

# Proteomic Analysis of Zebrafish Protein Recoding via mRNA Editing by ADAR Enzymes

Shamsudin S. Nasaev<sup>1,2</sup>, Anna S. Kopeykina<sup>3</sup>, Ksenia G. Kuznetsova<sup>4</sup>,  
Lev I. Levitsky<sup>5</sup>, and Sergei A. Moshkovskii<sup>1,3,a\*</sup>

<sup>1</sup>Federal Research and Clinical Center of Physical-Chemical Medicine, 119435 Moscow, Russia

<sup>2</sup>Institute of Biomedical Chemistry, 119121 Moscow, Russia

<sup>3</sup>Pirogov Russian National Research Medical University, 117997 Moscow, Russia

<sup>4</sup>University of Bergen, NO-5020, Bergen, Norway

<sup>5</sup>Talrose Institute for Energy Problems of Chemical Physics, Semenov Federal Research Center for Chemical Physics, Russian Academy of Sciences, 119334 Moscow, Russia

<sup>a</sup>e-mail: moshrfi@gmail.com

Received September 20, 2022

Revised October 8, 2022

Accepted October 8, 2022

**Abstract**—RNA editing by adenosine deaminases of the ADAR family can lead to protein recoding, since inosine formed from adenosine in mRNA is complementary to cytosine; the resulting codon editing might introduce amino acid substitutions into translated proteins. Proteome recoding can have functional consequences which have been described in many animals including humans. Using protein recoding database derived from publicly available transcriptome data, we identified for the first time the recoding sites in the zebrafish shotgun proteomes. Out of more than a hundred predicted recoding events, ten substitutions were found in six used datasets. Seven of them were in the AMPA glutamate receptor subunits, whose recoding has been well described, and are conserved among vertebrates. Three sites were specific for zebrafish proteins and were found in the transmembrane receptors astrotactin 1 and neuregulin 3b (proteins involved in the neuronal adhesion and signaling) and in the *rims2b* gene product (presynaptic membrane protein participating in the neurotransmitter release), respectively. Further studies are needed to elucidate the role of recoding of the said three proteins in the zebrafish.

DOI: 10.1134/S0006297922110098

**Keywords:** proteogenomics, shotgun proteomics, zebrafish, ADAR, RNA-dependent adenosine deaminase, RNA editing

## INTRODUCTION

Adenosine deaminases of the ADAR (adenosine deaminases acting on RNA) family have been found in most multicellular animals. Enzymatically active ADAR isoforms deaminate adenosine residues of ribonucleic acids at the 6th position of purine heterocycle, resulting in the adenosine conversion into non-canonical inosine. This natural post-transcriptional modification of RNA is common in animal cells [1].

The activity of ADARs depends on the presence of double-stranded (ds) segments in the RNA structure. ADARs have a domain that selectively binds dsRNA sequences. In humans and rodents, as well as some other

vertebrates, enzymatically active ADAR forms have been classified into two functional groups. In mice and humans, each of the groups includes one enzyme encoded by a single gene. Thus, human ADAR1 enzyme (*ADAR* gene) edits continuous regular stretches of dsRNA typically found in non-coding RNAs or non-coding parts of mRNAs, whereas human ADAR2 (*ADARB1* gene) is active towards shorter and irregular duplexes often located in the mRNA coding sequences [2].

How do these two types of ADAR enzymes function? The presence of double-stranded DNA or RNA in the cytoplasm elicits innate immune response even at comparatively low concentrations of these nucleic acids, since it may indicate a launch of the viral attack. The activity of ADARs towards continuous dsRNA structures neutralizes these sequences. Because inosine is more complementary to cytidine than to uracil (in contrast to

**Abbreviations:** ADAR, adenosine deaminase acting on RNA.

\* To whom correspondence should be addressed.

adenosine), dsRNA structures are destroyed after editing by ADARs. The main function of ADAR1 is neutralization of long duplexes in the nucleus and cytoplasm in response to the type I interferon-mediated response elicited by the increase in the cytoplasmic levels of dsRNA via the negative feedback mechanism. In this context, mammalian type I interferons induce a production of alternative long ADAR1 isoform that is transported to the cytoplasm, where it manifests its anti-interferon activity [3].

Unlike ADAR1, ADAR2 acts more locally. It fine-tunes the splicing of mRNAs and, in some cases, changes amino acid sequences of encoded proteins. Indeed, changes of the complementary base in mRNA can change the codon meaning if the corresponding position provides for such a substitution. Inosine residue behaves as guanosine, which results in single amino acid polymorphisms or, in rare cases, in stop codon skipping. Therefore, RNA editing by ADAR enzymes, more specifically, ADAR2 (at least in vertebrates), provides proteome recoding [3].

The extent of proteome recoding is different in various organisms. For example, in cephalopods, transcriptomes are edited and proteomes are recoded at a significant level, with several hundred amino acid substitutions found in different molluscan species [4]. *Drosophila*, which has only one ADAR enzyme (an ortholog of vertebrate ADAR2), is also characterized demonstrates an extensive RNA editing and proteome recoding: about seventy amino acid substitutions have been detected in the *Drosophila* proteome [5]. In comparison to mollusks and insects, mammals, such as primates and rodents, have a much lesser number of protein recoding events (as predicted from the transcriptomes), with only ~20 of them reliably identified in the proteome [6].

