

## Research Article

# Protein-Encoding RNA to RNA Information Transfer in Mammalian Cells: RNA-dependent mRNA Amplification. Identification of Chimeric RNA Intermediates and Putative RNA End Products

Sophia Rits<sup>1,2</sup>, Bjorn R. Olsen<sup>3</sup> and Vladimir Volloch<sup>3\*</sup>

<sup>1</sup>Division of Molecular Medicine, Children's Hospital, Boston, USA

<sup>2</sup>Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, USA

<sup>3</sup>Department of Developmental Biology, Harvard School of Dental Medicine, USA

## Abstract

Our initial unidirectional understanding of the flow of protein-encoding genetic information, DNA to RNA to protein, a process defined as the “Central Dogma of Molecular Biology” and usually depicted as a downward arrow, was eventually amended to account for the “vertical” information back-flow from RNA to DNA, reverse transcription, and for its “horizontal” side-flow from RNA to RNA, RNA-dependent RNA synthesis, RdRs. These processes, both potentially leading to protein production, were assumed to be strictly virus-specific. However, whereas this presumption might be true for the former, it became apparent that the cellular enzymatic machinery for the later, a conventional RNA-dependent RNA polymerase activity, RdRp, is ubiquitously present and RdRs regularly occurs in eukaryotes. The strongest evidence for the occurrence and functionality of RdRp activity in mammalian cells comes from viruses, such as hepatitis delta virus, HDV, that do not encode RdRp yet undergo a robust RNA replication once inside the host. Eventually, it became clear that RdRp activity, apparently in a non-conventional form, is constitutively present in most, if not in all, mammalian cells. Because such activity was shown to produce short transcripts, because of its apparent involvement in RNA interference phenomena, and because double-stranded RNA is known to trigger cellular responses leading to its degradation, it was generally assumed that its role in mammalian cells is restricted to a regulatory function. However, at the same time, an enzymatic activity capable of generating complete antisense RNA complements of mRNAs was discovered in mammalian cells undergoing terminal differentiation. Moreover, observations of widespread synthesis of antisense RNA initiating at the 3' poly(A) of mRNAs in human cells suggested an extensive cellular utilization of mammalian RdRp. These results led to the development of a model of RdRp-facilitated and antisense RNA-mediated amplification of mammalian mRNA. Here, we report the *in vivo* detection in cells undergoing terminal erythroid differentiation of the major model-predicted identifiers of such a process, a chimeric double-stranded/pinhead structured intermediates containing both sense and antisense RNA strands covalently joined in a rigorously predicted and uniquely defined manner. We also report the identification of the putative chimeric RNA end product of mRNA amplification. It is heavily modified, uniformly truncated, yet retains the intact coding region, and terminates with the OH group at both ends; its massive cellular amount is unprecedented for a conventional mRNA transcription product and it translates into polypeptides indistinguishable from the translation product of conventional mRNA. Moreover, we describe the occurrence of the second Tier of mammalian RNA-dependent mRNA amplification, a physiologically occurring, RdRp-driven intracellular PCR process, “iPCR”, and report the detection of its distinct RNA end products. Whether mammalian mRNA amplification is a specialized occurrence limited to extreme circumstances of terminal differentiation in cells programmed for only a short survival span or a general physiological phenomenon was answered in the companion article Volloch et al. *Ann Integr Mol Med.* 2019;1(1):1004. by the detection of major identifiers of this process for mRNA encoding  $\alpha 1$ ,  $\beta 1$ , and  $\gamma 1$  chains of laminin, a major extracellular matrix protein abundantly produced throughout the tissue and organ development and homeostasis and an exceptionally revealing indicator of the range and scope of this phenomenon. The results obtained introduce the occurrence of RNA-dependent mRNA amplification as a new mode of genomic protein-encoding information transfer in mammalian cells and establish it as a general physiological phenomenon.

**Keywords:** RNA-dependent amplification of mammalian mRNA; Physiologically occurring intracellular PCR, iPCR; RNA-dependent RNA polymerase, RdRp; Chimeric RNA; Sense-strand RNA; Antisense-strand RNA

## Introduction

The initial indications of mammalian RNA-dependent mRNA amplification were obtained in studies of the kinetics of globin mRNA

**Citation:** Rits S, Olsen BR, Volloch V. Protein-Encoding RNA to RNA Information Transfer in Mammalian Cells: RNA-dependent mRNA Amplification. Identification of Chimeric RNA Intermediates and Putative RNA End Products. *Ann Integr Mol Med.* 2019; 1(1): 1003.

**Copyright:** © 2019 Vladimir Volloch

**Publisher Name:** Medtext Publications LLC

**Manuscript compiled:** August 22<sup>nd</sup>, 2019

**\*Corresponding author:** Vladimir Volloch, Department of Developmental Biology, Harvard School of Dental Medicine, USA, E-mail: vladimir.volloch@gmail.com; vladimir\_volloch@hms.harvard.edu

synthesis in differentiating murine erythroid cells, which strongly suggested the occurrence of cytoplasmic *de novo* production of globin mRNA [1]. In these experiments, unexpectedly, in short, 90 seconds, labeling pulses 85% of radioactivity incorporated into globin mRNA was accounted for in the cytoplasm [1]. The labeled molecules appeared to be of a genuine cytoplasmic origin. Indeed, in the nuclei, globin RNA is synthesized in the form of a 15S precursor, which is processed to its mature 9S size with a half-life of about 10 minutes. Following the pulse, only mature size radioactive globin RNA was seen in the cytoplasm and only precursor-size molecules in the nuclei [1]. Cytoplasmic globin RNA was labeled uniformly and not by the end addition to preexisting molecules, consistent with *de novo* synthesis utilizing mature cytoplasmic globin mRNA as the initial template. Moreover, a high concentration of actinomycin D completely inhibited incorporation into nuclear globin RNA but had little effect

on globin RNA labeling in the cytoplasm suggesting the involvement of a distinct enzymatic activity in cytoplasmic RNA synthesis [1]. Two additional lines of evidence suggested a possible involvement of RNA-dependent RNA polymerase, RdRp, an enzymatic activity previously detected in and isolated from rabbit reticulocytes [2]. One was the detection of globin antisense RNA, of a size comparable with globin mRNA and apparently containing a poly(U) segment, presumably a complement of the 3'-terminal poly(A) of globin mRNA. Another was the detection of the erythropoietic differentiation-specific RdRp enzymatic activity [1].

Findings described above were substantiated more directly in studies with cytoplasts, the *in vitro*-enucleated erythroid cells [3]. The availability of cytoplast preparations from differentiating erythroid cells, shown to be free of contaminating nucleated cells, allowed definitive establishment of the occurrence of cytoplasmic synthesis of both positive- and negative-strand globin RNA. Cytoplasts were further selectively permeabilized, enabling the use of substances that do not readily cross the plasma membrane. By using mercury-substituted CTP as a substrate, it was possible to physically separate newly synthesized globin RNA by adsorption to thiol-agarose and show that the cytoplasmic incorporation of radioactive precursors into both sense and antisense globin RNA constituted a *de novo* synthesis. The system allowed experiments establishing that cytoplasmic globin RNA synthesis required the presence of  $Mg^{++}$ , was inhibited by  $Mn^{++}$  and showed no response to  $Zn^{++}$  in contrast to viral RNA replicases. Cytoplasmic globin RNA synthesis was resistant to actinomycin D, alpha-amanitin and rifampicin but inhibited by the rifampicin derivative AF/ABDP and by aurointricarboxylic acid. The results also established that a *de novo* DNA synthesis was not involved in the observed phenomenon: Synthesis of RNA occurred in permeabilized cytoplasts not only without the addition of dNTPs but also when ddNTP was added.

The apparent presence in antisense globin RNA of poly(U), presumably a transcript of 3'-terminal poly(A) of mRNA, indicated that synthesis of an antisense strand may initiate on the poly(A) of mRNA as a template. The precedence for such initiation is known for viral RdRs where it is primed by a uridylated protein [4]. Such a mechanism may or may not be utilized by cellular RdRp but it is a feasibility. A mechanism for the second stage of RNA-dependent globin mRNA synthesis, the initiation of synthesis of sense globin RNA on the antisense RNA strand template, was suggested in a study of the generation of double-stranded murine globin cDNA by RNA-dependent DNA polymerase (RdDp), a process that involves the formation of covalently linked antisense and sense strands in a hairpin configuration [5]. It examined a sequence of molecular events, in particular the self-priming, that initiates synthesis of the sense strand. Upon completion of reverse transcription of globin mRNA and the removal of the RNA template by RNase H activity associated with RdDp, the 3' terminus of the antisense strand snaps back to form a stable double-stranded self-priming structure which is extended by RdDp to generate the sense strand. The self-priming event is enabled by strong complementarity of 14-nucleotide 3'-terminal segment of the antisense strand with an internal segment of the same molecule corresponding to a portion of the 5'untranslated region, 5'UTR, of mRNA located just upstream of the translation start site. Surprisingly and informatively, the strong complementarity between the terminal and the internal complementary elements occurs within the antisense but not the sense strand. This is because A:C mismatches on the sense strand correspond to stable T:G base pairs on the antisense

strand. The self-priming-enabling complementarity within the globin antisense strand is highly preserved throughout a vast evolutionary distance from marsupials to humans [5], with conservation of not only the occurrence of complementary elements but also their positions within the antisense segment corresponding to the 5'UTR of mRNA. Although nucleotide sequences in the 5' UTRs of adult globin mRNAs diverged substantially during evolution, the complementary relationship of the 3' terminal and the internal elements of the antisense strand, as well as the position of the internal element within a segment corresponding to the 5'UTR of mRNA, remained preserved, strongly suggesting that their functionality is physiologically important.

If cellular RdRp activity utilizes the same self-priming arrangement as RdDp to extend the antisense globin RNA strand into a sense-orientation molecule, the resulting pinhead RNA structure would have to be cleaved to separate the globin RNA strands. Where the cleavage may occur within the pinhead structure was indicated by a detailed characterization of antisense globin RNA in mouse erythroid tissues [6]. This study made use of a multistep procedure in which a molecular tag is attached to cellular RNA by ligation with a defined ribooligonucleotide. The act of ligation preserves the termini of RNA molecules, which become the junctions between cellular RNA and ligated ribooligonucleotide. It also unambiguously preserves the identity of cellular RNA as a sense or antisense molecule through all subsequent manipulations [7]. These manipulations included RT-PCR-mediated amplification utilizing a tag-specific and RNA-of-interest-specific primers, and cloning followed by nucleotide sequencing. This approach resulted in identification and characterization of antisense beta-globin RNA molecules in erythroid tissues of anemic mice. The antisense RNA was shown to be fully complementary to spliced globin mRNA, indicative of the template/transcript relationship. At the 5' end antisense globin RNA terminates with a uridylate stretch, reflecting the presence of poly(A) at the 3' end of the sense globin RNA and indicating that synthesis of the former initiates within the 3'-terminal poly(A) region of the latter. With respect to the structure of their 3' termini, the detected antisense globin RNA could be divided into two classes of interest. One class represented a minor population and consisted of full-size antisense molecules corresponding precisely to globin mRNA. The other class represented the major globin antisense population and consisted of 3'-truncated molecules. The truncation was not random; 3'-truncated antisense molecules lacked predominantly fourteen 3'-terminal nucleotides, the same 14-nucleotide segment that constitutes a 3'-terminal complementary element of globin antisense self-priming structure.

