Вавиловский журнал генетики и селекции. 2018;22(4):438-444
DOI 10.18699/VJ18.380

The characteristics of miRNA binding sites in mRNA of *ZFHX3* gene and its orthologs

A.M. Kondybayeva[1], A.N. Akimniyazova[2], S.U. Kamenova[1], A.T. Ivashchenko[2]

[1] Asfendiyarov Kazakh National Medical University, Almaty, Kazakhstan
[2] Al-Farabi Kazakh National University, Almaty, Kazakhstan

Transcription factor gene *ZFHX3* is one of the candidate genes involved in stroke development. The ZFHX3 protein contains oligopeptides encoded by trinucleotide repeats (TNRs). TNR variability is considered to be one of the causes of the disease, but their biological function has not yet been established. We assume that TNRs are the binding sites of miRNA to mRNA and are involved in regulation of *ZFHX3* gene expression. The characteristics of miRNA–mRNA interaction were determined using MirTarget software. It has been shown that the first TNR in mRNA of the human *ZFHX3* gene consists of the seven consecutive miR-12-32603-3p binding encoding polyGlu. The ZFHX3 protein of human polyGlu contains 30 Glu. In the orthologous proteins of 36 animal species the length of polyGlu varied from 27 Glu to 33 Glu. Negatively charged polyGlu of the ZFHX3 transcription factor probably interacted with positive DNA-binding proteins. The following mRNA region of the *ZFHX3* gene contained the binding sites for miR-17-39416-3p, miR-5-15733-3p, miR-9-20317-3 encoding polyAla by 15 Ala lengths. In the 33 ZFHX3 orthologous proteins polyAla had the same length. The mRNA region of the human *ZFHX3* gene with binding polysite of miR-1322-3p encoded polyGln consisting of 19 Gln. In the 41 orthologs of the ZFHX3 protein the length of polyGln varied from seven Gln to 23 Gln. The binding sites of miR-2-6184-3p, miR-5-14114-5p and miR-19-43437-5p were located with overlapping nucleotides sequences, and encode polyPro. In ZFHX3 human polyPro consisted of 12 Pro. In the orthologs, polyPro contained from 10 Pro to 14 Pro. The binding sites of miR-17-39416-3p, miR-9-20317-3p, miR-1-1819-3p, miR-5-15733-3p, miR-6-17815-3p, miR-18-39953-5p, miR-2-6862-5p, miR-1260b and miR-X-48174-3p in human ZFHX3 encoded polyGly by 22 Gly length. In the 28 orthologs of ZFHX3 the length of polyGly decreased to 11 Gly. The TNR regions could simultaneously bind several miRNAs, which increased the dependence of gene expression on miRNA. The oligopeptides encoded by the binding polysites of miRNA in mRNA in the orthologous ZFHX3 proteins were flanked by conserved oligopeptides.

Keywords: miRNA; mRNA; *ZFHX3* gene; ZFHX3 oligopeptides; stroke.

HOW TO CITE THIS ARTICLE:
Kondybayeva A.M., Akimniyazova A.N., Kamenova S.U., Ivashchenko A.T. The characteristics of miRNA binding sites in mRNA of *ZFHX3* gene and its orthologs. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2018;22(4):438-444. DOI 10.18699/VJ18.380

Received 29.01.2018
Accepted for publication 29.04.2018

e-mail: a_ivashchenko@mail.ru

Характеристики полисайтов связывания миРНК с мРНК гена *ZFHX3* и его ортологов

А.М. Кондыбаева[1], А.Т. Акимниязова[2], С.У. Каменова[1], А.Т. Иващенко[2]

[1] Казахский национальный медицинский университет им. С.Д. Асфендиярова, Алматы, Казахстан
[2] Казахский национальный университет им. аль-Фараби, Алматы, Казахстан

Ген транскрипционного фактора *ZFHX3* относится к числу кандидатных генов, участвующих в развитии инсульта. В белке ZFHX3 имеются олигопептиды, кодируемые повторами тринуклеотидов (ПТН). Изменчивость ПТН считают одной из причин заболеваний, однако их биологическая функция не установлена. Мы предполагаем, что ПТН являются сайтами связывания миРНК с мРНК и участвуют в регуляции экспрессии гена *ZFHX3*. Характеристики взаимодействия миРНК с мРНК находили по программе MirTarget. Показано, что первый ПТН в мРНК гена *ZFHX3* человека состоит из семи последовательно расположенных сайтов связывания miR-12-32603-3p, кодирующих полиGlu. В белке ZFHX3 человека полиGlu содержит 30 Glu. В ортологичных белках 36 видов животных длина полиGlu изменялась от 27 до 33 Glu. Отрицательно заряженный полиGlu транскрипционного фактора ZFHX3, вероятно, взаимодействует с положительно заряженными белками, связанными с ДНК. Следующий участок мРНК гена *ZFHX3* содержит сайты связывания miR-17-39416-3p, miR-5-15733-3p, miR-9-20317-3p, которые кодируют полиAla длиной 15 Ala. В 33 ортологичных белках ZFHX3 полиAla имел одинаковую длину. Участок миРНК гена *ZFHX3* человека с полисайтом связывания miR-1322-3p кодирует полиGln, состоящий из 19 Gln. В 41 ортологе белка ZFHX3 длина полиGln изменялась от 7 до 23 Gln. Сайты связывания miR-2-6184-3p, miR-5-14114-5p и miR-19-43437-5p расположены с наложением нуклеотидных последовательностей и кодируют полиPro. В ZFHX3 человека полиPro состоял из 12 Pro. У ортологов он содержал от 10 до 14 Pro. Сайты связывания miR-17-39416-3p, miR-9-20317-3p, miR-1-1819-3p, miR-5-15733-3p, miR-6-17815-3p, miR-18-39953-5p, miR-2-6862-5p, miR-1260b и miR-X-48174-3p кодировали у ZFHX3 человека полиGly длиной 22 Gly. В 28 ортологах ZFHX3 длина полиGly уменьшалась до 11 Gly. Участки ПТН могут одновременно связывать несколько миРНК, что увеличивает зависимость экспрессии генов от миРНК. Олигопептиды, кодируемые полисайтами связывания миРНК в мРНК, в ортологичных белках ZFHX3 фланкированы консервативными олигопептидами.