Despite the limited scale of recoding in mammals, the loss-of-function mutations in the mammalian ADAR2 genes are usually lethal. The lethal phenotype of transgenic mice with the knocked out of *Adarb1* gene (ADAR2) was reversed by a single amino acid substitution in the gene of the ionotropic AMPA glutamate receptor subunit 2 (*Gria2*). During embryonic development, residue 607 of this protein is recoded from Gln to Arg, which results in 100% recoding in adult individuals. The recoding leads to a decrease in the conductivity of the corresponding ion channel essential for the normal brain functioning. Introduction of the Q607R substitution in the receptor molecule by genome mutagenesis rescued the lethality in *Adarb1*-knockout mice [7].

Several other sites recoded by RNA editing have been functionally characterized in mice and various cells types. For example, an amino acid substitution in filamin A (*FLNA*) resulting from RNA recoding was found to affect the tone of vascular smooth muscles [8]. Evolutionary conserved Ile-to-Val substitution in coatomer subunit alpha (*COPA*) has the tumor suppressor effects and plays a role in the development of human hepatocellular carcinoma [9].

The scope of proteome recoding via RNA editing by ADARs has been typically estimated based on the transcriptome data. However, it is unclear how many editing events lead to the recoding of translated proteins. It is believed that many editing events are the side effects of enzymatic reactions and have no biological significance [10]. Apparently, the editing events that result in protein recoding and can be detected in the proteome are more likely to have the functional consequences. That is why the results of transcriptome-wide analysis of RNA editing by ADARs can be used for searching the amino acid polymorphisms in the proteomic data. Since a common approach to the identification of protein sequences is a search for matches between the mass-spectrometry data and protein sequences predicted from the genomes or transcriptomes, identification of editing events requires that the theoretical, consensus proteome should be supplemented with the recoded sequences [11].

Several reports have been published on analysis of recoding sites in the proteome level. In the pioneering work, Liscovitch-Brauer et al. [4] demonstrated extensive transcriptome rearrangements ADARs in cephalopods and confirmed the corresponding changes in the mollusk proteomes [4]. We found recoded proteins in the *Drosophila* proteome [5] and revealed that the recoding patterns were different at different metamorphosis stages [12]. We also identified recoded sites in the murine and human brain proteins [6], which were conserved between these two mammal species but did not overlap with the recoded sites in the *Drosophila* brain proteome. In another important work, Peng et al. [13] conducted a large-scale analysis of various cancer cell proteomes and identified the corresponding recoded sites [13].

In addition to the well-studied RNA editomes of cephalopods, *Drosophila*, mice, and humans, RNA editing by ADARs was recently described in the transcriptome of another well-established model species, zebrafish (*Danio rerio*) [14]. This freshwater fish of the Cypriniformes order native to the tropical Asia, has been used in laboratories since 1960s for a wide range of biomedical applications. The studies in the zebrafish have played an important role in developmental biology, neuroscience, and investigation of diseases of vertebrates [15]. The aim of this study was to reliably identify recoded protein sites in the zebrafish proteome using the transcriptome data and a workflow implemented in our earlier studies. Identification of the recoding sites at the proteome level is essential for further research, as these sites are more likely to be functionally important in comparison to the sites detected only at the transcriptome level.

## MATERIALS AND METHODS

**Development of the database with predicted sites recoded by RNA editing for the proteomic search.** As a reference, we used the zebrafish proteome available from

Uniprot (accession number UP000000437\_7955 [16]). The non-synonymous substitutions in the mRNA exons predicted from the zebrafish transcriptome were taken from Buchumenski et al. [14]. The VCF file was composed from the said substitutions, which was annotated with the snpEff program [17] using the zebrafish genome assembly, version GRCz11.99 [18]. Therefore, the recoded residues were associated with the proteomic database records. In total, substitutions in the zebrafish proteins were predicted in 116 positions; 63 of them had two alternative variants.

Using implemented in-house Python scripts, protein sequences were converted into tryptic peptides with one possible missed cleavage site. Recoded proteins were then converted into peptides using the same approach. The resulting database contained both genome-encoded and recoded peptides. At the first stage, 116 recoded sites produced 471 tryptic peptides (taking into account peptides with one missed trypsin cleavage site). Since mass-spectrometry does not allow identification of very short or very long peptides, the final list included peptides 7 to 40 amino acid residues in length. After exclusion of 139 short and long peptides, the final list of recoded peptides contained 332 species.

Using the Pyteomics program library, the decoy sequences were generated as reversed target sequences with the conservation of the C-terminal residue and added to the database [19].

**Proteomic datasets.** Six shotgun proteomic datasets of zebrafish tissues [21-25] were selected from the ProteomeXchange repository for reprocessing [20] (see Table 1 for more details).

**Proteomic search.** Original RAW files were converted into mzML files using ThermoRawFileParser [26]. The IdentiPy search engine [27] was tuned specifically for each dataset. Its configuration file was set as follows,

according to the methods used for the acquisition of mass spectra in the original works:

- PXD023967 – product accuracy: 0.01 Da, fixed: camC, variable: oxM, acetyl-
- PXD030733 – product accuracy: 0.5 Da, fixed: camC, variable: oxM, acetyl-
- PXD009612 – product accuracy: 0.01 Da, fixed: camC, variable: oxM
- PXD005630 – product accuracy: 0.5 Da, fixed: camC, variable: oxM, camM
- PXD016847 – product accuracy: 0.5 Da, fixed: camC, variable: oxM, acetyl-
- PXD014228 – product accuracy: 20 ppm, fixed: camC, variable: oxM

The peptides were identified using IdentiPy in the automatic parameter optimization mode. Scavenger post-search tool [28] was used to validate the recoded peptides and to filter them using the group specific false discovery rate (FDR 1%).

