

Influence of homo-repeats on the aggregation properties of proteins from eukaryotic and bacterial proteomes and codon usage for homo-repeats in human proteome

O.V. Galzitskaya^{1*}, G.S. Novikov², N.V. Dovidchenko¹, M.Yu. Obanov¹

¹ *Institute of Protein Research, Institutskaya Street 4, Pushchino, Russia*

² *St. Petersburg Academic University, St. Petersburg, Russia*

* *e-mail: ogalzit@vega.protres.ru*

Key words: homo-repeat, codon usage, proteome, splicing site; disease

The influence of homo-repeats with lengths larger than four on the aggregation properties of proteins has been studied across 122 eukaryotic and bacterial proteomes. It has been found that proteins with homo-repeats are on average longer than in the whole database. The ability of proteins with homo-repeats to aggregate cannot be explained only by the presence of long homo-repeats in the proteins. There should be other characteristics of proteins increasing the aggregation property including such as appearance of homo-repeats in pairs in the same protein.

We have found the biases for codon usages for some amino acids in homo-repeats for human proteome and for all amino acids when the same codon is used for homo-repeats. Similar results are obtained for human proteins with homo-repeats associated with diseases. Moreover, for proteins associated with diseases (from the HraDis database), the fraction of proteins for which the same codon is used for homo-repeats is larger than for proteins from the human proteome. We are the first who demonstrated for human proteome that in some cases the splicing sites correspond to the homo-repeats and these sites more often appear at the *C*-terminal part of the homo-repeats.

Acknowledgements: This work was supported by the programs “Molecular and Cellular Biology” (01201353567) and by the Russian Science Foundation (No. 18-14-00321).