A compendium of observations and considerations described above suggested a model for RNA-dependent amplification of mammalian mRNA [6] whose depiction is embedded in Figure 1 ("model" panel) below. In this postulated process, mature spliced cytoplasmic mRNA is transcribed by RdRp. Such transcription initiates within the 3' poly(A) region of mRNA and produces an antisense strand containing poly(U) at the 5' end and terminating at the 3' end with a complement of the 5' terminus of the mRNA molecule. The subsequent transcription of the antisense strand was postulated to occur via self-priming and extension of its 3' terminus. Cleavage within a single-stranded loop of the resulting pinhead molecule and the unwinding of a double-stranded structure would then produce a 5'-truncated sense strand RNA terminating with poly(A) at the 3'end and containing an antisense segment at the 5' terminus, and a 3'-truncated antisense RNA strand. If self-priming

occurs within the segment corresponding to the 5' untranslated region of mRNA, the resulting sense strand would contain the entire protein coding information of the original mRNA and would translate into a polypeptide indistinguishable from the conventionally produced protein. In the case of globin mRNA amplification, the observed predominant 14-nucleotide 3'-terminal truncation of the antisense globin RNA strongly suggested that the cleavage occurs mainly at the 3' end of the single-stranded loop. The version of a model formulated in this study [6] could not specify the order in which the cleavage and the unwinding occur but the results of the present study allowed to define it, as described below.

This proposed mechanism remained largely hypothetical because no convincing evidence in support of the mechanism could be obtained. Indeed, the detection and characterization of globin antisense RNA [6] does not constitute decisive evidence for the occurrence of the proposed RNA amplification mechanism since it could have been generated for other purposes, and the identification of its predominant 3'-terminal truncation in a position consistent with the postulated cleavage at the 3' end of a single-stranded loop of the pinhead molecule that would define the 5' terminus of the RNA end product [6] is, on its own, an indirect indication. But what would constitute a convincing evidence? At the core of the proposed mechanism is the uniquely defined extension of a self-primed antisense strand. Therefore, the detection of the major recognizable attribute of this mechanism, an antisense molecule extended into a sense strand in a rigorously predicted and positionally uniquely defined manner, would provide definitive evidence for the proposed mechanism of the mammalian mRNA amplification process. Two types of postulated RNA molecules would, due to their intrinsic structure, contain such evidential information. One is the double-stranded/pinhead chimeric intermediate consisting of a sense and an antisense strands covalently joined in a precise and defined way, a direct precursor of the RNA end-product. The other is a chimeric RNA end-product containing a defined antisense fragment at its 5' terminus. Due to their transient nature, chimeric intermediates could be a rarity, but the chimeric RNA end product could be expected to represent the majority of globin RNA sequences. In both cases the region of interest is the junction between the sense and the antisense portions. Thus, the present study searched, using next generation RNA sequencing, for the occurrence of such chimeric sense/antisense junction sequences and other relevant RNA molecules of interest in murine erythroid cells.

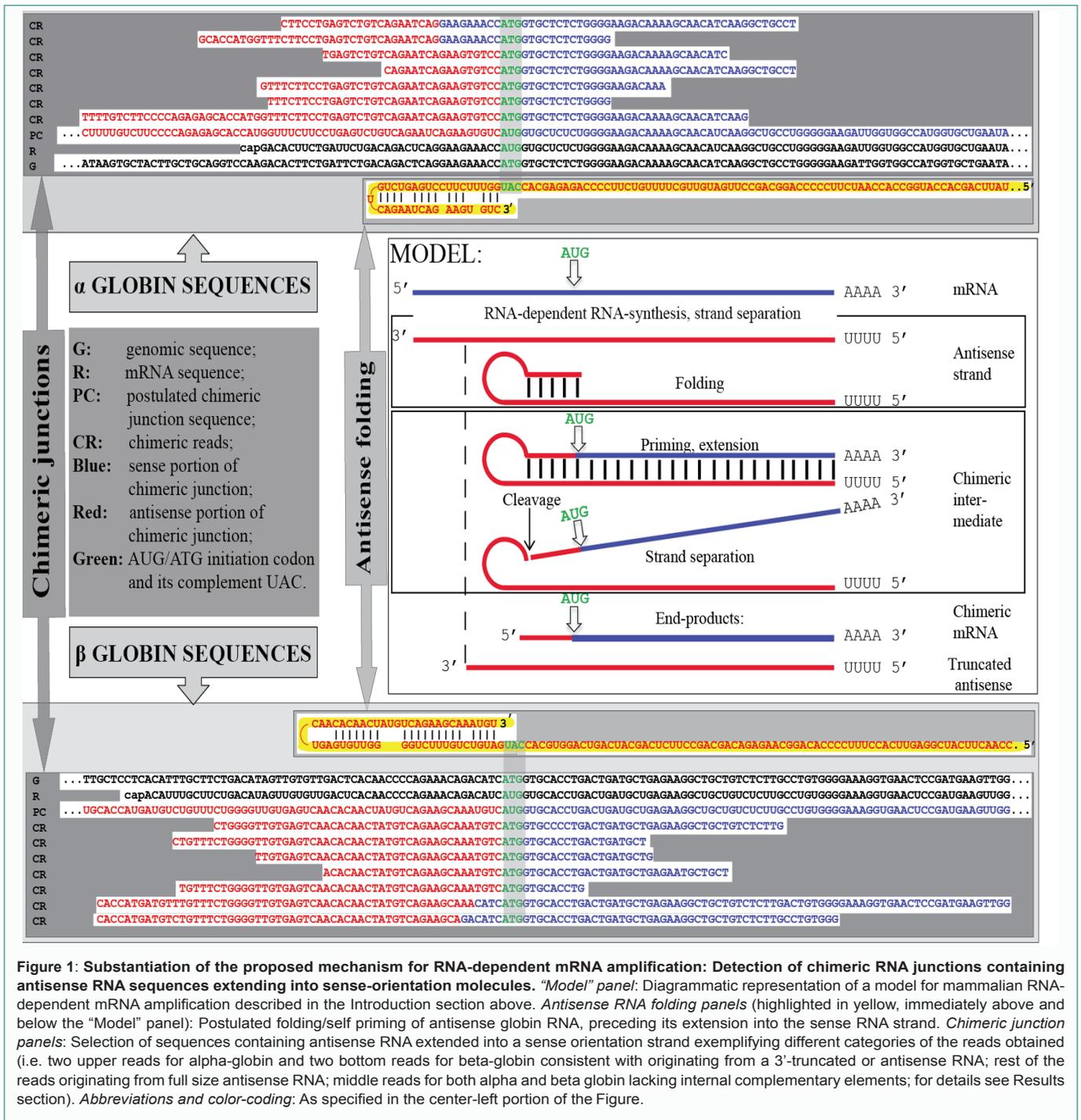
## Results

### Substantiation of the proposed "chimeric" pathway of RNA-dependent mRNA amplification: Detection of chimeric antisense/sense junction sequences

Two types of general searches were conducted to detect globin RNA sequences of interest. In the first, the occurrence of the model-predicted chimeric sense/antisense globin RNA molecules was tested by next generation RNA sequencing. Cytoplasmic RNA from spleen cells of anemic mice was used to generate two types of sequencing libraries, an Illumina TruSeq RNA library and a NEB Ultra Directional RNA library for Illumina, that were analyzed on a MiSeq sequencer; similar reads were obtained with both types of sequencing libraries. The resulting reads were aligned with or blasted against appropriate references, and sequences of the fragments of interest were extracted from raw data and analyzed. A selection of chimeric fragments detected for both alpha and beta globin RNA is presented in Figure 1, together with alpha and beta globin genomic reference sequences, conventional globin mRNA sequences, postulated chimeric sequences,

and folded antisense sequences whose 3' extension would generate predicted chimeric sequences. The Figure also contains a schematic summary of the postulated RNA amplification process that results in chimeric molecules. As shown in Figure 1, two types of chimeric reads were obtained for both alpha and beta globin sequences. One type represents the predicted chimeric junctions resulting from the extension of the full-size self-primed antisense RNA. The other type (two upper reads for alpha-globin and two bottom reads for beta-globin) represents antisense/sense junction sequences consistent with originating from a 3'-truncated antisense RNA retaining a portion of its terminal complementary element apparently sufficient for self-priming and extension. The importance of the first type of antisense/sense junction sequences is self-evident, it constitutes a direct and decisive evidence of the occurrence of mRNA amplification process. The significance of the second type of junction sequences and its mechanistic origin are addressed below in the Discussion section. Whereas the observed "full size" chimeric antisense/sense junctions for beta-globin RNA sequences were exactly as predicted, the antisense portions of the "full size" alpha-globin chimeric junctions uniformly contained at their 3' termini an additional "C" not encoded in the genome. Its origin, implications and significance are also discussed below.