**Visualization of mass spectra.** The mass spectra of the peptides of interest were visualized with the xiSPEC spectrum viewer [29] and examined manually. The coverage of the corresponding amino acid substitution by the MS/MS fragments confirmed the validity of the peptide spectrum match.

**Annotation of recoded sites in terms of the protein spatial structure.** Amino acid substitutions were annotated using various open access resources. First, the recoded proteins were annotated using the UniProt [16] and neXtProt [30] databases. Zebrafish proteins lacking known spatial structure were aligned with the mammalian proteins using Blastp [31]. The spatial structures were analyzed with the PyMOL software [32]. The effects of the substitutions on the spatial structure stability were estimated using I-Mutant2.0 [33], MUpro [34], and iStable [35].

**Table 1.** Proteomic datasets of zebrafish tissues that were used to search for protein sites recoded via RNA editing by ADARs

ProteomeXchange repository accession number in [20]	Biomaterial analyzed	Detector for MS/MS spectrometry
PXD005630 [21]	synaptosomes and postsynaptic density from adult fish brain	LTQ (linear ion trap)
PXD009612* [22]	whole embryos on day 6 after fertilization	Orbitrap
PXD014228 [23]	adult brain	Orbitrap
PXD016847 [24]	adult brain	LTQ
PXD023967*.**	motor neurons of adult fish	Orbitrap
PXD030733 [25]	adult brain	LTQ

Note. All data were obtained by liquid chromatography/mass spectrometry of trypsin-digested proteins using Orbitrap detectors, at least, for precursor ion analysis. The accuracy of the m/z value estimation in the MS/MS spectra recorded using the linear ion trap was approximately one order of magnitude lower in comparison with those recorded using Orbitrap. The search was performed after addition of the recoding sites from Buchumenski et al. [14] to the database.

\* No recoded sites that met the selection criteria set in this work were found in these datasets.

\*\* Data were available in the repository, but not described in publications.

## RESULTS AND DISCUSSION

**Datasets and filtering of identified recoded sites.** In contrast to the data obtained for humans and mice, few datasets are available for the zebrafish proteome. Besides, these datasets should satisfy a set of criteria essential for successful implementation of the proteomics and genomics methods used in this and other works [6]. For example, for reprocessing, we used the data of the label-free shotgun proteomic analyses recorded using Orbitrap mass spectrometers (Table 1). Because of the deficit of available proteomic datasets, we analyzed the high-resolution data of tandem mass spectrometry acquired with an Orbitrap detector (as in most recent studies), as well as the data obtained using earlier hybrid machines (Orbitrap Velos or Elite) with the linear ion traps (LTQ). The MS/MS spectra obtained with the latter devices had lower mass tolerance, which was taken into account when setting the proteomic search parameters.

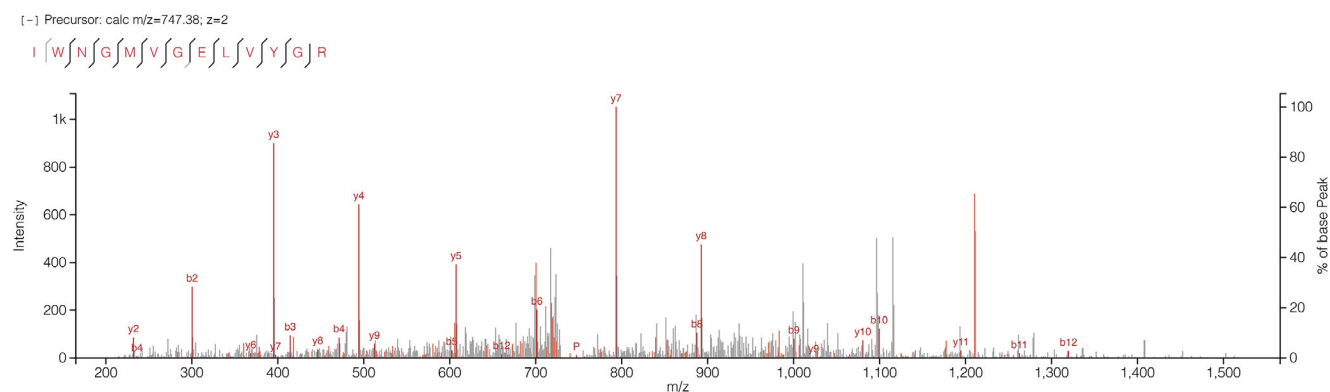
Due to the fact that zebrafish is a common model in neuroscience, most of selected datasets contained data from various brain preparations, including synaptic fraction [21] and cultured motor neurons. One of the used proteomic dataset was from 6-day-old whole zebrafish embryos [22] (Table 1). According to the existing understanding of RNA editing, the central neural system is a promising study object, since in vertebrates, most editing events followed by the protein recoding occur in the neural tissues, where the ADAR2 isoform is highly expressed [3].

The use of the datasets with the low-resolution MS/MS data prompted up to more carefully select the filtration criteria for the identification of recoded peptides. Thus, the following inclusion criteria were used: identification of the recoded site in at least in two datasets; identification in two different peptides formed by incomplete tryptic cleavage; confirmation of identification by manual examination of the mass spectra visualized with the xiSPEC viewer [29]. In addition, peptides with incomplete tryptic cleavage were excluded from analysis, if not confirmed by identification of peptides produced by complete cleavage, as these peptides ensured higher false discovery rates (FDR) [36].