Ключевые слова: миРНК; мРНК; ген *ZFHX3*; олигопептиды ZFHX3; инсульт.

Some genes have been identified in stroke and they are considered as candidates to determine various subtypes of stroke (Mineharu et al., 2006; Kurzepa et al., 2014; Inose et al., 2015; Wu et al., 2017). These include the *ZFHX3* gene encoding the transcription factor involved in the regulation of expression of many genes. Disruption in *ZFHX3* gene expression was identified in stroke, atherosclerosis and other cardiovascular diseases (Liu et al., 2014; Martin et al., 2014; Chauhan et al., 2016; Hauer et al., 2017). Gene *ZFHX3*, being a transcription factor, can manifest its function in a variety of ways, and be a cause of different stroke subtypes. The gene contains trinucleotide repeats that can participate in expression of its function (Sobczak et al., 2010). It has been shown that changes in the expression of *ZFHX3* gene and other candidate stroke genes correlate with miRNA level variation (Dhiraj et al., 2013; Ji et al., 2016; Liang, Lou, 2016; Qingfeng et al., 2016; Chen et al., 2017). MicroRNAs are effective regulators of gene expression, so it is required to establish which miRNA can regulate the expression of the *ZFHX3* gene. MicroRNA interaction with mRNA is determined by the physicochemical properties of these molecules. Unfortunately, in the publications related to the subject only a few substantiated assumptions were made that led to significant errors in determining the binding sites of miRNA in mRNA and interpreting of the obtained results. Existing programs for detecting miRNA binding sites in the mRNA of target genes, unfortunately, predict many false-positive sites (Peterson et al., 2014). To search for the binding sites, we used the MirTarget program, determining the characteristics of miRNA–mRNA interaction, including the detection of multiple miRNA binding sites (Ivashchenko et al., 2014). To improve the binding sites identification their presence in the mRNA of orthologous genes was verified (Ivashchenko et al., 2013; Atambayeva et al., 2017). The identification was necessary to be aware which organisms can be selected as model organisms in the study of stroke and other cardiovascular diseases. The conducted research will help determine the effect of miRNA on the expression of *ZFHX3* and other candidate genes involved in stroke development.

Materials and methods

The nucleotide mRNA sequences of the *ZFHX3* gene and its orthologs were taken from GenBank (http://www.ncbi.nlm.nih.gov). The following abbreviations of species names were used: *Acinonyx jubatus* – *Aju*, *Ailuropoda melanoleuca* – *Ame*, *Alligator mississippiensis* – *Ami*, *Anasplatyrhynchos* – *Apl*, etc. (Supplementary table 1)[1]. miRNAs were taken from miRBase Release 21 (http://mirbase.org) and the article (Londin et al., 2015) (Supplementary table 2).

The search for miRNA target genes was carried out in the MirTarget program (Ivashchenko et al., 2014). The program defines the origin of miRNA binding sites with mRNA, the location of sites in the 5′-untranslated region (5′UTR), the protein-coding region (CDS) and the 3′-untranslated region (3′UTR) of mRNA, the free energy of hybridization (ΔG, kJ/mole) and schemes of miRNA nucleotides interaction with mRNA. The ΔG/ΔGm (%) ratio was calculated for each site, where ΔGm is the free miRNA binding energy of a fully complementary nucleotide sequence. The miRNA binding sites were selected to have the ΔG/ΔGm ratio of more than 85 %, taking into account the miRNA length and ΔG value. The position of binding sites was indicated from the first nucleotide of 5′UTR in mRNAs. The MirTarget program took into account the interactions of miRNA nucleotides with mRNA target genes not only between adenine (A) and uracil (U), guanine (G) and cytosine (C), but also between A and C, G and U, via a single hydrogen bond. The distance between A and C was equal to the G-C, A-U, and G-U distances. The numbers of hydrogen bonds in the G-C, A-U, G-U and A-C interactions were found to be 3, 2, 1 and 1, respectively. The free binding energies of these nucleotide pairs were accepted as the same values (3 : 2 : 1 : 1). The program determines the interaction of mRNA with miRNA over its entire length and allows one unpaired nucleotide in mRNA, but not in miRNA, since it is bound in the RISC complex.