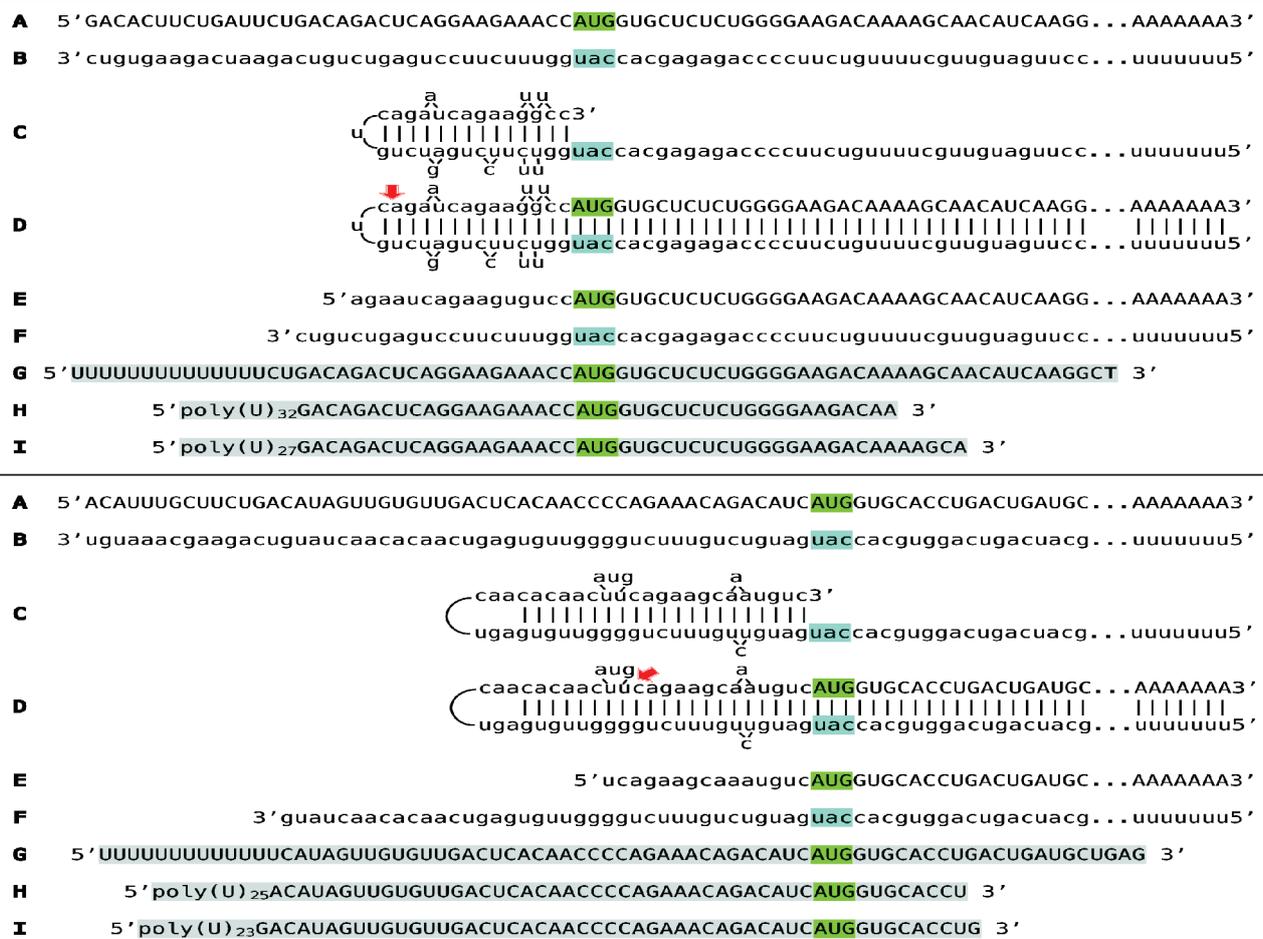
As described above in the Introduction section, during reverse transcription, provided the antisense strand (cDNA) contains the required complementary elements and a reverse transcriptase is associated with the RNase H activity, self-priming can occur and a fragment containing both sense and antisense components may be generated [5]. For several reasons, it can be stated with a high degree of certainty that the chimeric globin RNA fragments detected in the present study did not arise during the reverse transcription or the enrichment steps of sequencing libraries construction but were present in the initial cytoplasmic RNA preparations. First, reverse transcriptase enzymes employed in preparation of sequencing libraries were RNase H-negative mutants. Second, even if a hairpin structure were created during the reverse transcription stage, it would represent a dead end in the library preparation process because no adaptor could be ligated. Third, chimeric sequences were observed not only with a non-stranded Illumina TruSeq RNA library but also with a stranded NEB Ultra Directional RNA sequencing library, where the second cDNA strand is eliminated as a part of the library construction. Fourth, and most important, is the intrinsic geometry of the observed chimeric reads. Indeed, the structure of many of the chimeric fragments detected in the present study and of all chimeric fragments shown in Figure 1 decisively rules out the possibility of their generation during library preparation. This is because they have, when folded by matching the complementary elements discussed above, 3'-protruding sense-orientation globin-specific segments. Since the length of the antisense component 5' of the position of self-priming structure in these fragments is significantly smaller than that of the sense-oriented segment, the former could not serve as a template for the latter during the library preparation process thus corroborating the occurrence of the observed chimeric molecules in the initial RNA preparations. Fifth, some of the chimeric fragments sequenced, in addition to exhibiting 3' globin-specific protrusions, also lack the entire internal complementary element (for example, middle reads for both alpha and beta globin RNA junction sequences in Figure 1), therefore their generation at the reverse transcription stage of library construction would be impossible. The inadvertent generation of a chimeric fragment, either by intramolecular or



intermolecular processes, during the PCR enrichment of a sequencing library can also be ruled out. For this to occur, the 3'-terminal portion of the antisense strand should anneal to the internal portion of the same or another antisense strand and be extended into a sense strand. In sequencing library preparation, the PCR enrichment step follows the adaptor ligation step. If antisense strand contains an adaptor, its 3'-terminal complementary element cannot be extended, even if it anneals either intramolecularly or intermolecularly. If an antisense strand does not contain an adaptor, it potentially can anneal and be extended but cannot be sequenced.

#### Detection of 5'-polyuridyliated/5'-truncated globin sense strand RNA sequences

The second type of general search was conducted for 5'-truncated, sense-orientated globin RNA sequences containing poly(U) segments appended to their 5' termini. The primary reason for this search was an observation by Kapranov and co-investigators [8] of a class of such RNA molecules in human cells where a robust synthesis of antisense RNA initiating at the 3'-terminal poly(A) segments of their sense counterparts was also detected. We reasoned that a connection might exist between these two observations and, as discussed below, it appears that it indeed does. In this type of search, cytoplasmic RNA from spleen cells of anemic mice was used to generate stranded NEB Ultra Directional RNA libraries for Illumina, that were analyzed on a MiSeq sequencer. The resulting reads were blasted against appropriate references, and sequences of the fragments of interest were extracted



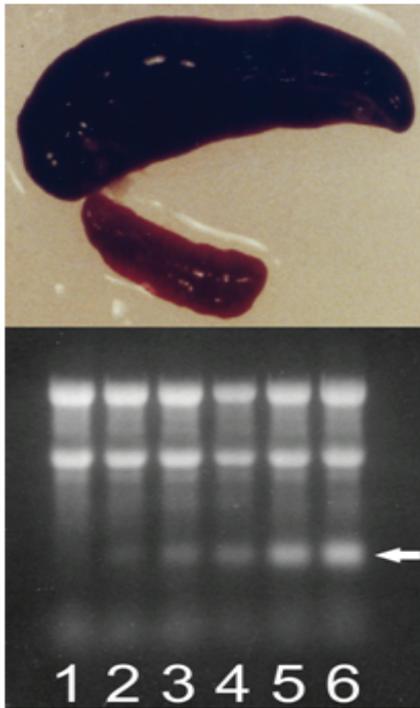
**Figure 2: Detection of 5'-truncated sense strand globin RNA sequences containing 5'-terminal poly(U) segments.** Uppercase letters – nucleotide sequence of the sense strand; lowercase letters – nucleotide sequence of the antisense strand. Highlighted in green – “AUG” translation initiation codon on the sense strand; highlighted in blue – “uac” complement of translation initiation codon on the antisense strand. Red arrows: position of cleavage of the chimeric intermediate. **A-F**: Projected pathways of the chimeric Tier of globin mRNA amplification. **A**: Conventional genome-originated mRNA; **B**: Antisense complement of conventional mRNA; **C**: Folding of the antisense strand into self-priming configuration, 3'terminal “c” is a transcript of the 5'cap G of mRNA; **D**: Extension of self-primed antisense strand into sense-oriented sequence. **E**: Projected chimeric RNA end product. **F**: Projected cleavage-generated 3'-truncated antisense RNA. **G-I**: Selection of detected 5'-truncated sense strand sequences containing 5'-terminal poly(U) segments (highlighted in grey). Note that truncations are clustered around the nucleotide positions corresponding to the projected truncations of the antisense RNA by-product of the chimeric amplification pathways. *Top panel*: Alpha-globin RNA sequences; *bottom panel*: Beta-globin RNA sequences.

from raw data and analyzed. A selection of 5'-terminal poly(U)-containing fragments detected for both alpha and beta globin RNA is presented in Figure 2 (lines G, H, I in both top, alpha-globin, and bottom, beta-globin, panels) together with projected chimeric pathways for alpha- and beta-globin mRNA amplification (lines A-F). In all detected sequences globin portions are 5'-truncated. The truncations are not random. Interestingly, they are mostly clustered around the nucleotide positions corresponding to the projected predominant truncations of the antisense RNA by-product of the chimeric amplification pathways of both alpha- and beta-globin RNA. The implications and significance of these observations, and their relevance to the processes of RNA-dependent mRNA amplification are addressed below in the Discussion section.

### Complete and striking lack of the chimeric RNA end product of RNA-dependent mRNA amplification: Possible explanation

Whereas, due to their transient nature, chimeric intermediates could be a rarity and hard to detect, the chimeric RNA end products

of RNA-dependent mRNA amplification are expected to be highly ubiquitous and represent the overwhelming majority of globin RNA sequences. They are well defined, at least in the case of beta-globin sequences, by the predominant truncations of their antisense portions and therefore should be easy to distinguish. Strikingly, and conspicuously however, *none* of such chimeric sequences characteristic for the end products of the chimeric cycle of globin mRNA amplification were detected. The variable length of antisense portions of the chimeric sequences obtained, some extending well into the region corresponding to the coding portion of the sense strand, as well as the amount of detected chimeric fragments, only about 0.001% of total reads, indicate that all detected fragments are those of double-stranded/pinhead chimeric intermediates randomly cleaved during the construction of sequencing libraries, and not those of the single-stranded chimeric RNA end product. It may seem nonsensical to suggest that the chimeric RNA end product molecules are ubiquitously present in the population but are *concealed* from detection by a proverbial “coat of invisibility” or by its molecular



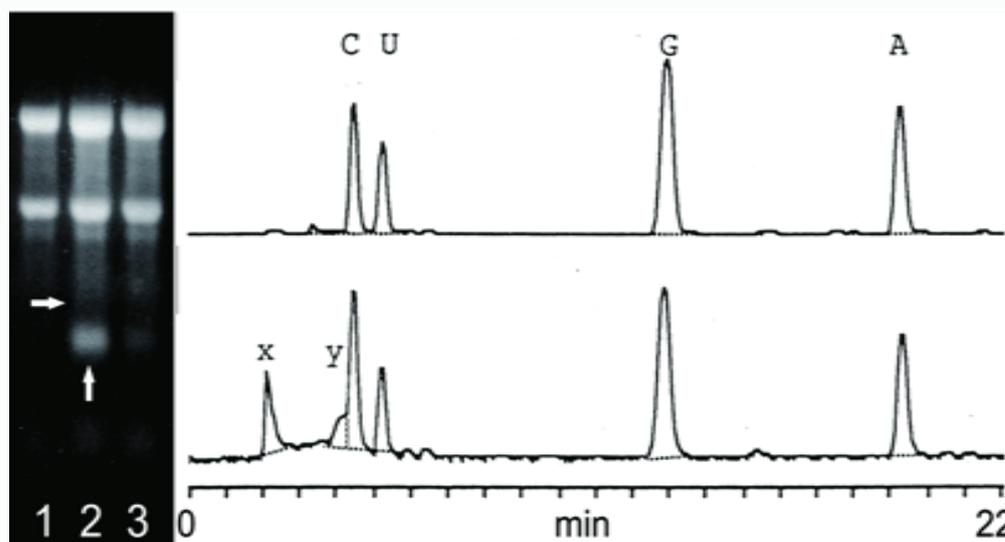
**Figure 3: During induced anemia mouse spleen converts to an erythropoietic organ and produces new type of RNA.** Hemolytic anemia was induced by daily intra-peritoneal injections of phenylhydrazine hydrochloride. Spleen cells were lysed with non-ionic detergent, nuclei were removed, cytoplasmic fraction was treated with proteinase K in the presence of SDS, and RNA was extracted with phenol/chloroform and treated with DNase. Note that Trizol extraction of whole cells cannot be used because pepRNA (putative end product of mRNA amplification) separates with the DNA/protein fraction during such extraction. Cytoplasmic RNA samples from spleen cells collected at different stages of hemolytic treatment were resolved on denaturing agarose/methylmercury hydroxide gels and visualized by ethidium bromide staining. *Top panel:* Conversion of spleen into an erythropoietic organ. At seven days post-induction of hemolytic anemia the spleen mass increases almost 20-fold and nearly 90% of spleen cells are erythroid with characteristics of basophilic erythroblasts. *Upper image:* Spleen at seven days post-induction. *Lower image:* Spleen from untreated animal. *Bottom panel:* Proportion of pepRNA increases with the duration of hemolytic treatment. Lanes 1-6: Number of daily phenylhydrazine injections. Splens were collected and processed 24 hours after final injection. Two prominent upper bands in each lane: 28S and 18S ribosomal RNA. Horizontal arrow: PepRNA. Qualitatively similar results were obtained in mice with induced hemorrhagic anemia [6].

equivalent. But, in fact, as described below, it appears that this is indeed the case. The conspicuous absence of potentially highly ubiquitous RNA end product of amplification suggested a possibility that it might not be detectable by a reverse transcription-based method of analysis employed in the present study. The indirect, not sequencing-based, detection, identification and characterization of such class of globin mRNA is described below.