A broader distribution of the  $m/z$  values in the low-resolution tandem mass spectra makes it more difficult to curate the visualized mass spectra manually. Such mass spectra obtained by the chromatography/mass spectrometry of complex mixtures usually contain multiple peaks unrelated to the target peptide. However, in some cases, the peaks of interest have a higher intensity. Thus, Fig. 1 illustrates a correspondence between the mass spectrum and the peptide of the glutamate receptor subunit *gria4b*. As one can see from Fig. 1, the most intense peaks of the mass spectrum are associated with the molecular masses theoretically predicted for the genome-encoded peptide with the K/R substitution. The rest of the mass spectra representing recoded peptides and confirmed by manual examination are shown in Fig. S1 in the Online Resource 1.

**Recoded sites identified in the zebrafish proteomes.** Recoded protein sites were found in 4 of 6 datasets processed in this work. All four datasets were generated for the central nervous system. In total, 10 recoded sites were found of 116 sites predicted from transcriptome [14] (one of the sites yielded two recoding variants; see Table 2). This number was even less than the number of recoded sites reliably identified in the central neural systems of mice and humans (14 and 18, respectively) [6]. However, the search database containing recoded sites predicted from the RNA-seq data from the zebrafish was significantly smaller. Moreover, the mammalian datasets were significantly larger in terms of the number of acquired mass spectra.

As mentioned above, the classical and the best studied example of mammalian protein recoding is substitutions in the NMDA glutamate receptor subunits. Among them, the Gln-to-Arg replacement influences the conductivity of the corresponding ion channel and is essential for the development and functioning of the central nervous system. At the same time, the existing shotgun proteomics methods, which use trypsin as a protease of choice, do not allow to detect this substitution reliably. However, the transcripts of the glutamate receptors of the same type contain other edited sites whose existence was confirmed by the analysis of mammalian proteomes [6].



**Fig. 1.** Tandem mass spectrum of the glutamate receptor subunit *gria4b* peptide visualized with the xiSPEC software [29]. The most intense peaks are associated with the molecular masses theoretically predicted for the corresponding peptide with K/R substitution.

**Table 2.** Zebrafish protein sites recoded via RNA editing by ADAR enzymes and identified in the shotgun proteomic data

Gene	Protein	Amino acid substitution resulting from recoding	Peptides identified	Confirmed by mass spectrometry	Datasets containing identified sites (for details, see Table 1)	Identification of alternative genome-encoded site
Recorded sites in glutamate receptor subunits						
<i>gria2a</i> and/or <i>gria3b</i>	glutamate receptor subunit (flop isoform)	R756G and/or R764G	GSAL[G]NAVNLA VLK, GYGVA TPKGSAL[G]NAVNLA VLK	+	PXD014228 (brain), PXD005630 (synaptosomes, postsynaptic density), PXD030733 (brain)	+
<i>gria2a</i> and/or <i>gria3b</i>	glutamate receptor subunit (flip isoform)	R756G and/or R764G	GSAL[G]TPVNLAVLK	+	PXD005630 (synaptosomes, postsynaptic density)	+
<i>gria2b</i>	glutamate receptor subunit (flip isoform)	R760G	GSSL[G]TPVNLAVLK	+	PXD005630 (synaptosomes, postsynaptic density)	+
<i>gria2b</i>	glutamate receptor subunit (flop isoform)	R760G	GSSL[G]NAVNLA VLK	+	PXD014228 (brain)	+
<i>gria4a</i>	glutamate receptor subunit (flip isoform)	R760G	GSQL[G]TPVNLAVLK	+	PXD014228 (brain), PXD005630 (synaptosomes, postsynaptic density)	+
<i>gria4b</i>	glutamate receptor subunit (flop isoform)	R766G	GSQL[G]SAVNLA VLK	+	PXD005630 (synaptosomes, postsynaptic density)	-
<i>gria4b</i>	glutamate receptor subunit	K493R	IWNGMVGELVYG[R]	+	PXD005630 (postsynaptic density)	+
Recorded sites in other proteins						
<i>astn1</i>	astrotactin 1	K935E	TLTPYAYIASLLDSGA[E]IRMAAGVR	-	PXD005630 (synaptosomes)	+
<i>astn1</i>	astrotactin 1	K935R	TLTPYAYIASLLDSGA[R]	-	PXD005630 (synaptosomes, postsynaptic density)	+
<i>nrg3b</i>	neuregulin 3b	T10A	TALGA[A]METMTLEEPGGEQASPRAPGPLR, MSERTALGA[A]METMTLEEPGGEQASPR	-	PXD005630 (synaptosomes, postsynaptic density), PXD030733 (brain)	-
<i>rims2b</i>	regulating synaptic membrane exocytosis 2b	S489G	LIGISTSYMVSQSER, SSDSVDVSDV SAMSSASRL[G]STSYMVSQSER	+	PXD016847 (brain), PXD005630 (postsynaptic density)	-



Seven of 10 recoded sites were identified in the zebrafish sequences homologous to the mammalian receptor subunits and partially identical to them. They were found in both flip and flop alternatively spliced isoforms [37] of the *gria2a*, *gria2b*, *gria4a*, and *gria4b* gene products (Table 2). These isoforms have been identified in many vertebrates [37]. Among the recoded sites, we found the R/G substitutions at positions 760–766 (depending on the specific gene), that have been also described for the mammalian proteomes. Another identified substitution was K493R in the *gria4b* gene product (the mass spectrum of the corresponding peptide is shown in Fig. 1). No recoded sites homologous to this substitution have been found in the proteomes of other species.

The homologs of other recoded sites observed in mammals were mostly absent in the zebrafish transcriptome [14], except the Q/G substitution in the *cadpsa* gene product, which is present in the editomes and recoded proteins of mice and humans. However, we did not detect this substitution in the zebrafish proteins.