Results and discussion

The miRNA binding sites in the mRNA of the *ZFHX3* gene were detected in 5′UTR, CDS, 3′UTR (Supplementary tables 3 and 4) and the vast majority of miRNA binding sites are located in the CDS. Table presents the characteristics of miRNA interaction with the mRNAs of the human *ZFHX3* gene. Some miRNAs had a single binding site in mRNA that was located separately from the binding site of the same miRNA or together with the binding sites of other miRNAs. Some miRNAs had two or more sequential binding sites, overlapped with their nucleotide sequences, we called polysites (sequentially located binding sites of the same miRNA). When polysites were located in the CDS, they encoded a sequence of one amino acid, for example, polyGln in *HTT*, *ATXN1*, *ATXN7*, *TBP*, *AR* genes, polyGly in *EVX*, *HOX*, *HOXD13, GATA* genes, polyGlu in *FXN* gene, polyPro in *ALG13*, *DIAPH1*, *FMNL2*, *PCLO*, *SRPK2*, *ZFHX4* genes (http://www.ncbi.nlm.nih.gov) etc. If the binding sites of two or more different miRNAs were located with overlapping nucleotide sequences, this mRNA site was called a multiple site. Messenger RNAs of the *ZFHX3* gene contain both poly- and multiple sites.

None of the known programs for search of miRNA binding site has been able to search for such sites. Each of the sites included in multiple miRNA binding sites can have the same miRNA interaction characteristics with mRNA. However, there are polysites and multiple miRNA binding sites including a binding site with higher mRNA–miRNA interaction characteristics. This causes a longer residence time for the RISC complex in this binding site. The remaining sites increase the probability of catching the RISC complex. With a large number of binding sites in polysites and multiple sites, mRNA can bind two or more RISC complexes including miRNA. The identification of all miRNAs that bind to mRNA of *ZFHX3* gene is necessary to evaluate their effect on the expression of *ZFHX3* gene and to determine the dependence of its expression on the expression of other genes that are targets for these miRNAs.

The simplistic understanding of the relationship between miRNA and a target gene is widely spread and states that if the binding of miRNA to the gene mRNA is established, the problem is regarded as solved. However, for this miRNA, there may be other target genes with stronger interaction, or the mRNA of one gene can interact with several miRNAs.

[1] Supplementary tables 1–16 are available in the online version of the paper: http://www.bionet.nsc.ru/vogis/download/pict-2018-22/appx7.pdf

Characteristics of miRNA interaction in CDS mRNAs of *ZFHX3* gene

miRNA	The beginning of site, nt	ΔG, kJ/mole	ΔG/ΔGm, %	The length, nt
miR-2-8257-5p	969	–125	88	23
miR-20-23817-3p	1810	–132	86	24
miR-12-32603-3p	2062–2113 (7)	–108 ÷ –115	86–92	23
miR-6891-3p	2120	–106	93	21
miR-17-39416-3p	2978–2996 (4)	–115	87	22
miR-1-1819-3p	2982	–121	88	23
miR-5-15733-3p	2986–3004 (3)	–127 ÷ –129	86–87	24
miR-9-20317-3p	2986–3001 (4)	–127 ÷ –129	86–87	24
miR-1322	5836–5858 (7)	–85	83	19
miR-11-29856-3p	6025	–110	93	23
miR-2-6184-3p	6793–6794 (2)	–110 ÷ –117	85–90	23
miR-5-14114-5p	6793–6802 (3)	–119 ÷ –123	86–89	23
miR-19-43437-5p	6798	–110	87	23
miR-3692-5p	8410	–119	90	24
miR-1322	10262–10271 (4)	–87 ÷ –91	85–90	19
miR-9-25681-5p	10261–10264 (2)	–100 ÷ –102	85–87	21
miR-1322	10798–10823 (4)	–83 ÷ –87	81–85	19
miR-671-5p	10874	–119	90	23
miR-6779-5p	10868	–113	93	21
miR-17-39416-3p	11193–11232 (12)	–113 ÷ –121	85–92	22
miR-9-20317-3p	11201–11234 (9)	–127 ÷ –136	87–91	24
miR-1260b	11201–11228 (6)	–89 ÷ –93	86–90	19
miR-X-48174-3p	11207–11216 (2)	–121	85	24
miR-1-1819-3p	11211–11229 (2)	–119	86	23
miR-18-39953-5p	11212–11230 (3)	–123	86	23
miR-6-17815-3p	11214	–127	86	24
miR-2-6862-5p	11218	–118	86	23
miR-5-15733-3p	11234–11240 (2)	–127 ÷ –129	86–87	24
miR-3960	11461–11483 (5)	–102 ÷ –113	81–90	20
miR-5-14114-5p	11461	–121	88	23
miR-19-21199-3p	11462–11468 (2)	–134 ÷ –138	85–88	25
mir-1-2121-3p	11463	–142	91	25
miR-1-2770-3p	11463	–123	87	24
miR-19-33623-3p	11463	–136	91	24
miR-2-6184-3p	11481	–113	87	23

Note: The number of miRNAs binding sites is indicated in parentheses.

Therefore, it is necessary, at least computationally, to know which miRNAs can bind to the genomic mRNA genes. This task can be successfully solved using our program that predicts miRNA – mRNA binding sites and determines the quantitative characteristics of this interaction.

In the mRNA protein-coding region of the *ZFHX3* gene, the miR-2-8257-5p binding site is located first and encodes the PSARPPPP octapeptide (Supplementary table 5). This binding site is present in the mRNA orthologs of the *ZFHX3* gene and encodes an identical peptide or it is different by one or more amino acids. The QTYMEHHC peptide in front of it and the LREESASD peptide behind it are absolutely conserved in the ZFHX3 protein of 23 animals.

