Biochimie 93 (2011) 1916–1921

Contents lists available at ScienceDirect

Biochimie

journal homepage: www.elsevier.com/locate/biochi

Review

Functional evidence of post-transcriptional regulation by pseudogenes

Enrique M. Muro*, Nancy Mah, Miguel A. Andrade-Navarro

Max-Delbrück Center for Molecular Medicine (MDC), Robert Rössle Str. 10, 13125 Berlin, Germany

ARTICLE INFO

Article history:
Received 14 April 2011
Accepted 19 July 2011
Available online 27 July 2011

Keywords:
Pseudogenes
ncRNA
Natural antisense transcripts
siRNA
miRNA

ABSTRACT

Pseudogenes have been mainly considered as functionless *evolutionary relics* since their discovery in 1977. However, multiple mechanisms of pseudogene functionality have been proposed both at the transcriptional and post-transcriptional level. This review focuses on the role of pseudogenes as post-transcriptional regulators. Two lines of research have recently presented strong evidence of their potential function as post-transcriptional regulators of the corresponding parental genes from which they originate. First, pseudogene genomic sequences can encode siRNAs. Second, pseudogene transcripts can act as indirect post-transcriptional regulators decoying ncRNA, in particular miRNAs that target the parental gene. This has been demonstrated for PTEN and KRAS, two genes involved in tumorigenesis. The role of pseudogenes in disease has not been proven and seems to be the next research landmark. In this review, we chronicle the events following the initial discovery of the 'useless' pseudogene to its breakthrough as a functional molecule with hitherto unbeknownst potential to influence human disease.

1. Introduction

Pseudogenes are "*genomic loci that resemble real genes, yet are considered to be biologically inconsequential because they harbor premature stop codons, deletions/insertions and frameshift mutations that abrogate their translation into functional proteins*" [1]. Pseudogenes originate from gene templates (parental genes) either by retrotransposition of the parental gene's mRNA (processed pseudogenes that have no introns and, in principle, no upstream DNA regulatory regions) or as the product of genome duplication (non-processed or duplicated pseudogenes, which may contain all the parental gene introns and their upstream DNA regulatory regions). DNA sequences of pseudogenes evolve faster than those of their respective parental genes due to mutations, insertions and deletions that prevent the production of a functional protein [2].

Parental genes of pseudogenes can be either currently active genes or genes that were only active in an ancient genome. In the latter case, pseudogenes are clearly "genomic relics" because they constitute the only remains of a once functional gene [3,4]. These "genomic relics" were once protein-coding genes that are no longer able to produce a functional protein. For example, some olfactory receptor (OR) genes in human were inactivated as the human olfactory ability became increasingly limited. While these human OR genes lost their protein-coding ability through pseudogenization, a large proportion of the orthologous OR genes remained functional in other mammals with superior olfactory capabilities [5,6]: about 400 protein-coding ORs remain in human, compared to 1000 in mouse [7]. Another example of a pseudogene whose parental gene is no longer extant involves the loss of the primate ability to synthesize vitamin C. L-gulonolactone oxidase (*GULO*), which is necessary for vitamin C synthesis, is present as a functional gene in most mammals, but it is a pseudogene in primates [8].

Other works have used pseudogenes for comparative studies, for example to study the loss of hemoglobin in different Antarctic icefishes [9]. More recently, we have used prokaryotic pseudogenes as markers of functionally less important genes to demonstrate for the first time that functionally less important genes tend to be located at the end of operons while the more important genes tend to be located toward operon starts [10].

Historically, pseudogenes were not considered functional because their transcripts were generally non-coding, which essentially equated to irrelevancy in a protein-centric world, where the old dogma simply viewed RNA as an intermediate molecule in the protein production process [11]. Taking into account the recent discoveries on the function of pseudogene transcripts that we will describe later, we will see that epithets such as "dead genes" or "junk DNA" are misnomers for pseudogenes.

In this review, we will focus on studies that unearth novel functions of transcribed pseudogenes, which eventually lead to the discovery of their function as post-transcriptional regulators.

* Corresponding author. Tel.: +49 30 9406 4227; fax: +49 30 9406 4240.
E-mail address: enrique.muro@mdc-berlin.de (E.M. Muro).

0300-9084/$ – see front matter © 2011 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.biochi.2011.07.024

2. Pseudogene discovery

The first pseudogene was reported in 1977, in a genomic region coding for the oocyte-type 5S RNA of *Xenopus laevis* [12]. The name, *pseudogene*, was given because: *"this homologous structure was nearly as long as, and almost exact of, the gene itself"*. The pseudogene had a truncated 5′ end and 14 bp mismatches in comparison with its parental gene. In terms of its functionality, the possible role of the pseudogene as a "transcribed spacer" was then discussed, but the authors stressed that they thought pseudogenes were just "*relics of evolution*", a term that since then has been frequently used.

The next three years witnessed the sequencing of pseudogenes from globin genes of different species (rabbit, human, mouse, goat and sheep) [13–19]. The evidence showed that although these pseudogenes had comparable motifs to those previously observed in annotated genes, including a transcription initiation site, mRNA 5′ capping motifs, start and stop codons, and canonical polyadenylation signals (PAS), frameshifts in the pseudogene introduced a number of premature stop codons, abrogating the translation of a full-length functional protein [19]. The fact that many globin genes possessed corresponding pseudogenes in five mammalian species, intriguingly implied some kind of functionality for these new genomic sequences. In ref. [19] it was proposed, without evidence, that if pseudogenes had any functionality this could be due to the use of the transcription machinery in the production of useless transcripts, a process that was described as '*diverting genes*'. As an alternative, the same work proposed, for the first time, the possible function of pseudogenes as *antigenes* (sources of antisense transcripts).