#### Identification of the putative RNA end product of globin mRNA amplification

Deficiency of red blood cells in mammals, due to blood loss or

induced chemically, by selective degradation of circulating red blood cells with phenylhydrazine injections [9,11], stimulates conversion of the spleen into an erythropoietic organ [9-11]. This splenic reaction is particularly dramatic in rodents. As shown in Figure 3, (*top panel*), at seven days post-induction of hemolytic anemia in mice the spleen mass increases nearly 20 fold, and it consists almost entirely of basophilic erythroblasts [11]. Electrophoresis of total cytoplasmic RNA from such spleens on denaturing methylmercury hydroxide/agarose gels shows the presence of an increasingly pronounced band that is slightly shorter than globin mRNA (Figure 3, *bottom panel*) and whose emergence and cellular levels are associated both with the duration of hemolytic

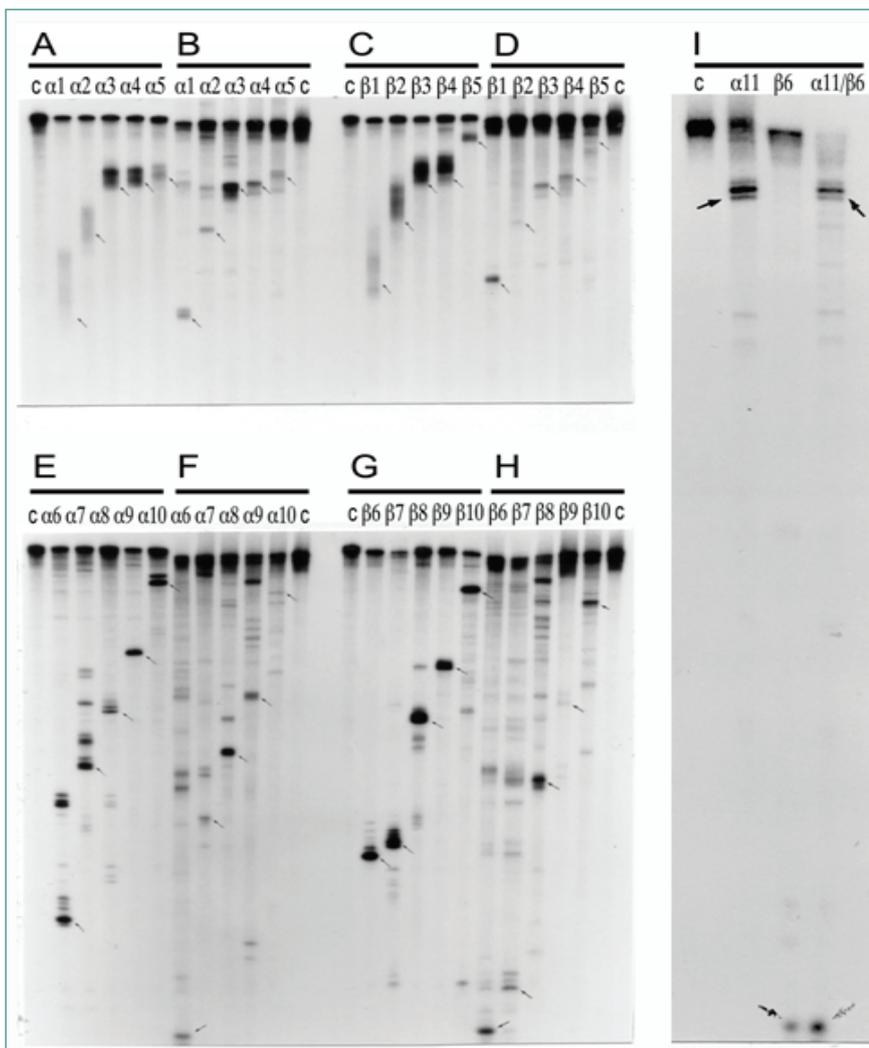


**Figure 4: pepRNA (putative end product of mRNA amplification) contains modified nucleotides.** *Left panel:* Ethidium bromide-stained total cytoplasmic RNA resolved on a denaturing methylmercury hydroxide/agarose gel; Lane 1 – RNA from normal spleen; Lane 2 – RNA from anemic spleen, Lane 3 – RNA from reticulocytes of anemic mice. Two prominent upper bands in each lane: 28S and 18S ribosomal RNA. Vertical arrow: PepRNA; horizontal arrow: Estimated position of conventional globin mRNA. *Right panels:* OD<sub>260</sub> elution profiles of nucleoside digests (as described in Methods section) of control RNA (*upper panel*) and pepRNA (*lower panel*) resolved by HPLC. A, C, G, U: Conventional nucleotides; X, Y: Modified nucleotides.

(or hemorrhagic, not shown) treatment and with hemoglobin accumulation.

This band is resistant to DNase and disappears following RNase or alkaline treatments. Based on the size of this RNA, its kinetics and abundance, and in light of results described below, it was designated pepRNA (*putative end product* of globin mRNA amplification). Eventually, erythroid cells mature into reticulocytes where the levels of hemoglobin are maintained at a plateau [12], and are released into the bloodstream; during this erythroblast/reticulocyte transition the cellular pepRNA levels drop sharply (Figure 4, *left panel*). The pepRNA band produced no signal upon stringent hybridization of Northern blots with globin-specific probes [6]. Since nucleotide modifications can interfere with hybridization properties of nucleic acids [13], the nucleoside composition of pepRNA was analyzed by HPLC (Figure 4, *right panel*). In addition to the nucleoside peaks seen in digests of control RNA, two new peaks were evident in the profile of pepRNA (Figure 4, *right panel*). Mass-spectrometry produced mass numbers of 304 and 320 for the new peaks and standard numbers for the four conventional peaks (267, 283, 244 and 243 for A, G, U and C, respectively). Together, the new peaks comprised 18%, nearly one fifth, of all nucleoside residues. Multiple considerations discussed below indicated that the modified nucleotides are adenosine and guanosine with the same size modifying group with a mass of 37 appended to both.

When Northern blots were hybridized with globin-specific probes under low stringency conditions, the pepRNA band produced a signal; however, increased background precluded any claim of specificity. Therefore, we analyzed the pepRNA using hybridization with globin-specific oligonucleotide probes at relatively low stringency with added levels of specificity, such as direct nucleotide sequencing by primer extension and RNase H mapping analysis. Several synthetic oligonucleotides complementary to alpha or beta globin RNA were used as primers in reverse transcription sequencing of pepRNA and conventional globin RNA. PepRNA was eluted from gels and depleted of regular globin RNA by high stringency hybridization with immobilized globin-specific probes [14,15]. Conventional globin mRNA, isolated from reticulocytes or eluted from the same gel as pepRNA, was also purified by high stringency hybridization. Sequences obtained with all globin-specific oligonucleotide primers used and with conventional globin RNA were of expected length, but all extension reactions with pepRNA, when they occurred, terminated after incorporation of only few nucleotides. When a set of random primers was used with pepRNA, only very short labeled fragments were produced whereas the expected distribution of variable length labeled molecules was seen with conventional globin mRNA. Sequences obtained with pepRNA and defined primers appeared to be globin, but specificity could not be claimed due to their insufficient length.



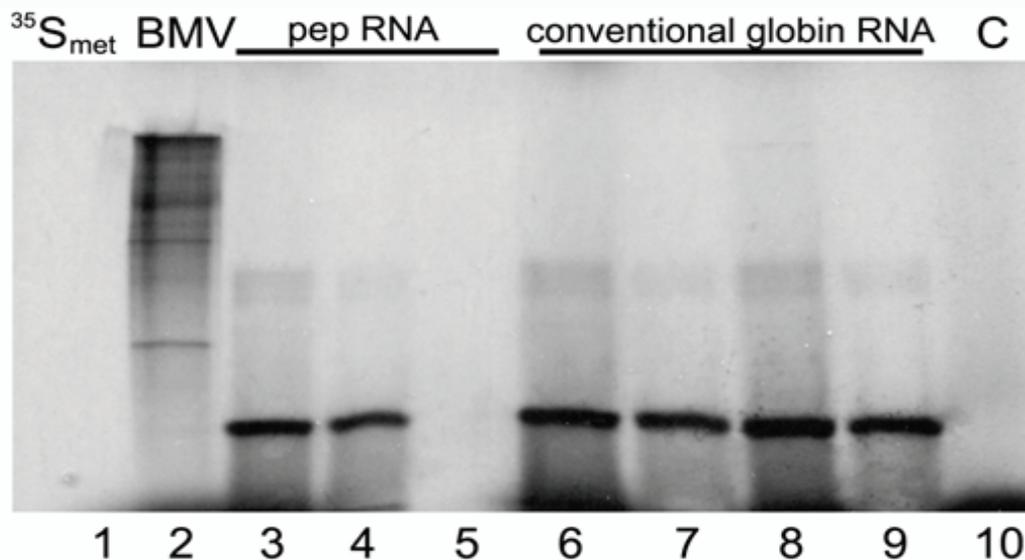
**Figure 5: Comparative RNase H mapping analysis of pepRNA (*putative end product* of mRNA amplification) and conventional globin mRNA.** PepRNA was eluted from methylmercury hydroxide/agarose gels and depleted of regular globin RNA by high stringency hybridization with immobilized globin-specific probes. Conventional globin mRNA, isolated from reticulocytes or eluted from the same gel as pepRNA, was also purified by high stringency hybridization with immobilized globin-specific probes. Purified RNA was labeled at either 5' or 3' terminus. RNase H digest was performed as described in Methods section and mapping analysis was carried out by electrophoresis on 8% polyacrylamide-urea gels. *Panels A, C, E, G, I:* Regular globin mRNA. *Panels B, D, F, H, I:* PepRNA. *Panels A-D:* 3' labeled RNA. *Panels E-I:* 5' labeled RNA. Lanes  $\alpha 1$ - $\alpha 11$ : Alpha-globin-specific oligonucleotides used in RNase H digest assays. Lanes  $\beta 1$ - $\beta 10$ : Beta-globin-specific oligonucleotides used in RNase H digest assays. Lane  $\alpha 11/\beta 6$ : Oligonucleotides  $\alpha 11$  and  $\beta 6$  were used jointly in RNase H digest reaction. Sequences and positions of oligonucleotide probes used in digest assays are specified in Methods section. Lane "c": Control, no oligonucleotide added.