It was found recently that the zebrafish editome contains many other recoding substitutions absent in mammals [14]. We were able to reliably identify three such sites in three proteins, respectively. Notably, all of them were situated in the protein fragments that could not be resolved by special structure prediction systems, such as AlphaFold 2 [38] and Swiss-model [39], since these fragments lacked sufficient similarity with experimentally resolved structures of the corresponding homologs from other organisms.

Astrotactin 1 (*astn1*) was characterized by the substitution of Lys to Arg or Glu at position 935. This transmembrane receptor regulates neuronal adhesion and participates in the migration of juvenile neuroblasts during brain development [40]. The extracellular part of this receptor contains the EGF (epidermal growth factor)-like domains and the fibronectin domain, as well as the recoded site, which is situated outside of the said domains. Supposedly, the recoded site is located in a close vicinity to the molecule fragment essential for the ligand binding, as indirectly indicated by the proximity of glycosylation sites. The recoded site of astrotactin is exposed and not covered by other structures. Interestingly, the recoded residue is more or less conserved and is also present in mammalian orthologs, where, however, it is not recoded. At the same time, the homology of the sequences surrounding the recoded residue is not enough to predict their folding based on the resolved structures. Evaluation of the effect of this substitution on the protein stability by three different methods produced conflicting results. Although lysine replacement is supposed to have little effect on the local conformation, it may modify the binding affinity toward potential ligands.

Neuregulin 3b (*neu3b*) is another neuronal transmembrane protein that participates in the development of neural tissue. Like astrotactin, it shares some features with

EGF. Thus, in mammals, neuregulin ortholog binds to a member of the EGF receptor family [41]. The Thr-to-Ala recoding was found at position 10, close to the N-terminus of the molecule. The N-terminal part of neuregulin is exposed into the extracellular space and, as often occurs in proteins, is structurally unstable. Analysis of protein stability after a single amino acid substitution with MUPro [34] and i-Stable [35] predicted a decrease in the protein structure stability after the replacement, while I-Mutant [33] reported no difference in the stability after the substitution. Since the N-terminus of neuregulin is unstable, the substitution is supposed to have no significant effect on the receptor spatial structure. However, like in astrotactin, it may still influence potential protein–protein interactions. In contrast to the recoded site in astrotactin, the N-terminus of neuregulin 3 is not conserved and has nothing in common with the mammalian orthologs.

While astrotactin 1 and neuregulin 3 have some common functions in providing adhesion and other receptor-mediated interactions in developing neurons, the role of the third recoded zebrafish protein is different. The product of the *rims2b* gene is unrelated to glutamate receptors; it is located under the presynaptic membrane and is associated with the presynaptic vesicles. It is involved in the calcium-dependent exocytosis, i.e., in the synaptic activity. The S489G substitution is in the non-conserved part of the sequence, which is absent in the mammal orthologs and, therefore, has not been spatially characterized. All three methods used for the stability prediction indicated that thus substitution most probably decreases the structural stability of the protein.

Interestingly, presynaptic membrane proteins participating in the same exocytosis pathway (including the ortholog of the *cadpsa* gene product mentioned above) are extensively recoded in the *Drosophila* nervous system [12]. It was suggested that RNA editing in this poikilothermal insect contributes to the adaptation of its nervous system to the functioning at changing ambient temperature [42]. One may hypothesize a similar function of RNA editing and protein recoding in zebrafish, which is also poikilothermic. However, the confirmation of this hypothesis requires extensive and labor-consuming *in vivo* experimentation in transgenic fish.

## CONCLUSIONS

RNA editing by ADAR enzymes is an evolutionary ancient cellular process characteristic for most Eumetazoans. During evolution, enzymatic adenosine deamination has acquired two different functions. The first one is inactivation of potentially immunogenic RNA duplexes. The second function is recoding of selected proteins by introduction of amino acid substitutions in their sequence. At first sight, the purpose of recoding seems confusing, because functional amino acid substitutions could be