The binding site of miR-20-23817-3p was located at the beginning of mRNA nucleotide sequence and encoded the PAGSAAGP octapeptide in the orthologous proteins of 20 animal species (Supplementary table 6). In the orthologous proteins of other animal species, the miR-20-23817-3p binding site

included nucleotide substitutions and encoded octapeptides different from PAGSAAGP. However, the characteristics of miR-20-23817-3p interaction with mRNA of *ZFHX3* gene of these species were comparable to those for the first group of animals, which indicated the functional ability of miR-20-23817 to regulate the *ZFHX3* gene expression in these animal species. Notably, the oligopeptides MEGEEAL and EQPQAGLL adjoining to the oligopeptide encoded by the miR-20-23817-3p binding sites are absolutely conserved in the ZFHX3 protein of studied animal species.

For miR-12-32603-3p seven binding sites were identified in the mRNA of the *ZFHX3* gene (see Table). They encoded polyGlu replacing two Glu positions with Ala (Supplementary table 7). The presence of polysite for miRNA binding site promoted capture of the RISC complex and its further advantageous finding in the position of the site with the maximum free energy of mRNA–miRNA interactions. The site for miR-12-32603-3p was located in the region from 2080 nt for which ΔG and ΔG/ΔGm values were maximum: –115 kJ/mole and 92 %, respectively. MicroR-6891-3p binding site encodes the EEEEDE hexapeptide. It should be noted that in the 2D structure of the encoding polyGlu sites there was no intramolecular interaction of nucleotides and the entire polysite nucleotide sequence was capable of binding several RISC complexes. The mRNA region adjacent to the miRNA binding polysites from the 3′-end encoded D(E)EGCKGLF oligopeptide, and the region from the 5′-end encoded FSEKA(V)EPA oligopeptide (see Supplementary table 7). MicroR-12-32603-3p binding sites encodes from 27 to 33 amino acids in the ZFHX3 orthologous proteins. The functional role of polyGlu in the ZFHX3 protein has not been described in the literature. It is possible that hydrophilic polyGlu imparts high solubility to the transcription factor ZFHX3 and ensures its interaction with positively charged DNA-binding proteins, freeing up DNA for transcription. It means that the region of miR-12-32603-3p binding sites performs two functions: a) provides suppression of the synthesis of the transcription factor ZFHX3 by miRNA and as a result blocks the expression of ZFHX3-dependent genes, b) without or with a lower concentration of miR-12-32603-3p, compared to the concentration of the mRNA of *ZFHX3* gene, the polysite encodes a polyGlu necessarily for the functioning of the transcription factor. The synonymic codons of miR-12-32603-3p binding sites in mRNA of *Hsa*, *Mmu* and *Hgl* are used in the $GAG_{19}GAA_8GCG_2$, $GAG_{18}GAA_9GCG_1GCA_1$ and $GAG_{20}GAA_9GCG_1GCA_1$ ratios, respectively. In the third position of the codons for Glu and Ala, guanine is used.

The next region of mRNA gene containing the binding sites of miR-17-39416-3p, miR-5-15733-3p, miR-9-20317-3p, was located in the region from 2978 nt to 3027 nt (see Table). From the 5′-end to the multiple binding sites adjoins the site of mRNA region, encoding the absolutely conservative GGEQVFSH octapeptide (Supplementary table 8). From the 3′-end of the binding sites, a less homologous oligopeptide was encoded. The binding sites of miR-17-39416-3p encoded the TAGAAAAA, GAAAAAVA, AAAAVAAA and AAVAAAAA oligopeptides. The binding sites of miR-5-15733-3p encoded the AAAAAVAA, AAAVAAAA and AAAAAAAN octapeptides. The binding sites of miR-9-20317-3p encoded the AAAAAVAA, AAAVAAA, AVAAAAA and VAAAAAAA oligopeptides. Therefore, the mRNA region of *ZFHX3* gene, with overlapping multiple sites, was a compact target for three miRNAs that competed with each other for the binding site. The highest value of free binding energy (ΔG = –129 kJ/mole) was in miR-5-15733-3p and miR-9-20317-3p, which indicates a significant effect on the mRNA translation of the *ZFHX3* gene with an equal concentration of miRNA and mRNA. The ZFHX3 protein regions encoded by the binding sites of miRNAs studied above were identical in mice and rats and differed a little from those in humans (see Supplementary table 8). While the functional role of polyAla in the ZFHX3 protein remains unknown, the 2D structure of the encoding polynucleotide in mRNA makes it easy to bind miRNA due to the small number of intramolecular interactions of the nucleotides of this mRNA region. Synonymic codons miR-12-32603-3p binding sites in mRNA of *Cca* are $GCG_6GCA_6GCC_1GCT_1$, *Hsa* – $GCG_9GCA_1GCC_1GCT_1$, *Mmu* – $GCG_2GCA_3GCC_7GCT_1$, *Ppa* – $GCG_{11}GCA_1GCC_1GCT_1$, *Rno* – $GCG_2GCA_4GCC_6GCT_1$.