For RNase H mapping, RNA was labeled at 5' or 3' ends, prehybridized with an excess of different oligodeoxynucleotide probes, one at a time unless alpha and beta globin probes were used together, and treated with RNase H. Maps of fragments, obtained using the same sets of alpha and beta globin-specific oligonucleotide probes with conventional globin RNA or with pepRNA, are shown side by side in Figure 5. In experiments with 3'-labeled conventional globin mRNA (Figure 5A-D), the noticeable feature was a diffuse poly(A) tail of variable length, clearly seen with the short fragments produced by RNase H cleavage. In parallel experiments with pepRNA, fragments produced by RNase H cleavage had the same electrophoretic mobility as corresponding fragments of regular globin RNA but with uniformly short poly(A) tail. In RNase H cleavage maps from experiments with regular globin RNA or with pepRNA labeled at the 5' terminus, the pepRNA-originated fragments exhibited patterns of electrophoretic mobility on gels similar to those of corresponding conventional globin RNA-originated fragments (Figure 5E-H). However, fragments of pepRNA were uniformly shorter than corresponding fragments obtained with regular globin RNA; apparently, 20-25 nucleotides shorter for alpha- and 30-35 nucleotides shorter for beta-globin RNA. Oligonucleotide probes differed in their efficiency of mediating RNase H cleavage of pepRNA. When two, one alpha-globin-specific and another beta-globin-specific, "efficient" probes were used together, complete cleavage of pepRNA by RNase H was observed, attesting to its homogeneity (Figure 5I).

Observations made during RNase H mapping analysis led us to conclude that pepRNA terminates with the (OH) group at both 3' and 5' termini. In these experiments, the 5' ends of the regular globin RNA and of pepRNA were labeled by polynucleotide kinase-mediated addition of radioactive phosphate. Before labeling, conventional

globin RNA was pretreated with pyrophosphatase to remove the cap and with phosphatase to remove the remaining 5' phosphate and expose the (OH) group. When these pretreatments were omitted, incorporation of radioactive phosphate was suppressed, apparently for the lack of substrate. No pretreatment, however, was needed for labeling of the pepRNA and the kinetics of kinase reactions with pretreated regular globin RNA was similar to that with non-pretreated pepRNA; when pepRNA was pretreated, there were neither increase in incorporation or its kinetics nor changes in the mapping outcomes. These observations indicate that the bulk of pepRNA terminates at the 5' end with the (OH) group. 3' end-labeling of RNAs was carried out using labeled pCp and T4 RNA ligase. Similar levels of incorporation were obtained with both conventional globin RNA and pepRNA. However, the kinetics of incorporation was strikingly slower with the latter than with the former. These observations suggest that RNA ligase may have difficulties recognizing the 3' termini of pepRNA as a substrate; their implications and significance are discussed below.

The above results indicate that pepRNA is a variant of globin mRNA. According to RNase H mapping results, it contains a complete coding region. Whether it can be translated into globin polypeptides was tested in cell-free translation assays followed by immunoprecipitation and gel analysis. With pepRNA, a single major translation product that could be immunoprecipitated with sheep antisera against murine adult globins, and co-migrated with the translation product of conventional globin RNA, was seen (Figure 6). The possibility that the translation product of pepRNA was due to contamination with regular globin RNA was ruled out by depleting pepRNA of regular globin RNA, prior to the translation assay, by high stringency hybridization with immobilized globin-specific probes, and testing it, after binding to nitrocellulose, by high stringency



**Figure 6: pepRNA (putative end product of mRNA amplification) directs synthesis of polypeptides indistinguishable from the translation product of conventional globin mRNA.** PepRNA was eluted from methylmercury hydroxide/agarose gels and depleted of regular globin RNA by high stringency hybridization with immobilized globin-specific probes. Conventional globin mRNA, isolated from reticulocytes or eluted from the same gel as pepRNA, was also purified by high stringency hybridization with immobilized globin-specific probes. Cell-free translation was carried out in the presence of methylmercury hydroxide and radioactive methionine as described in Methods. Translation products were analyzed on 15% polyacrylamide-SDS gels directly or after immunoprecipitation with sheep antisera against adult murine globin.  $^{35}\text{S}_{\text{met}}$ :  $^{35}\text{S}$ -labeled methionine; BMV: Brome mosaic virus RNA. Lane 1:  $^{35}\text{S}_{\text{met}}$ ; Lane 2: BMV. Lanes 3-5: PepRNA; Lanes 6, 7: Reticulocyte globin RNA; Lanes 8, 9: Conventional globin RNA from anemic spleen. Lanes 4, 7, 9: Translation product immunoprecipitated with sheep antisera against adult mouse globins; Lane 5: Preimmune serum was used in immunoprecipitation. Lane 10, C: Control, no RNA added.

hybridization with labeled globin-specific probes; this yielded a negative result.

## Discussion

### Validation of major predictions of the proposed mechanism for mammalian mRNA amplification

The results described above strongly support the proposed mechanism for RNA-dependent amplification of globin mRNA. The decisive substantiation of the mechanism comes from identification of its major recognizable attribute, chimeric RNA consisting of both sense and antisense sequences covalently joined in a rigorously predicted and uniquely defined manner. These results constitute conclusive evidence for the occurrence of RNA-dependent synthesis of mRNA in mammalian cells. All results obtained with pepRNA, chief among them the capacity to direct synthesis of globin polypeptides, are consistent with the predictions of the proposed mechanism and, cumulatively, indicate that pepRNA is the end product of RNA-dependent synthesis of globin mRNA. The term “putative” is used here only in deference to its inability to hybridize with a probe and to support reverse transcription and, consequently, conventional nucleotide sequencing analysis. These properties, apparently conferred by nucleotide modifications, have two possible explanations. One is that modifications interfere with the formation of complementary bonding. Another is that modifications make this type of RNA “sticky” and mediate the formation of hard to unfold complexes or secondary structures. The second explanation is supported by the following observations. When cytoplasmic RNA from anemic spleen cells is resolved on agarose/formaldehyde gel (where pepRNA is neither denatured nor separated), transferred to a membrane and hybridized with globin-specific labeled probe, in addition to an expected signal with conventional globin mRNA, an order of magnitude stronger signal is seen with ribosomal bands. There is no signal with similarly processed RNA from normal spleen. It appears that pepRNA form complexes with rRNA where “sticky” modifying groups are sequestered enabling hybridization with a probe, an observation that explains translation of pepRNA despite its inability to hybridize in free form: engagement of a codon with rRNA at the ribosome A site apparently sequesters modifying groups and makes its presentation to and hybridization with an anticodon feasible. PepRNA/rRNA interaction appears to be non-specific. When anemic spleen is extracted with Trizol (guanidinium complex) [16,17], pepRNA is largely lost from the RNA fraction, yet when phenol/chloroform-extracted cytoplasmic RNA from anemic spleen is re-extracted with guanidinium, pepRNA is fully retained and can be separated on methylmercury hydroxide/agarose gel. It appears therefore that upon disruption of tissue, pepRNA complexes with DNA, the predominant nucleic acid species amenable, due to its rather linear structure, to forming complexes whereas when cytoplasmic fraction is prepared, the predominant species for pepRNA to associate with is rRNA. The only currently known method to release pepRNA from its interactions and to denature and separate it is treatment with methylmercury hydroxide. In our experience, pepRNA, apparently due to its “stickiness”, could be lost in storage; it is preferable, therefore, to analyze it in fresh preparations.

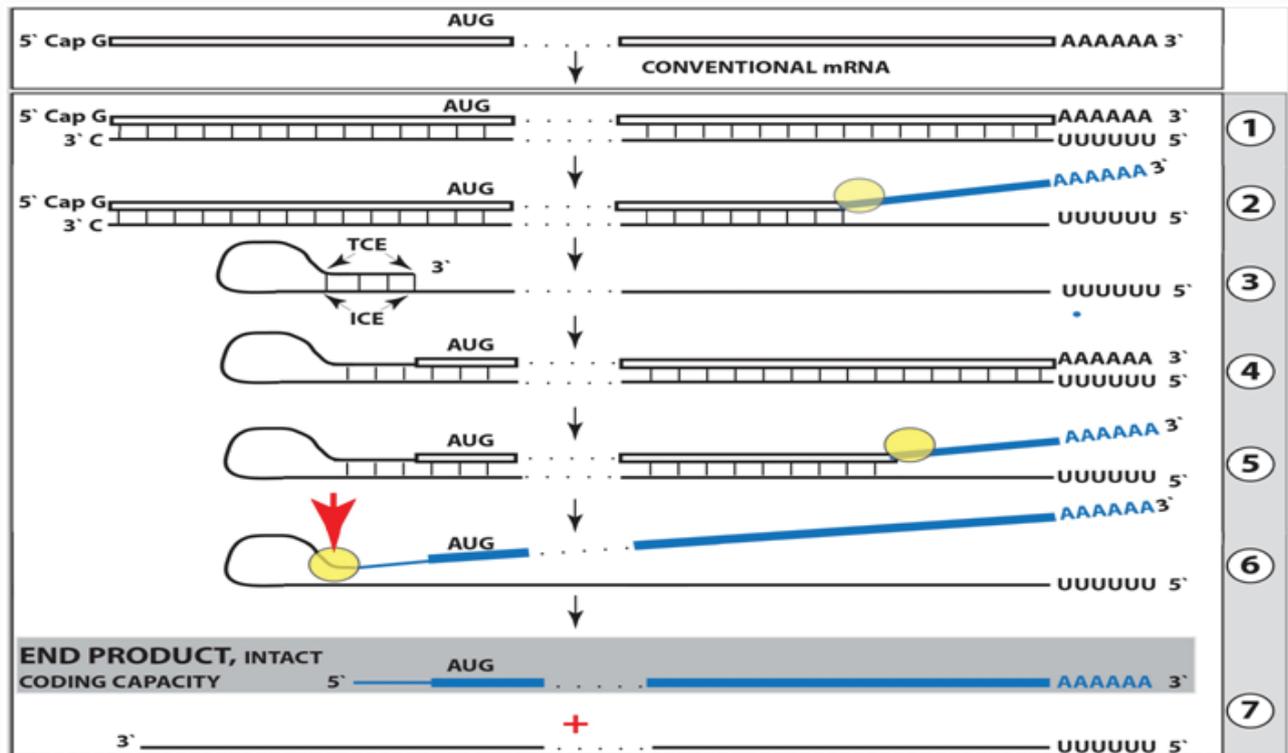
### The chimeric Tier of mammalian RNA-dependent mRNA amplification

Results of the present study warrant the re-consideration of a model for mammalian RNA-dependent mRNA amplification. Its

current understanding in light of these results is reflected in Figure 7 and can be summarized as follows. The amplification process occurs in the cytoplasm and starts with transcription of the antisense complement from a conventional mature spliced mRNA template, initiated at its 3' poly(A) (Figure 7, Step 1). Generation of a complete antisense transcript requires the presence of an eligible RNA template and a compatible polymerase activity. The enzymatic activity central to the mammalian mRNA amplification process is RdRp. The ubiquitous presence of conventional RdRp was shown in numerous eukaryotic organisms but not in mammals [18]. *However, the ability of viruses not encoding the RdRp, such as HDV, to robustly replicate in mammalian cells attested to the cellular presence and functionality of this activity* [18,19]. The RdRp activity in mammalian cells appears to be non-conventional; two possible candidates for this role are the RNA polymerase II complex or its components [20,21] and RdRp activity of the TERT complex [22], both ubiquitously present in all cells. Under regular circumstances, the RdRp activity in mammalian cells produces only short antisense RNA transcripts. For example, a widespread synthesis of diverse short antisense RNA transcripts initiating at the 3' poly(A) of mRNA was observed in human cells [8]. On the other hand, RdRp activity isolated from rabbit reticulocytes [2] was able to produce, in assays, long antisense RNA transcripts. Subsequent studies identified full-length antisense transcripts of globin mRNA in erythroid cells [6].