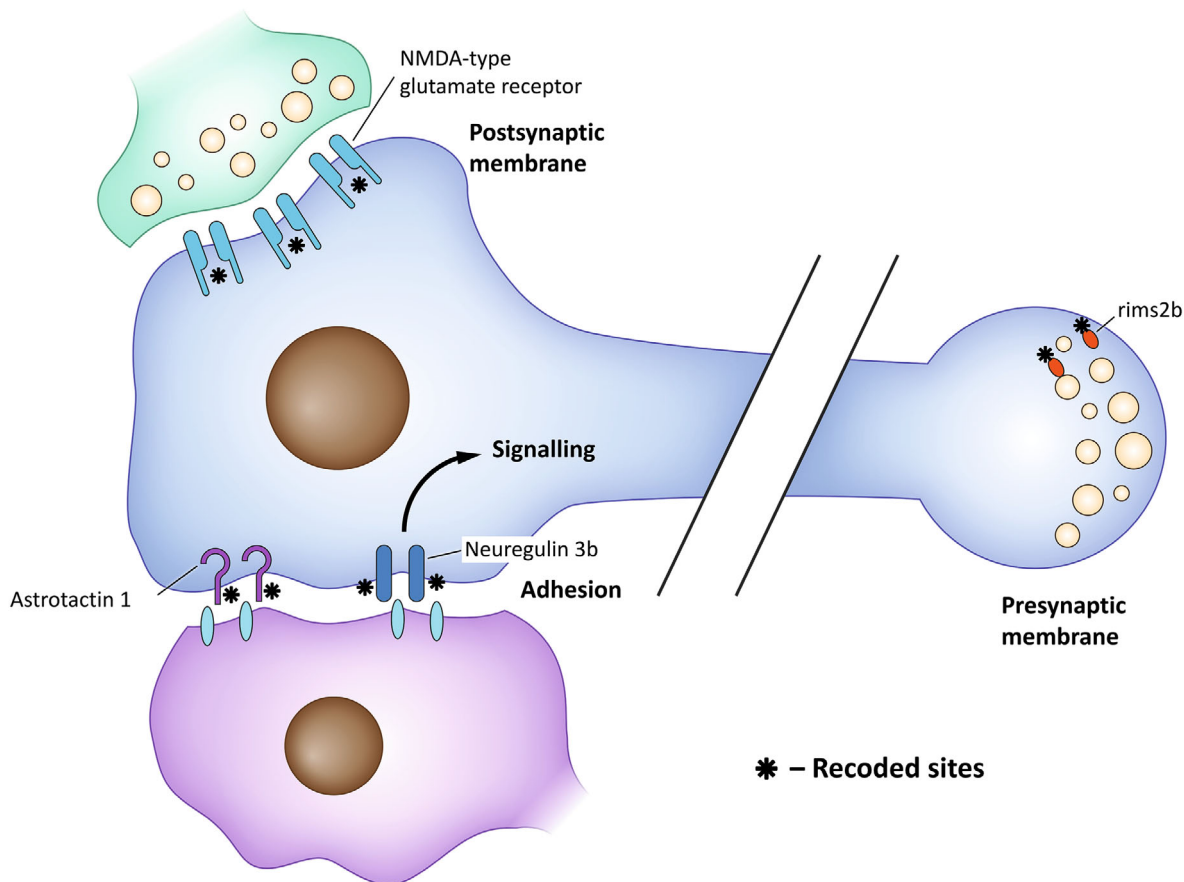


Fig. 2. Hypothesized location of recoded proteins in the zebrafish neuron.

formed by mutations in the genome, and protein functions are often modulated by multiple post-translational modifications. However, recoding as a mechanism of structural and functional modification of some proteins has been fixed in the course of evolution and, for example, is used to regulate protein function during development.

We have systematically studied protein recoding at the protein level in different species. Analysis of six proteomic datasets of the zebrafish, a model fish species, yielded as few as ten recoding sites only. According to the published data, mRNA editing in this species is less common than in mammals. At the proteome level, most recoding events were found in the glutamate receptor subunits, which should be considered as a case of conserved protein recoding via RNA editing in vertebrates. The recoded sites identified in three other proteins were specific for the zebrafish (according to existing data). Although these amino acid substitutions unlikely affect the spatial structure of the corresponding proteins, they may influence the intermolecular interactions. Out of recently described examples, a conserved and at first sight non-significant Val-to-Ile substitution of valine to isoleucine in human coatomer subunit alpha was shown to modulate the malignant potential of hepatocellular carcinoma [9]. Interestingly, all three zebrafish-specific amino acid substitutions were in the sequences less conserved in the

zebrafish than in other vertebrates. It is possible that this taxon-specific recoding is more typical for the evolutionary younger, changing protein fragments, although it is yet unknown whether this recoding is conserved among, e.g., Actinopterygii or other taxa. We believe that considerations on the evolution in the context of RNA editing and consequent protein recoding should be left for the transcriptomic studies, which provide the authors more options for the comparison between the species [43]. The hypothesized location of recoded zebrafish proteins in the neuron is shown in Fig. 2.

The continuation of this work may be an experimental verification of the hypothesis on the functional significance of recoded sites found in the adhesion and synaptic proteins. Besides, very few extensive proteomic zebrafish datasets are available for comparing with mouse or human databases, so one may continue generating those larger datasets. At the moment, the dataset for the isolated synaptic components produced most results with respect to protein recoding via RNA editing, but unfortunately, this dataset was generated using low-resolution MS/MS spectra [21]. Further generation of high-quality proteomic data from different zebrafish cells and tissues will promote our understanding of protein recoding via RNA editing by the ADAR enzymes in this important model species.

**Contributions.** S. S. Nasaev and A. S. Kopeykina created the database of recoded sites for the proteomic search, performed the proteomic search and spatial structure analysis, and processed the results; K. G. Kuznetsova and L. I. Levitsky set up the search methods for the data reprocessing and wrote the manuscript; S. A. Moshkovskii conceived the project, supervised analysis of the experimental results, and wrote the manuscript.

**Acknowledgments.** The authors thank the Core Facility of the Federal Research and Clinical Center of Physical-Chemical Medicine for providing computing resources for the study.

**Funding.** The work was supported by the Russian Science Foundation (project no. 20-15-00072).

**Ethics declarations.** The authors declare no conflict of interest. The article contains no description of studies with human or animal participants performed by any of the authors.

**Electronic supplementary material.** The online version contains supplementary material available at <https://doi.org/10.1134/S0006297922110098>.