Another mRNA region with the miR-1322 binding polysite encoded polyGln (Supplementary table 9). The first binding site encoded the RQQQQQ hexapeptide. In humans polyGln consisted of 19 Gln, and in rats and mice – from 14 and 13 Gln, respectively, which could affect the expression degree of the polyGln function. The miR-1322 binding sites in the mRNA of the *ZFHX3* gene of all studied animal species were located between the nucleotide sequences encoding the conservative LADMIAS (from the 5′-end) and AQTLAQAQ (from the 3′-end) oligopeptides. Only in Xtr polyGln Gln changed to Leu. The function of polyGln remains unknown. We assume that the polysite binds to miR-1322, so the encoded polyGln of ZFHX3 transcription factor can interact with negatively charged regions of DNA-binding proteins. Another our assumption is that the polyGln variability of ZFHX3 protein is associated with the development of several diseases, but the way it happens remains unknown (Liu et al., 2014; Martin et al., 2014; Chauhan et al., 2016; Hauer et al., 2017). Synonymic codons miR-12-32603-3p binding sites in mRNA of *Cfa* are CAA_6CAG_{21}, *Cfe* – CAA_2CAG_{20}, *Hsa* – $CAA_{13}CAG_6$, *Phu* – CAA_5CAG_{18}, *Ssc* – CAA_7CAG_{14}, *Tch* – CAA_5CAG_{14}.

The binding site of miR-11-29856-3p encoded the absolutely conservative heptapeptide TETLLQL, presented in the protein of 55 species of animals (Supplementary table 10). Absolutely conservative heptapeptide LLPHFPMT adjoined the N-terminus of the heptapeptide, and the variable region of the ZFHX3 protein – the C-terminus. The high conservation of miR-11-29856-3p binding site suggested the important role of this miRNA in regulating the expression of ZFHX3 gene even in the early stages of animal evolution since its binding site is present in *Xtr*.

Supplementary table 11 presents the variability data of multiple binding sites miR-2-6184-3p, miR-5-14114-5p and miR-19-43437-5p encoding the PPPPPPP heptapeptide for an equal length. While the 5′-end of the multiple sites encoded absolutely conserved octapeptide PLRPQTPE, the LPAAPPQP octapeptide was located at the 3′-end. As far as stroke markers are concerned, it is possible to assume there is an association between miR-5-14114-5p and miR-2-6184-3p with the target gene *ZFHX3* whose binding sites have the ΔG values of –123 kJ/mole and –117 kJ/mole, respectively. The functional value of polyPro is unknown. Synonymic codons

of miR-2-6184-3p, miR-5-14114-5p and miR-19-43437-5p binding sites in mRNA of *Hsa* were $CCA_6CCU_3CCC_3$, *Mmu* – $CCA_2CCU_4CCC_3CCG_3$, *Rno* – $CCA_4CCU_2CCC_5CCG_1$, *Xtr* – $CCA_4CCU_8CCC_1$.

The absolute conservativeness of the SNPLLASQ octapeptide encoded the miR-3692-5p binding site in the mRNA of *ZFHX3* gene of 57 animal species (Supplementary table 12). The 5′-end of the binding site encoded the absolutely conserved PLRPQTPE octapeptide. From the 3′-end of the SNPLLASQ octapeptide, the highly conserved octapeptide LLSGAIPQ was located in most animal species. With ΔG = –119 kJ/mole and the ΔG/ΔGm value of 90 %, the association of miR-3692-5p with the target gene *ZFHX3* could serve as a stroke marker. MicroR-3692-5p was co-expressed with the gene of the ZDHHC14 transcription factor, which is expressed in many tissues as an oncosupressor (Yeste-Velasco et al., 2014).

In the region from 10798 nt to 10890 nt, three binding sites of miR-1322-3p separated by codons into a RQL tripeptide and a KV dipeptide were located in the human gene mRNA (Supplementary table 13). In *ZFHX3* gene mRNA of other animal species, these miR-1322-3p binding sites differed significantly in the number of codons and in the number of encoded Gln, accordingly. For example, in the rat *ZFHX3* gene the first polysite encoded 15 Gln, the second polysite encoded 8 Gln and the third polysite – again 15 Gln. Therefore, the rat could not be an adequate model for studying diseases in which polyGln is involved in the ZFHX3 protein. The longest polypeptide of 33 Gln was encoded by the second binding site in the *ZFHX3 Pma* mRNA.

The miR-6779-5p and miR-671-5p binding sites of the human *ZFHX3* gene mRNA overlapped with nucleotide sequences encoding the QTPVPP and PVPPGAP, oligopeptides, respectively (Supplementary table 14). Substitution of Pro for Ser, Gln, Thr or Ala in different animal species was a reflection of changes in one nucleotide in the binding site. However, the energy of miR-6779-5p and miR-671-5p interaction with the *ZFHX3* gene mRNA varied insignificantly. QTPVPPGAP-bound oligopeptides QQPKAS and SPDKDPAK were highly conserved in the vast majority of species, so they probably perform a highly significant function in the ZFHX3 protein.

Supplementary table 15 shows the variability of the amino acids encoded by the binding sites of nine miRNAs in the theorthologous *ZFHX3* gene mRNA. The mRNA region located from 11193 nt to 11264 nt overlapping of the nucleotide sequences of the nine miRNA binding sites encoded polyGly with one substitution of Gly by Ser. The twelve miR-17-39416-3p binding sites encoded the GGGGGGSG, GGGGSGGG, GGSGGGGG, GSGGGGGG, SGGGGGGG octapeptides and the seven GGGGGGGG octapeptides. The binding sites in positions between 11223 nt and 11232 nt had the greatest free binding energy ΔG equal to –121 kJ/mole and the ΔG/ΔGm value of 92 %. Nine miR-9-20317-3p binding sites in the mRNA of the human *ZFHX3* gene encoded the GGGSGGGG oligopeptide, seven GGGGGGGG oligopeptides and a GGGGGGGS oligopeptide. The two miR-1-1819-3p binding sites and one miR-6-17815-3p binding site encoded GGGGGGG. The miR-18-39953-5p and miR-2-6862-5p binding sites encoded one of the GGGGGGG heptapeptide. MicroR-1260b had six binding sites, encoding one GGSGGG and five GGGGGG hexapeptides. miR-5-15733-3p binding sites encodes GGGGGGGS and GGGGGSYH. Two binding sites of miR-X-48174-3p encoded oligopeptides GSGGGGGG and GGGGGGGG. Synonymic codons of miRNA binding sites in mRNA of *Cca*, *Nga* and *Lve* were used in the ratios $GGC_{15}GGU_3$, $GGC_{16}GGU_1$ and $GGC_{13}GGU_3$, respectively. In the third position of synonymous codons, only pyrimidines were used.