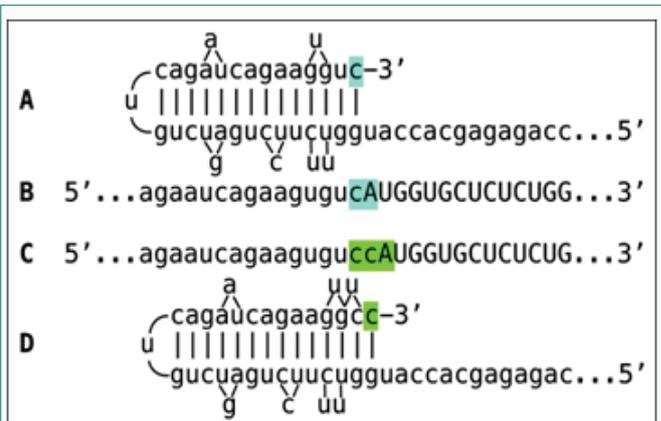
It could be argued that the component responsible for the production of long antisense transcripts in mammalian cells is a processivity conferring co-factor of RdRp activity that is induced under special circumstances when over-production of specific proteins is required. The notion of a processivity co-factor is strongly supported by studies of HDV replication in “normal” (i.e. apparently lacking processivity co-factor) mammalian cells. Within the framework of the above considerations, *the ability of RdRp-deficient viruses to use RdRp activity of mammalian cell for their replication implies that they should encode a processivity co-factor of cellular RdRp*. In case of HDV, it appears to be hepatitis delta antigen HDAG, the only protein encoded by HDV. HDAG is essential [23] both for production of long transcripts by cellular RdRp, and for viral replication. In its absence only short transcripts are generated [23]. *These observations provide a proof of concept for the notion of RdRp processivity co-factor, central to our understanding of mammalian mRNA amplification*. Identification of a cellular homolog of HDAG, DIPA [24,25] suggests directions for a search for the cellular RdRp processivity co-factor.

As for the template eligibility, the only major prerequisite for a potential RNA template appears to be the presence of the poly(A) segment at its 3' terminus [6,8]. The vast majority of mammalian mRNA species contains 3'-terminal poly(A) segments. The notion that many, or possibly most, of them could be eligible templates for RdRp was suggested in earlier studies [6]. Subsequent observations by Kapranov et al. [8] showed a widespread synthesis of antisense RNAs initiating, apparently indiscriminately, at the 3' poly(A) of diverse mRNAs in human cells. This, apparently undiscerning, RdRp template eligibility of the bulk of mammalian mRNA species raises questions with regard to mechanisms underlying the manifestly stringent specificity of the mRNA amplification process as seen, for example, in erythropoietic differentiation [6]. The specificity of the amplification process appears to be determined, as discussed below, at the 3' terminus of an antisense transcript by its ability or inability to support production of a complementary sense strand molecule [6,26,27,73].



**Figure 7: Projected stages of the chimeric Tier of RdRp-facilitated, antisense RNA-mediated amplification of mammalian mRNA.** *Top panel:* Conventional, genome-originated mRNA molecule. *Bottom panel:* Projected stages of antisense RNA-mediated mRNA amplification. Boxed line – sense strand RNA. Single line – antisense strand RNA. “AUG” – functional translation initiation codon (could be other than “AUG”). “TCE”– 3'-terminal complementary element; “ICE”– internal complementary element, both on the antisense RNA strand. Yellow circle – helicase/modifying activity complex. Blue lines (both single and boxed) – RNA strand modified and separated from its complement by a helicase complex. Red arrowhead – position of cleavage of the chimeric intermediate. Step 1: Synthesis of antisense strand; step 2: Strand separation; step 3: Folding of antisense strand into self-priming configuration; step 4: Extension of self-primed antisense RNA; step 5: Strand separation; step 6: Cleavage of the chimeric intermediate; stage 7: End-products of amplification. Note that chimeric RNA end product retains the intact coding capacity of conventional mRNA.

The observed uniformly short 3'poly(A) region of the pepRNA, apparently transcribed from 5'poly(U) of the antisense strand indicates that synthesis of the antisense strand starts within the poly(A) region of the sense strand, in a relatively narrow distance range from the 3'UTR of a conventional mRNA. The manner of antisense RNA initiation remains to be determined; it may involve priming with a uridylylated protein, as in viral systems [4] or occur by a different mechanism. Interestingly, at the 3'terminus of the antisense strand, *RdRp appears to be capable of transcribing the cap "G" of conventional mRNA despite its inverted orientation.* The rationale for this conclusion is based on the finding that antisense/sense junction sequences of chimeric alpha-globin RNA amplification intermediates uniformly contain at the 3' termini of their antisense portions a “C” not encoded in the genome (shown in Figure 1); it is summarized in Figure 8. If the cap “G” is not transcribed, the antisense strand would terminate with the 3' “c” corresponding to the transcription start site of mRNA (panel A; highlighted in blue). In such a case, the antisense RNA folding/self-priming configuration would be as shown in panel A, and following the extension of self-primed antisense RNA, the antisense/sense junction would consist of the “c/A” (highlighted in blue) as depicted in panel B. The experimental results presented in Figure 1 are different. They show that the sequence of the antisense/sense junction is, in fact, “cc/A” (panel C; highlighted in green). Since the genomic sequence upstream of the transcription start site cannot account for the additional 3'-terminal “c” (panel D; highlighted in green) in the antisense strand, the remaining possibility is that the “c” in question is a transcript of the cap “G” of the sense strand and



**Figure 8: RNA-dependent RNA polymerase can transcribe the cap “G” of mRNA.** Data shown are adapted from Figure 2, top panel. Uppercase letters – nucleotide sequence of the sense strand; lowercase letters – nucleotide sequence of the antisense strand. “c” highlighted in blue–3'-terminal nucleotide of the antisense strand corresponding to the transcription start site of mRNA; “cA” highlighted in blue–the projected junction structure in the absence of the cap “G” transcription. “c” highlighted in green–transcript of the cap “G”; “ccA” highlighted in green – the resulting junction structure when cap “G” is transcribed. **A:** Projected self-priming configuration of the antisense strand in the absence of the cap “G” transcription. **B:** Projected nucleotide sequence of the sense/antisense junction in the absence of the cap “G” transcription. **C:** Detected nucleotide sequence of the sense/antisense junction. **D:** Self-priming configuration of the antisense strand as defined by experimental results. Note that the genomic sequence upstream of the TSS cannot account for the additional 3'-terminal “C” in the antisense strand.

that the antisense folding into a self-priming configuration occurs as shown in panel **D**. In the case of beta-globin chimeric RNA junctions, the antisense folding is such (Figure 1) as to either accommodate the cap-transcribed “C” or, if the cap is not transcribed, to account for the “C” in question as the first nucleotide transcribed during the extension. To distinguish between these two possibilities is not possible.

The resulting double-stranded sense/antisense structure is then separated (Figure 7, Step 2) into single-stranded molecules by a helicase/modifying activity. The detection of non-modified chimeric sequences (Figure 1), together with previous observation of non-modified antisense molecule [6], suggest that nucleotide modifications are introduced post-transcriptionally and occur only within the sense-orientated RNA strand. Considering that the only general feature strictly specific for the sense strand is its 3'poly(A) tail, it can be suggested that the poly(A) region of the sense component of double-stranded RNA is both a recognition site and a starting point for a separating/modifying activity. Assuming that modifications facilitate strand separation, taking into account the presence of the terminal 3'poly(A)/5'poly(U) sense/antisense segment in a double-stranded complex, and considering the necessity to prevent its snapback after strand separation, it is likely that one of the modified nucleotides is an altered adenosine. This possibility is consistent with the slow rate of pCp ligation and with the observation that the modified pepRNA species does not bind to oligo (dT). Moreover, an independent argument can be made that the two modified nucleotides are altered adenosine and guanosine. Only these two bases have the same mass number differential, 16, as do the two modified bases, consistent with the possibility that the same modifying group with a mass of 37 is appended to both A and G. *Considering that methylmercury hydroxide, capable of denaturing the modified pepRNA, is known to interact with the nucleotide's NH group during denaturation, and allowing for the mass requirement, one possible candidate for the modifying group of pepRNA is LiCH<sub>2</sub>NH<sub>2</sub>.*

The determination that both modified bases are purines indicates that the modification likely occurs at position(s) not available or not present in pyrimidine nucleotides. One likely candidate is purine position 3 in the proximity to the Watson-Crick interphase, a position used in pyrimidines for sugar attachment. It should be emphasized that while the delineation of the chemical nature of nucleotide modification(s) remains important, accomplishing this will not, on its own, augment our understanding of the origin and the identity of the RNA of interest, but will pave the way to removing or manipulating these modifications so as to allow unimpeded analysis.

The generation of a sense strand on an antisense template occurs via extension of the 3' terminus of a self-primed antisense template and requires the presence within the antisense transcript of two spatially independent complementary elements whose occurrence in antisense globin molecules was found to be evolutionary conserved across mammalian species [5]. One of these is the strictly 3'-Terminal Complementary Element (TCE), the other is the Internal Complementary Element (ICE). These elements (Figure 7, Step 3) must be complementary to an extent sufficient to form a priming structure but may contain both mismatches and unconventional G/U pairings. Generation of a sense strand also requires the thermodynamic feasibility, enhanced/enabled by the occurrence of these two complementary and topologically compatible elements, of the antisense strand folding into a self-priming configuration.

The requirement for terminal localization of the TCE appears to be stringent; an overhang of even a single nucleotide diminishes self-priming [5]. The point that the antisense strand remains unmodified following the separation from its template is crucial since it allows complementary interactions and the formation of a self-priming structure that otherwise would be impeded.

