \le c_{\rm cr}$ cannot form predictable secondary structures, while the secondary structures of RNAs with $c \gg c_{\rm cr}$ are unique but rather unstable; therefore, one expects that alphabets with number of letters slightly more than critical to be the best for the biological purposes.