As it has been mentioned above, the miR-9-20317-3p binding site encoded the AAAAAAA oligopeptide (see Supplementary table 8). In the mRNA region from 11193 nt to 11264 nt the miR-9-20317-3p binding site encoded the GGGGGGGG oligopeptide (see Supplementary table 15). This explains why, for example, the nucleotide sequence gcggcggcggcggcggcggcg can be translated into three reading frames, where the first reading frame corresponds to the AAAAAAA oligopeptide, the second reading frame – to the RRRRRR oligopeptide and the third reading frame – to the GGGGGG oligopeptide. For this reason, the miR-5-15733-3p and miR-17-39416-3p binding sites in the region encoded polyAla, and in the mRNA region from 11193 nt to 11264 nt – polyGly. The assumption that a single nucleotide sequence of binding polysites can encode the oligopeptide in the three reading frames has been confirmed in the study of miR-1322-3p binding with the mRNA of 48 human genes that demonstrated the way of polyGln, polySer and polyAla were encoded by miR-1322-3p (Niyazova et al., 2015; Atambayeva et al., 2017). Synonymic codons of the miRNA binding sites in the mRNAs of *Hsa*, *Csa* and *Pab* were used in the ratios $GGC_{18}GGU_2AGU_1$, $GGC_{15}GGU_5AGU_1$ and $GGC_{16}GGU_2AGU_1$, respectively. In the third position of binding sites synonymic codons only pyrimidines were used. In *Cca*, *Nga* and *Lve* mRNAs were no Ser among polyGly, and synonymic codons kept the advantageous use of GGC codons: $GGC_{15}GGU_3$, $GGC_{16}GGU_1$, $GGC_{13}GGU_3$. In the third position of synonymous codons, only pyrimidines are used.

From 11461 nt to 11503 nt seven miRNA had one to five mRNA binding sites (see Table). Supplementary table 16 shows the variability of the amino acids encoded by the binding sites of these miRNAs. Two miR-19-21199-3p binding sites encoded the PPPPSAAAP and PPSAAAPSS nonapeptides. The miR-1-2121-3p and miR-19-33623-3p binding sites encoded the same PPPPSAAAP nonapeptide. One binding site of miR-5-14114-5p encoded the PPPPSAAA octapeptide. The miR-1-2770-3p and miR-2-6184-3p binding sites encoded the PPPPSAAA and AAPSSASP octapeptides. Three miR-3960 binding sites encoded heptapeptides PPPPSAA, PPPSAAA and PPSAAAP. The binding sites of seven miRNAs in rat and mouse mRNAs were not different from those in humans; only an Ala codon was converted into a Ser one, which had little effect on the energy of miRNA interaction with mRNA in this region, so rats and mice can serve as adequate models for studying the effect of these miRNAs on the *ZFHX3* gene expression.

The transcription factor ZFHX3 gene is unique for several reasons. Its mRNA contains binding sites for 26 miRNAs, so its expression depends on several host genes that are co-expressed with their miRNA (Hoeppner et al., 2009; Ma et al., 2011; He et al., 2012; Li et al., 2015), or are dependent on the expression of miRNA genes that are located between protein-coding genes.

For some miRNAs in mRNA there are polysites encoding oligopeptides: polyGlu, polyAla, polyPro, polyGly, polyGln.

A mRNA contains several regions that include multiple binding sites for several miRNAs with overlapped nucleotide sequences. Most of the studied binding sites are located in the protein-coding region, which allows us to assume the function of the encoded oligopeptides. Changes in codon usage frequencies and their location in binding sites encoding oligopeptides do not significantly affect the energy of miRNA interaction with the *ZFHX3* gene mRNA. For three miRNAs there are two homologous sites separated by 8215 nt, containing miRNA binding sites. These regions encoded oligopeptides in two different reading frames corresponding to the binding sites of these miRNAs. These duplicating regions of miRNA binding sites increase the control of these miRNAs for the *ZFHX3* gene expression, which indicates that miRNA binding sites are "indifferent" to what they encode. In some cases, the encoded oligopeptides can perform the necessary function for the ZFHX3 protein. For example, polyGlu oligopeptide can interact with positively charged DNA-binding proteins.

When selecting miRNA and mRNA associations as a marker of the disease one should meet the folowing conditions: (1) high binding energy of miRNA with mRNA; (2) high ΔG/ΔGm value; (3) the presence of multiple binding sites in mRNA for miRNA; (4) certain degree of conservatism of miRNA binding sites in orthologous gene mRNAs. On these grounds, it is proposed to use the miR-20-23817, miR-9-20317-3p, miR-19-21199-3p, miR-19-33623-3p and miR-1-2121-3p associations with the *ZFHX3* gene mRNAs as markers for the diagnosis of stroke and other diseases for which *ZFHX3* gene is the candidate gene causing their development. These miRNAs were associated with mRNA of gene with a free energy ΔG value equal to –130 kJ/mole and more.

