The increasing usability of synthetic genetic constructs ranging from biotechnology and gene therapy strategies to its role as an alternative information carrier increases the importance of DNA sequence adaptation to environmental factors and genetic machinery of the host. We suggest an improved reverse translation algorithm that optimizes the genetic sequence encoding a given protein to the host organism's biochemical features which significantly depend on the biophysical properties of their living environment. The optimization strategy is based on the adjustment of the local GC content variations according to a recently suggested superstatistical model which represents a long DNA molecule by a series of consecutive ∼150 bp DNA segments where nucleotides are allocated randomly, corresponding to the local equilibrium scenario, while the fractions of different nucleotides in consecutive segments alternate in a long-range correlated manner. To adjust the synthetic patch to the neighboring DNA fragments, the local GC content variability in the vicinity of the patching point is also taken into account and could be solved by adding linker sequence(s) either up- or downstream of the patch, or both. In case when the synthetic DNA patch has to encode functional genes on both strands the optimization problem could be solved only iteratively because each of directions from 5′ to 3′ and from 3′ to 5′ should be adjusted separately and thus affect each other. We tested our algorithm computationally using 130 complete genome sequences of both pro- and eukaryotic organisms. We believe that the suggested algorithm could contribute to the further improvement of stability and the adaptation abilities of long synthetic DNA patches in various heterologous systems.

We thank the Ministry of Education and Science of the Russian Federation for the financial support of this work.