At the time, the idea that pseudogenes had no functionality was entrenched in the minds of scientists: "*a pseudogene is a DNA segment with high homology with a functional gene but containing nucleotide changes such as frameshift and nonsense mutations that prevent its expression*" [20,21]. Reflecting this, the work reporting the first algorithm that calculated the rate of nucleotide substitution within pseudogenes bore the modest title: "*pseudogenes as a paradigm of neutral evolution*" [21]. At that time, it was observed that many pseudogenes had an intronless DNA sequence, so it was thought that their origin could be the result of reverse transcription [20]. It was later demonstrated (thanks to computational analyses of complete genome sequences of many organisms) that there are around five times more human processed pseudogenes than non-processed pseudogenes [22].

In 1985 a comprehensive review was published on the topic of pseudogene function by Elio F. Vanin [23]. He suggested that "*pseudogene be used only to describe sequences found to be both related and defective*". Reflecting the predominance of processed pseudogenes over non-processed ones, the review is focused on the former, reporting 17 pseudogenes at the time. Their genetic defects were predominantly shown to consist of point mutations and indels (insertions and deletions) that lead to a change in the reading frame resulting in premature in-frame stop codons. Some of these pseudogenes had none of these changes, e.g., the processed pseudogene in rat *RC9 cytochrome c*.

The existence of processed pseudogenes with intact coding sequences raised an important question. Could processed pseudogenes actually code for a functional protein? Due to the random genomic location of processed pseudogenes it was thought unlikely that a pseudogene be so lucky to find itself in a promoter region capable of initiating its transcription. 5′ regulatory regions for pseudogenes such as the mouse rpL3204A [24] and the *RC9 cytochrome c* in rat [25] were noted, though no functional protein was ever found for those pseudogenes. Almost two decades passed before the first evidence of pseudogene translation to a functional protein was published in 2002: PGAM3, a protein coded from a (processed) pseudogene found in primate white blood cells (human, chimpanzee and macaque) [26]. Later, other examples of pseudogenes that are translated to truncated proteins were found, namely PsiCx43 [27] and NANOGP8 [28], which expand the regulatory possibilities of pseudogene transcripts.

3. The rise of pseudogene functionality

The next most notable novelty in the field was the discovery that the immune system from chicken, human and rabbit diversifies its response using DNA sequence from pseudogenes through somatic gene conversion mechanisms, by which a DNA segment from the pseudogene is transferred to another immune system gene without modifying the pseudogene sequence [29]. This is the case of the immunoglobulin V_H gene segments [30]. Pseudogenes were identified as repositories of genetic variability but not as having a biological function *per se*.

It was not until 1999 that a pseudogene transcript was first reported to have an active biological function: the post-transcriptional regulation of neural nitric oxide synthase (nNOS) by an antisense transcript encoded by its own pseudogene (pseudo-NOS) [31]. The pseudogene itself is a natural antisense transcript (NAT) that is 145 bp long and shares ~80% complementarity with respect to the parental gene's transcript. This enabled its association to the mRNA of NOS, which prevented its translation and therefore regulated nNOS protein synthesis. This has consequences in neural intercellular signaling. Experimental verification was performed both *in vitro* and *in vivo* in *Lymnaea stagnalis*, a freshwater snail. The working hypothesis of a pseudogene acting as an antigene had been suggested previously ([19] and [32]), but here it had been demonstrated for the first time.

The second report describing biological activity of a pseudogene transcript appeared later in 2003 for the pseudogene Makorin1-p1 [33]. This work had a large influence triggering an important review [34], which we will describe below.

The study by Hirotsune and coworkers [33] concluded that a transcribed pseudogene from Makorin1 (Makorin1-p1) was regulating Makorin1 in mouse, even though some fragmented open reading frames impeded the protein translation of the pseudogene. The authors defined pseudogenes to be "*a gene copy that does not produce a functional full length protein*" [33]. It was observed that a transgene-insertion mutant mouse showed bone deformity and polycystic kidneys, and it was claimed that the insertion reduced the transcription of Makorin1-p1, which was imprinted and affected the mRNA regulation of Makorin1.

In 2006 these results were thoroughly refuted as it was shown that Makorin1-p1 is neither imprinted nor expressed [35]: both Makorin1-p1 alleles are methylated and therefore it is a silent pseudogene, reestablishing the idea that mammalian pseudogenes are only "*evolutionary relics*".

Nonetheless, in 2003, between the Makorin1-p1 work and its refutation, a very relevant review by Balakirev and Ayala was written proposing that pseudogenes can be *potogenes*, DNA sequences that have the potential to evolve to a new gene [34]. This idea had actually been previously proposed shortly after the discovery of the existence of pseudogenes [20].

This review [34] focused on the evolution of pseudogenes. *Drosophila melanogaster* was used as an example for the study of pseudogene evolution, mainly due to the extensive experience of the authors in that organism, although pseudogenes are not as frequent in *Drosophila* as in mammals. This added a different and interesting perspective on the topic. For instance, *Drosophila* pseudogenes were noted for having more synonymous mutations than deleterious mutations, as well as some conserved functional regions, suggesting that they could actually code for proteins.