In the experiments establishing the role of miRNA in the regulation of gene expression, it is necessary to consider the following circumstances, which in many publications are not taken into account. MicroRNAs in a complex with RISC bind to mRNA and manifest themselves as inhibitors of translation, therefore the description of their action used in biochemistry is applicable.

The miRNA interaction mRNA is determined by the physicochemical properties of these molecules. Unfortunately, in the study of this interaction, little-grounded assumptions were made that led to significant errors in the determination of miRNA binding sites in mRNA and interpretation of the obtained results. Existing programs for identification of miRNA binding sites in mRNA of target genes, unfortunately, predict many false-positive sites (Peterson et al., 2014). To increase the reliability of the predicted binding sites, we have studied orthologous genes of different animal species (Atambayeva et al., 2017).

Conclusion

The mRNA of the *ZFHX3* gene contains binding sites for 26 miRNAs. This fact indicates a strong dependence of the *ZFHX3* gene expression on miRNA. The binding sites of these miRNAs are found in the protein-coding region of the *ZFHX3* gene mRNA and encode the oligopeptides located between the conservative flanking ZFHX3 protein oligopeptides. Identical nucleotide sequences of binding sites of one miRNA located in different locations of mRNA encode different oligopeptides in different reading frames. The mRNA of the *ZFHX3* gene contains consecutively located binding sites for one or more miRNAs that encode oligopeptides.

The identified miRNAs can be used as a marker of disease development involving *ZFHX3* gene. The impact effectiveness of each miRNA on the expression of *ZFHX3* gene depends on the free energy of their interaction and on the ratio of miRNA concentrations, since the miRNA binding sites (RISC) coincide or overlap. In addition, the ratio of miRNA and mRNA of *ZFHX3* gene concentrations is very important. The prolonging length of the miRNA binding polysites increases the probability of the interaction of one or more miRNAs with mRNA of *ZFHX3*. The detection of miRNA binding sites in mRNA of paralogous genes allows to determine the adequacy of experimental animals for studying the diseases caused by the effect of miRNA.

Acknowledgments

We would like to thank our study individuals for their generous participation in this study. This study was supported by a grant from the Ministry of Education and Science, Kazakhstan Republic, SRI of Biology and Biotechnology Problems, al-Farabi Kazakh National University.

Conflict of interest

The authors declare to have no conflict of interest.

References

Atambayeva S., Niyazova R., Ivashchenko A., Pyrkova A., Pinsky I., Akimniyazova A., Labeit S. The binding sites of miR-619-5p in the mRNAs of human and orthologous genes. BMC Genomics. 2017; 18(1):1-10. DOI 10.1186/s12864-017-3811-6.

Chauhan G., Debette S. Genetic risk factors for ischemic and hemorrhagic stroke. Curr. Cardiol. Rep. 2016;18(12):1-11. DOI 10.1007/s11886-016-0804-z.

Chen Y., Song Y., Huang J., Qu M., Zhang Y., Geng J., Zhang Z., Liu J., Yang G.Y. Increased circulating exosomal miRNA-223 is associated with acute ischemic stroke. Front. Neurol. 2017;8:57. DOI 10.3389/fneur.2017.00057.

Dhiraj D., Chrysanthou E., Mallucci G., Bushell M. miRNAs-19b, -29b-2* and -339-5p show an early and sustained up-regulation in ischemic models of stroke. PLoS One. 2013;8(12):e83717. DOI 10.1371/journal.pone.0083717.

Hauer A., Pulit S., van den Berg L., de Bakker P., Veldink J., Ruigrok Y. A replication study of genetic risk loci for ischemic stroke in a Dutch population: a case-control study. Sci. Rep. 2017;7(1):1-5. DOI 10.1038/s41598-017-07404-4.

He C., Li Z., Chen P., Huang H., Hurst L., Chen J. Young intragenic miRNAs are less coexpressed with host genes than old ones: Implications of miRNA host gene coevolution. Nucleic Acids Res. 2012; 40(9):4002-4012. DOI 10.1093/nar/gkr1312.

Hoeppner M., White S., Jeffares D., Poole A. Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. Genome Biol. Evol. 2009;1:420-428. DOI 10.1093/gbe/evp045.

Inose Y., Kato Y., Kitagawa K., Uchiyama S., Shibata N. Activated microglia in ischemic stroke penumbra upregulate MCP-1 and CCR2 expression in response to lysophosphatidylcholine derived from adjacent neurons and astrocytes. Neuropathology. 2015;35(3):209-223. DOI 10.1111/neup.12182.

Ivashchenko A., Berillo O., Pyrkova A., Niyazova R., Atambayeva Sh. MiR-3960 binding sites with mRNA of human genes. Bioinformation. 2014;10(7):423-427. DOI 10.6026/97320630010423.

Ivashchenko A., Issabekova A., Berillo O. miR-1279, miR-548j, miR-548m, and miR-548d-5p binding sites in CDSs of paralogous and

orthologous *PTPN12, MSH6,* and *ZEB1* genes. Biomed. Res. Int. 2013;2013:1-11. DOI 10.1155/2013/902467.