**Open access.** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Goncharov, A. O., Shender, V. O., Kuznetsova, K. G., Kliuchnikova, A. A., and Moshkovskii, S. A. (2022) Interplay between A-to-I editing and splicing of RNA: a potential point of application for cancer therapy, *Int. J. Mol. Sci.*, **23**, 5240, doi: 10.3390/ijms23095240.
- Luciano, D. J., Mirsky, H., Vendetti, N. J., and Maas, S. (2004) RNA editing of a miRNA precursor, *RNA*, **10**, 1174-1177, doi: 10.1261/rna.7350304.
- Goncharov, A. O., Kliuchnikova, A. A., Nasaev, S. S., and Moshkovskii, S. A. (2019) RNA editing by ADAR adenosine deaminases: from molecular plasticity of neural proteins to the mechanisms of human cancer, *Biochemistry (Moscow)*, **84**, 896-904, doi: 10.1134/S0006297919080054.
- Liscovitch-Brauer, N., Alon, S., Porath, H. T., Elstein, B., Unger, R., et al. (2017) Trade-off between transcriptome plasticity and genome evolution in cephalopods, *Cell*, **169**, 191-202.e111, doi: 10.1016/j.cell.2017.03.025.
- Kuznetsova, K. G., Kliuchnikova, A. A., Ilina, I. U., Chernobrovkin, A. L., Novikova, S. E., et al. (2018) Proteogenomics of adenosine-to-inosine RNA editing in the fruit fly, *J. Proteome Res.*, **17**, 3889-3903, doi: 10.1021/acs.jproteome.8b00553.
- Levitsky, L. I., Kliuchnikova, A. A., Kuznetsova, K. G., Karpov, D. S., Ivanov, M. V., et al. (2019) Adenosine-to-inosine RNA editing in mouse and human brain proteomes, *Proteomics*, **19**, e1900195, doi: 10.1002/pmic.201900195.
- Higuchi, M., Maas, S., Single, F. N., Hartner, J., Rozov, A., et al. (2000) Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2, *Nature*, **406**, 78-81, doi: 10.1038/35017558.
- Jain, M., Weber, A., Maly, K., Manjaly, G., Deek, J., et al. (2022) A-to-I RNA editing of Filamin A regulates cellular adhesion, migration and mechanical properties, *FEBS J.*, **289**, 4580-4601, doi: 10.1111/febs.16391.
- Song, Y., An, O., Ren, X., Chan, T. H. M., Tay, D. J. T., et al. (2021) RNA editing mediates the functional switch of COPA in a novel mechanism of hepatocarcinogenesis, *J. Hepatol.*, **74**, 135-147, doi: 10.1016/j.jhep.2020.07.021.
- Buchumenski, I., Roth, S. H., Kopel, E., Katsman, E., Feiglin, A., et al. (2021) Global quantification exposes abundant low-level off-target activity by base editors, *Genome Res.*, **31**, 2354-2361, doi: 10.1101/gr.275770.121.
- Moshkovskii, S. A., Ivanov, M. V., Kuznetsova, K. G., and Gorshkov, M. V. (2018) Identification of single amino acid substitutions in proteogenomics, *Biochemistry (Moscow)*, **83**, 250-258, doi: 10.1134/S0006297918030057.
- Kliuchnikova, A. A., Goncharov, A. O., Levitsky, L. I., Pyatnitskiy, M. A., Novikova, S. E., et al. (2020) Proteome-wide analysis of ADAR-mediated messenger RNA editing during fruit fly ontogeny, *J. Proteome Res.*, **19**, 4046-4060, doi: 10.1021/acs.jproteome.0c00347.
- Peng, X., Xu, X., Wang, Y., Hawke, D. H., Yu, S., et al. (2018) A-to-I RNA editing contributes to proteomic diversity in cancer, *Cancer Cell*, **33**, 817-828 e817, doi: 10.1016/j.ccell.2018.03.026.
- Buchumenski, I., Holler, K., Appelbaum, L., Eisenberg, E., Junker, J. P., et al. (2021) Systematic identification of A-to-I RNA editing in zebrafish development and adult organs, *Nucleic Acids Res.*, **49**, 4325-4337, doi: 10.1093/nar/gkab247.
- Ribas, L., and Piferrer, F. (2014) The zebrafish (*Danio rerio*) as a model organism, with emphasis on applications for finfish aquaculture research, *Rev. Aquaculture*, **6**, 209-240, doi: 10.1111/raq.12041.
- UniProt, C. (2021) UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Res.*, **49**, D480-D489, doi: 10.1093/nar/gkaa1100.
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., et al. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w<sup>1118</sup>; iso-2; iso-3, *Fly (Austin)*, **6**, 80-92, doi: 10.4161/fly.19695.



