TOWARD PREDICTING OF ARS ELEMENTS BY COMPUTER ANALYSIS OF THE DNA SEQUENCES

V. N. Konyshev, V. V. Lunyak, L. A. Noskin

We worked out the program for IBM-compatible computer which demonstrates combined distribution of several nucleotide patterns for lower and upper strands of a DNA (existing as ASCII file in computer). The program has a user-friendly interface and ensures: (1) visualization of the gene segments on display and line-printer; (2) modifying the display scale when necessary; (3) showing an actual position of the pattern related to the start of the analyzed gene or to the other sequences; (4) marking DNA-sequences on presence of introns, repeats, CDC zones in the gene by using information of GenBANK. The length (in bytes) of the gene to be analyzed and of patterns to be searched is not limited practically. Our program may be used for handling data from any standard GenBANKs after minimal modifications.

Applying above computer program we explored the possibility of predicting ARS elements using following patterns: (1) TCTCTTA and TCTATTA binding sites; (2) scaffold attachment region consensus sequence; (3) variations of yeast ARS consensus sequence; (4) binding sites CACTTAG for unknown nuclear protein, which we identified earlier for putative origin DARC146 and origin bona fide within DHFR gene. Scanning GenBANK we observed cluster distributions of these structural elements within several genes. Although there is no direct evidence for the involvement of indicated regions in the replication initiation in situ, we think that functional important region of genome could be associated in some way with a presence of the indicated functional elements. Proceed from correlation between mechanisms of DNA replication and activation of transcription, we suggest that presented program may be very convenient for search of regulatory regions in genome.