Ji Q., Ji Y., Peng J., Zhou X., Chen X., Zhao H., Xu T., Chen L., Xu Y. Increased brain-specific MiR-9 and MiR-124 in the serum exosomes of acute ischemic stroke patients. PLoS One. 2016;11(9):1-14. DOI 10.1371/journal.pone.0163645.

Kurzepa J., Kurzepa J., Golab P., Czerska S., Bielewicz J. The significance of matrix metalloproteinase (MMP)-2 and MMP-9 in the ischemic stroke. Int. J. Neurosci. 2014;124(10):707-716. DOI 10.3109/00207454.2013.872102.

Li N., Yang J., Cui L., Ma N., Zhang L., Hao L. Expression of intronic miRNAs and their host gene *Igf2* in a murine unilateral ureteral obstruction model. Braz J. Med. Biol. Res. 2015;48(6):486-492. DOI 10.1590/1414-431X20143958.

Liang T.Y., Lou J.Y. Increased expression of mir-34a-5p and clinical association in acute ischemic stroke patients and in a rat model. Med. Sci. Monit. 2016;22:2950-2955. DOI 10.12659/MSM.900237.

Liu Y., Ni B., Lin Y., Chen X.G., Fang Z., Zhao L., Hu Z., Zhang F. Genetic polymorphisms in *ZFHX3* are associated with atrial fibrillation in a Chinese Han population. PLoS One. 2014;9(7):101318. DOI 10.1371/journal.pone.0101318.

Londin E., Loher P., Telonis A.G., Quann K., Clark P., Jing Y., Hatzimichael E., Kirino Y., Honda S., Lally M., Ramratnam B., Comstock C.E.S., Knudsene K.E., Gomella L., Spaeth G.L., Hark L., Katz L.J., Witkiewicz A., Rostami A., Jimenez S.A., Hollingsworth M.A., Yeh J.J., Shaw C.A., McKenzie S.E., Bray P., Nelson P.T., Zupo S., Roosbroeck K.V., Keating M.J., Calin G.A., Yeo C., Jimbo M., Cozzitorto J., Brody J.R., Delgrosso K., Mattick J.S., Fortina P., Rigoutsos I. Analysis of 13 cell types reveals evidence for the expression of numerous novel primate- and tissue-specific microRNAs. Proc. Natl. Acad. Sci. USA. 2015;112(10):1106-1115. DOI 10.1073/pnas.1420955112.

Ma N., Wang X., Qiao Y., Li F., Hui Y., Zou C., Jin J., Lv G., Peng Y., Wang L., Huang H., Zhou L., Zheng X., Gao X. Coexpression of an intronic microRNA and its host gene reveals a potential role for miR-483-5p as an IGF2 partner. Mol. Cell. Endocrinol. 2011;333(1):96-101. DOI 10.1016/j.mce.2010.11.027.

Martin R., Owens W., Cunnington M., Mayosi B., Koref M., Keavney B. Chromosome 16q22 variants in a region associated with cardiovascular phenotypes correlate with ZFHX3 expression in a transcript-specific manner. BMC Genet. 2014;15:1-7. DOI 10.1186/s12863-014-0136-1.

Mineharu Y., Inoue K., Inoue S., Yamada S., Nozaki K., Takenaka K., Hashimoto N., Koizumi A. Association analysis of common variants of ELN, NOS2A, APOE and ACE2 to intracranial aneurysm. Stroke. 2006;37(5):1189-1194. DOI 10.1161/01.STR.0000217408.91389.4d.

Niyazova R., Berillo O., Atambayeva Sh., Pyrkova A., Alybaeva A., Ivashchenko A. miR-1322 binding sites in paralogous and orthologous genes. Biomed. Res. Int. 2015: 962637. DOI 10.1155/2015/962637.

Peterson S., Thompson J., Ufkin M., Sathyanarayana P., Liaw L., Congdon C. Common features of microRNA target prediction tools. Front. Genet. 2014;5:1-10. DOI 10.3389/fgene.2014.00023.

Qingfeng M., Haiping Zh., Zhen T., Rongliang W., Ping L., Ziping H., Shubei M., Yumin L., Jianping J. MicroRNA-181c exacerbates brain injury in acute ischemic stroke aging and disease. Aging Dis. 2016;7(6):705-714. DOI 10.14336/AD.2016.0320.

Sobczak K., Michlewski G., de Mezer M., Kierzek E., Krol J., Olejniczak M., Kierzek R., Krzyzosiak W. Structural diversity of triplet repeat RNAs. J. Biol. Chem. 2010;285(17):12755-12764. DOI 10.1074/jbc.M109.078790.

Wu Q., Wu H., Orlov Yu.L., Gegentana G., Huo W., Bragin A.O., Wu N., Suyalatu S., Zhao F., Zhao J., Tabikhanova L.E., Chen M., Bai H. Genetic polymorphisms and related risk factors of ischemic stroke in a Mongolian population in China. Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov Journal of Genetics and Breeding. 2017;21(5):581-587. DOI 10.18699/VJ17.275.

Yeste-Velasco M., Mao X., Grose R., Kudahetti S., Lin D., Marzec J., Vasiljević N., Chaplin T., Xue L., Xu M., Foster J., Karnam S., James S., Chioni A., Gould D., Lorincz A., Oliver R., Chelala C., Thomas G., Shipley J., Mather S., Berney D., Young B., Lu Y. Identification of ZDHHC14 as a novel human tumour suppressor gene. J. Pathol. 2014;232(5):566-577. DOI 10.1002/path.4327.