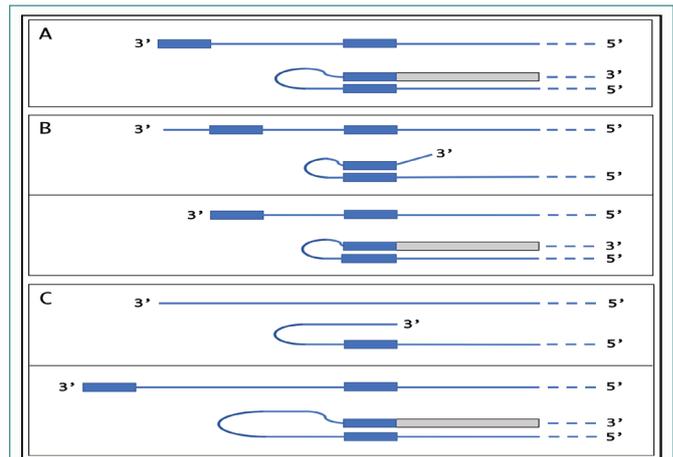
Provided that a self-priming structure is formed, the 3' end of the folded antisense strand is extended by RdRp into a sense-orientated molecule terminating with the poly(A) at the 3' end (Figure 7, Step 4), thus generating a hairpin-structured chimeric intermediate consisting of covalently joined sense and antisense strands. The double-stranded portion of the resulting structure is separated by a helicase activity invoked above, which mounts the 3'poly(A) of a newly synthesized sense strand component of the chimeric intermediate and proceeds along this strand in the 5' direction modifying the molecule as it advances (Figure 7, Step 5). When the helicase activity reaches a single-stranded portion of the hairpin structure, it, or an associated activity, cleaves the molecule either within the TCE, at a TCE/ICE mismatch, or immediately upstream of the TCE; a cleavage was shown to occur between the 5' hydroxyl group and the 3' phosphate (red arrow, Figure 7, Step 6). The cleavage releases the chimeric RNA end product and a 3'-truncated antisense RNA (Figure 7, Step 7). The conclusion that strand separation and associated nucleotide modifications precede the cleavage is validated by the non-occurrence of unmodified chimeric RNA end product, which would be abundantly present if the cleavage were to precede separation/modification process. If cleavage occurs at a mismatch close to the 3' end of the TCE and a substantial portion of the TCE were retained after the cleavage, the self-priming structure could *remain stable and be re-used*, apparently after 3'-P to 3'-OH phosphatase-mediated conversion, to generate another chimeric intermediate. In such a case, the site of antisense/sense RNA junction would shift upstream from the previous one by the size of the antisense truncation resulting from the initial cleavage. If such a process occurs more than once, *multiple antisense truncations are generated and every time the site of the antisense/sense RNA junction is shifted*. In fact, such a process was observed with chimeric intermediates of globin mRNA amplification. As can be seen in Figure 1, one shift of the antisense/sense junction was observed with alpha-globin RNA and two shifts were seen with beta-globin RNA thus explaining the origin of the observed chimeric junctions with a 3'-truncated antisense component. Eventually, the substantial portion of or the entire TCE would be cleaved off the antisense RNA and it could not be re-used. A possible antisense transcript of the chimeric RNA end product would also lack the TCE and therefore would be incapable of self-priming/extension. *The only molecule that can be productively and repeatedly re-used as a template in the chimeric Tier of amplification is a conventional mRNA. Therefore, for the chimeric amplification pathway to operate, it has to be presumed that conventional mRNA, once utilized and thus modified, can be transcribed by RdRp complex.*

In the chimeric pathway of mRNA amplification, the cleavage of the chimeric intermediate following the strand separation and the associated modification of its poly(A)-containing component of the double-stranded structure is the ultimate act in the generation of the chimeric mRNA end product. Consequently, it is formed already modified and, unlike its hairpin chimeric intermediate/precursor is never present in the unmodified form and therefore *cannot be detected* by sequencing methods dependent on the lack of modifications. This explains the striking absence of chimeric RNA end product sequences among detected chimeric reads. On the other

hand, the chimeric hairpin intermediate remains unmodified until the extension of a self-primed antisense strand into the sense-orientated molecule is concluded by the synthesis of the 3' poly(A) segment transcribed from the 5' poly(U) of the antisense strand and required for the commencement of strand separation/modification. Moreover, following the completion of the extension and the commencement of separation, the sense/antisense chimeric junction is not modified for the duration of the strand separation process downstream from the junction, thus generating a temporal window of opportunity and a certain steady-state level and enabling the detection of the yet unmodified junction segment by the conventional reverse transcription-based sequencing methods.

There is very little doubt that the observed pepRNA is a variant of globin mRNA. *But is it the projected chimeric RNA end product of the amplification process?* Multiple considerations strongly support an affirmative answer. First, the 5'(OH) group of pepRNA indicates that its 5' terminus is generated by cleavage, just as projected for the chimeric RNA end product. Second, 5' truncations of pepRNA seen with both alpha-globin probes (about 20-25 nucleotides) and beta-globin probes (about 30-35 nucleotides) match very closely with the sizes projected for 5' truncations of the chimeric RNA end product (20 nucleotides for chimeric alpha-globin RNA and 36 nucleotides for chimeric beta-globin RNA). Third, the massive cellular levels of pepRNA, shown to be homogeneously alpha and beta globin-specific (Figure 5) and to encode only globin polypeptides (Figure 6), are unprecedented for and vastly exceed the observed levels of even most abundant conventional mRNA transcripts. In differentiating mammalian erythroid cells, the mass of conventional globin mRNA is about 0.01% of that of ribosomal RNA [28,29]. *At its peak, pepRNA constitutes about 15% of ribosomal RNA (Figure 3), a 1500-fold increase and a clear indicator that an amplification process is at work.*

The functions of nucleotide modifications occurring during mRNA amplification remain to be elucidated. They could include the facilitation of RNA strand separation, stimulation of the cap-independent translation of the amplification end product, and regulation of its stability. The latter is suggested by a drastic decrease in levels of pepRNA, but not of conventional globin mRNA [10,11], during the erythroblast/reticulocyte transition, consistent with modifications marking pepRNA for degradation when it is no longer needed. There may also be a connection between nucleotide modifications of amplified mRNA and the activation of the mRNA amplification process. The amplified and heavily modified mRNA could behave in ways that are different spatially, qualitatively and quantitatively from those of conventional mRNA. For example, the amplification of mRNA species encoding secreted proteins could overwhelm the ER, cause ER stress, and trigger cell death. One cellular response to ER stress appears to be a shift of translation and secretion outside the ER [30]. It could be suggested, therefore, that nucleotide modifications of amplified mRNA may direct its translation and secretion of the resulting protein via pathways bypassing the ER, despite the presence of a signal peptide sequence. The initial ER stress resulting from increased transcription and subsequent translation of conventional mRNA encoding a secreted protein could be one of potentially multiple cellular events, probably certain types of stresses, which may trigger mRNA amplification processes. In this case, one can envision that conventional overproduction of a secreted protein induces ER stress and activates multiple transcription factors [31,32], which, in turn, activate mRNA amplification pathway thus facilitating overproduction and at the same time relieving the ER stress, since



**Figure 9: TSS shift as potential regulator of the eligibility of an mRNA for the amplification process.** Single line: 3' terminus of the antisense strand. Filled grey boxes: sense strand. Filled blue boxes: topologically compatible complementary elements on the antisense strand. **A:** One of the complementary elements is 3'-terminal; folding results in a self-priming structure that is extended into the sense strand. **B:** Both complementary elements are internal, no self-priming is possible; TSS shift in the downstream direction makes one of the elements 3'-terminal and allows self-priming and extension into the sense strand. **C:** There are no complementary elements/no self-priming; TSS shift in the upstream direction generates 3'-terminal complementary element and thus enables self-priming and extension. Note that processes depicted in panels **B** and **C** can occur in reverse, resulting in a loss, rather than the acquisition, of the eligibility.

in the mRNA amplification process, a portion of conventionally produced mRNA molecules, used as templates for the production of antisense RNA, is apparently modified during strand separation (Figure 7, step 2) and would be translationally processed outside the ER, alongside the modified chimeric RNA end product of mRNA amplification.

Globin genes contain the "TATA" regulatory element and, consequently, a fixed transcription start site (TSS). Globin antisense RNA contains both the TCE and the ICE and is eligible for RdRp-mediated amplification. On the other hand, with mRNA transcripts of TATA-less genes where transcription can be initiated from multiple positions, the eligibility for RdRp-mediated amplification could be regulated by a shift in the TSS position. The concepts of such a regulation are summarized in Figure 9. If the 3'-distant complementary element of the antisense strand is terminal (Figure 9, panel **A**), it can form a self-priming structure. If, however, both elements are internal (Figure 9, panel **B**), a downstream shift of the TSS position can make one of them 3'-terminal and thus enable self-priming. When, on the other hand, the 3' segment of the antisense strand has no topologically compatible complementary sequences, an upstream shift of the TSS position (Figure 9, panel **C**) can generate such an element and make the transcript eligible for amplification. The events diagrammed in panels **B** and **C** of Figure 9 can also occur in reverse, making previously eligible mRNAs ineligible for the amplification process.