18. Howe, K., Clark, M. D., Torroja, C. F., Tarrance, J., Bertelot, C., et al. (2013) The zebrafish reference genome sequence and its relationship to the human genome, *Nature*, **496**, 498-503, doi: 10.1038/nature12111.
19. Goloborodko, A. A., Levitsky, L. I., Ivanov, M. V., and Gorshkov, M. V. (2013) Pyteomics – a Python framework for exploratory data analysis and rapid software prototyping in proteomics, *J. Am. Soc. Mass Spectrom.*, **24**, 301-304, doi: 10.1007/s13361-012-0516-6.
20. Deutsch, E. W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J. J., et al. (2020) The ProteomeXchange consortium in 2020: enabling “big data” approaches in proteomics, *Nucleic Acids Res.*, **48**, D1145-D1152, doi: 10.1093/nar/gkz984.
21. Bayes, A., Collins, M. O., Reig-Viader, R., Gou, G., Goulding, D., et al. (2017) Evolution of complexity in the zebrafish synapse proteome, *Nat. Commun.*, **8**, 14613, doi: 10.1038/ncomms14613.
22. Watchon, M., Luu, L., Robinson, K. J., Yuan, K. C., De Luca, A., et al. (2021) Sodium valproate increases activity of the sirtuin pathway resulting in beneficial effects for spinocerebellar ataxia-3 *in vivo*, *Mol. Brain*, **14**, 128, doi: 10.1186/s13041-021-00839-x.
23. Faria, M., Prats, E., Gomez-Canela, C., Hsu, C. Y., Arick, M. A., 2nd, et al. (2019) Therapeutic potential of N-acetylcysteine in acrylamide acute neurotoxicity in adult zebrafish, *Sci. Rep.*, **9**, 16467, doi: 10.1038/s41598-019-53154-w.
24. Nonnis, S., Angiulli, E., Maffioli, E., Frabetti, F., Negri, A., et al. (2021) Acute environmental temperature variation affects brain protein expression, anxiety and explorative behaviour in adult zebrafish, *Sci. Rep.*, **11**, 2521, doi: 10.1038/s41598-021-81804-5.
25. Maffioli, E., Angiulli, E., Nonnis, S., Grassi Scalvini, F., Negri, A., et al. (2022) Brain proteome and behavioural analysis in wild type, BDNF<sup>+/−</sup> and BDNF<sup>−/−</sup> adult Zebrafish (*Danio rerio*) exposed to two different temperatures, *Int. J. Mol. Sci.*, **23**, 5606, doi: 10.3390/ijms23105606.
26. Hulstaert, N., Shofstahl, J., Sachsenberg, T., Walzer, M., Barsnes, H., et al. (2020) ThermoRawFileParser: modular, scalable, and cross-platform RAW file conversion, *J. Proteome Res.*, **19**, 537-542, doi: 10.1021/acs.jproteome.9b00328.
27. Levitsky, L. I., Ivanov, M. V., Lobas, A. A., Bubis, J. A., Tarasova, I. A., et al. (2018) IdentiPy: an extensible search engine for protein identification in shotgun proteomics, *J. Proteome Res.*, **17**, 2249-2255, doi: 10.1021/acs.jproteome.7b00640.
28. Ivanov, M. V., Levitsky, L. I., Bubis, J. A., and Gorshkov, M. V. (2019) Scavenger: a versatile postsearch validation algorithm for shotgun proteomics based on gradient boosting, *Proteomics*, **19**, e1800280, doi: 10.1002/pmic.201800280.
29. Kolbowski, L., Combe, C., and Rappsilber, J. (2018) xiSPEC: web-based visualization, analysis and sharing of proteomics data, *Nucleic Acids Res.*, **46**, W473-W478, doi: 10.1093/nar/gky353.
30. Lane, L., Argoud-Puy, G., Britan, A., Cusin, I., Duek, P. D., et al. (2012) neXtProt: a knowledge platform for human proteins, *Nucleic Acids Res.*, **40**, D76-83, doi: 10.1093/nar/gkr1179.
31. Madden, T. L., Tatusov, R. L., and Zhang, J. (1996) Applications of network BLAST server, *Methods Enzymol.*, **266**, 131-141, doi: 10.1016/s0076-6879(96)66011-x.
32. Schrödinger, L. L. C., and DeLano, W. *PyMOL*, URL: <https://www.pymol.org/pymol>.
33. Capriotti, E., Fariselli, P., and Casadio, R. (2005) I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure, *Nucleic Acids Res.*, **33**, W306-310, doi: 10.1093/nar/gki375.
34. Cheng, J., Randall, A., and Baldi, P. (2006) Prediction of protein stability changes for single-site mutations using support vector machines, *Proteins*, **62**, 1125-1132, doi: 10.1002/prot.20810.
35. Chen, C. W., Lin, J., and Chu, Y. W. (2013) iStable: off-the-shelf predictor integration for predicting protein stability changes, *BMC Bioinformatics*, **14 Suppl 2**, S5, doi: 10.1186/1471-2105-14-S2-S5.
36. Levitsky, L. I., Kuznetsova, K. G., Kliuchnikova, A. A., Ilina, I. Y., Goncharov, A. O., et al. (2022) Validating amino acid variants in proteogenomics using sequence coverage by multiple reads, *J. Proteome Res.*, **21**, 1438-1448, doi: 10.1021/acs.jproteome.2c00033.
37. Pei, W., Huang, Z., and Niu, L. (2007) GluR3 flip and flop: differences in channel opening kinetics, *Biochemistry*, **46**, 2027-2036, doi: 10.1021/bi062213s.
38. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., et al. (2021) Highly accurate protein structure prediction with AlphaFold, *Nature*, **596**, 583-589, doi: 10.1038/s41586-021-03819-2.
39. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., et al. (2018) SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Res.*, **46**, W296-W303, doi: 10.1093/nar/gky427.
40. Adams, N. C., Tomoda, T., Cooper, M., Dietz, G., and Hatten, M. E. (2002) Mice that lack astrotactin have slowed neuronal migration, *Development*, **129**, 965-972, doi: 10.1242/dev.129.4.965.
41. Miyagi, T., Wada, T., Yamaguchi, K., Hata, K., and Shiozaki, K. (2008) Plasma membrane-associated sialidase as a crucial regulator of transmembrane signalling, *J. Biochem.*, **144**, 279-285, doi: 10.1093/jb/mvn089.
42. Rieder, L. E., Savva, Y. A., Reyna, M. A., Chang, Y. J., Dorsky, J. S., et al. (2015) Dynamic response of RNA editing to temperature in *Drosophila*, *BMC Biol.*, **13**, 1, doi: 10.1186/s12915-014-0111-3.
43. Moldovan, M., Chervontseva, Z., Bazykin, G., and Gelfand, M. S. (2020) Adaptive evolution at mRNA editing sites in soft-bodied cephalopods, *PeerJ*, **8**, e10456, doi: 10.7717/peerj.10456.