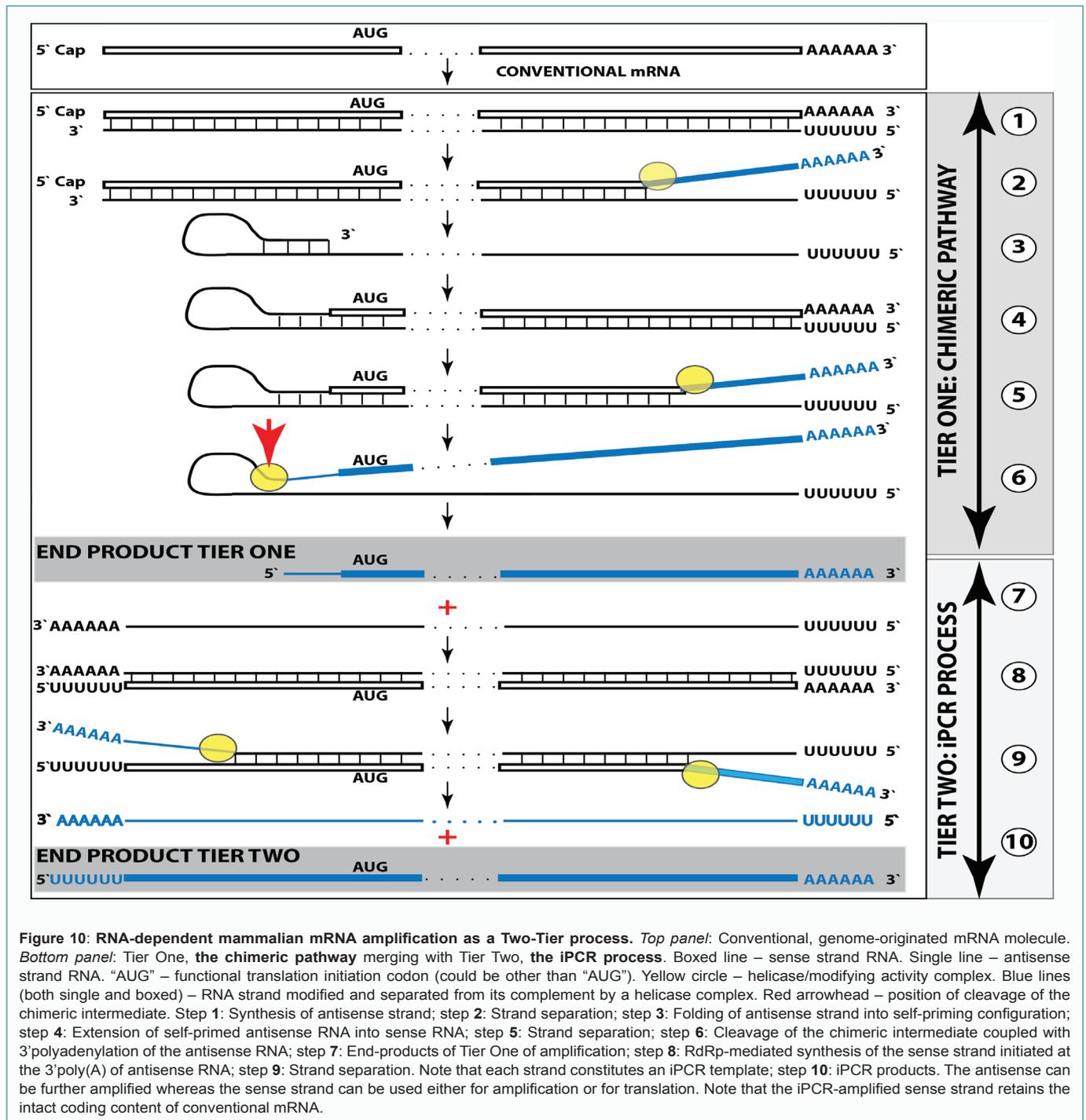
### Second Tier of mammalian mRNA amplification: Physiologically occurring polymerase chain reaction, iPCR

In addition to chimeric sequences of antisense RNA extended into a sense RNA, described above, the analysis of globin RNA molecules in murine erythroid cells identified sense globin RNA sequences

with non-conventionally templated 5'-terminal poly(U) added to the 5'UTR truncated at positions corresponding to potential cleavage sites within the antisense component of the chimeric intermediate (Figure 2). What could be the origin of such sequences? *Provided the presence of 3'poly(A) on an RNA molecule is necessary and sufficient for initiation of RNA-dependent RNA synthesis, another mRNA amplification paradigm, constituting the second Tier of RNA-dependent mRNA amplification, may be considered.* If an antisense transcript is polyadenylated at the 3' end by a known or a novel cellular poly(A) polymerase, it would become a valid template for RdRp. Since the antisense strand has, by virtue of initiation within the poly(A) of conventional mRNA, a poly(U) stretch at the 5' end, its transcription by RdRp would result in a sense strand with poly(U) at the 5' end and

poly(A) at the 3' end, also a legitimate RdRp template. Since strand separation mechanisms are in place and the described sequence of events can occur repeatedly, *the process will amount to an intracellular polymerase chain reaction, iPCR.* The obvious question regarding such a process is its specificity. If 3' polyadenylation of an antisense molecule was coupled with the cleavage of the chimeric intermediate, the specificity of iPCR would be equal to that of the initial chimeric amplification round.

Such Two Tier RNA-dependant mRNA amplification process is diagrammatically presented in Figure 10. Up to the step 6, all processes occur exactly as shown in Figure 7 and described above. In steps 6/7, the cleavage at a TCE/ICE mismatch or at the 5' end of the TCE (red



arrow) is coupled with polyadenylation of a newly created 3' terminus of the antisense strand. This event marks a conclusion of the chimeric cycle of amplification. One of its end products, a chimeric RNA marked in Figure 10 as "END PRODUCT, TIER ONE", is identical to the chimeric end product of Figure 7. The other end product is different. It is antisense strand, 3'-truncated in the same position as depicted in Figure 7, but in addition to the 5'-terminal poly(U), it also contains the 3'-terminal poly(A) (Figure 10, step 7). This molecule constitutes the initial template of a polymerase chain reaction. Indeed, just as in step 1, RdRp activity initiates transcription at the 3' poly(A) and generates the sense strand containing 5' poly(U) and 3' poly(A) (Figure 10, step 8). After strand separation (Figure 10, step 9) there are now two templates, each containing 3' poly(A) and 5' poly(U) (Figure 10, Step 10), and the iPCR is under way. Strand separation and associated nucleotide modification probably commence at 3' poly(A) as soon as it becomes double-stranded. Therefore, at steady state there would be many more non-modified poly(U)-containing than poly(A)-containing ends and because of problems with analyzing modified RNA, discussed above, their detection by conventional nucleotide sequencing would be much more likely.

Conceptually, amplification of a nucleic acid molecule by a polymerase chain reaction necessitates, beside the presence of the building blocks, the occurrence of a template, the priming arrangement for the initial nucleic acid strand and for its complement, a polymerase, and the arrangement for strand separation that doubles the number of template molecules in each cycle. In a conventional PCR reaction, a single-stranded DNA molecule serves as a template, oligonucleotides complementary to the initial DNA strand and to its complement in desired and appropriate positions function as primers (the "forward" and the "return" primers), DNA polymerase of choice extends the 3' end of a primer, generating a double-stranded molecule, and a thermal treatment separates the strands to enable the next cycle of the chain reaction.

In the cellular iPCR process, an eligible, i.e. 3'-terminal poly(A)-containing, single-stranded RNA molecule acts as a template, the process is driven by an RdRp that generates a complementary strand, the priming arrangements are reflected in the template eligibility requirements that are satisfied by the occurrence of poly(A) segment at the 3' termini of an RNA template as well as of its transcript/complement, and strand separation is carried out by a helicase activity. When a full-length 5' poly(U)-containing antisense strand is generated (transcribed by RdRp from a conventional mRNA molecule) and separated from its template (steps 1 and 2, Tier One, Figure 10), it is not an iPCR-eligible template because it lacks the 3'-terminal poly(A) segment. Instead, provided that it contains the 3'-terminal and internal complementary elements, it self-primers its extension into a sense strand molecule; thus generating an intermediate in the chimeric pathway of mRNA amplification. *It is the processing of this intermediate that has the potential to produce, in addition to chimeric RNA end product, an iPCR-eligible template.* This requires cleavage-coupled 3' polyadenylation of the antisense strand, which already contains 5'-terminal poly(U) transcribed from the poly(A) of a conventional mRNA progenitor molecule. The presence of the 3'-terminal poly(A) segment would allow the RdRp to initiate and to proceed with the synthesis of a sense strand complement that commences with the 5'-terminal poly(U) and concludes with the 3'-terminal poly(A) segments, also an iPCR-eligible template. The following separation of strands by a helicase activity would enable the next cycle of a polymerase chain reaction. Thus, the key feature

underlying the feasibility of iPCR is that both the initial cleavage/polyadenylation-released antisense molecule and its transcript are eligible RdRp templates. Potentially and purely hypothetically, if an antisense RNA transcribed from a conventional mRNA were polyadenylated at the 3' end, it would become an eligible iPCR template. This, however, is highly unlikely because such a process would completely lack specificity. *In the Two-Tier amplification process, because the generation of the initial iPCR template is coupled with and enabled by the concluding step of the previous Tier, the specificity of the iPCR process is as stringent as that of the preceding chimeric pathway,* which, in turn, is defined by the occurrence within an RdRp transcript of the TCE and the ICE features and by the ability of the initial antisense transcript to self-prime its extension. On the other hand, whether or not the second Tier of amplification, iPCR, occurs does not affect in any way the first Tier, a chimeric pathway.

The translationally functional end product of the iPCR amplification pathway is a *non-chimeric* sense strand RNA identical to a conventional genome-encoded mRNA in informational content and all other attributes except three – it contains modified nucleotides, it is 5' polyuridylylated, and it is 5' truncated. The truncations are no larger than the TCE of a corresponding antisense RNA, and since TCEs appear to be relatively short, so would the 5' truncations be. Mechanisms underlying the iPCR Tier of amplification could explain the not yet elucidated observations in human cells of a class of unconventional mRNA molecules that differ from their conventional counterparts only in two aspects: (a) their genome-encoded portions are truncated at the 5' end, typically by 14 to 18 nucleotides, not affecting their protein-encoding capacity, and (b) they contain poly(U) segments appended to their truncated 5' termini [8]. In terms of the Two-Tier mRNA amplification mechanisms, such molecules could be the products of the second Tier of the mRNA amplification process, and the extent of 5' truncations in their genome-encoded portions could reflect the average size of an antisense 3' truncation at the conclusion of the preceding chimeric pathway, i.e. the distance between the cleavage site within a chimeric intermediate and the 3' end of an antisense strand.

Whereas the occurrence and regulation of the chimeric pathway is entirely independent from the iPCR process, the occurrence of the latter depends on the completion of the former. The occurrence of the iPCR also requires that the last step of the chimeric pathway, a cleavage of the chimeric intermediate, is coupled with the 3' polyadenylation of the truncated antisense RNA strand. Arguably, the iPCR process may vary in a regulated manner in accordance with the needs for the production of a protein encoded by an mRNA being amplified. For example, although 5'-truncated polyuridylylated RNA, the identifier of the iPCR process, was detected in the globin system (Figure 2), no significant amount of this product was seen in the RNase H digest analysis (Figure 5). Both sequencing analysis and RNase H digest studies were conducted at the very late stage of splenic erythropoiesis. It is conceivable that at this stage the chimeric pathway is sufficient to maintain hemoglobin production and that the iPCR process occurred primarily at early stages of splenic erythropoiesis and only at low rate at late stages. Considering that the presence of the 3'-terminal poly(A) segment is a presumably single major template eligibility requirement for RNA-dependent RNA synthesis [6,8], and allowing for the operational presence of the RdRp, a key component of both the chimeric pathway and the iPCR process, it is reasonable to suggest that *the principal regulatory checkpoint for the entry into the Tier Two of mRNA amplification could be the rate of the chimeric intermediate*

*cleavage-coupled 3' polyadenylation of the antisense RNA strand.* In the extreme case, if the polyadenylation (but not the cleavage) was completely suppressed, so would the Tier Two be, but neither the outcome nor the efficiency of the Tier One process would be affected.

In the iPCR process, if, as indicated above, RdRp initiates transcription within the poly(A) (rather than at its 3' end), the poly(A)/poly(U) shortening will occur in successive iPCR rounds until template eligibility would be lost unless poly(A) elongation takes place simultaneously with its shortening. This is in contrast to the chimeric pathway where repeated use of conventional mRNA as a template for antisense RNA synthesis apparently has no impact on the length of its 3' poly(A). Additional augmentation of the iPCR's efficiency could occur at the level of the iPCR template availability, which is comprised of two components: stability of both sense and antisense strands produced in the reaction and the utilization of the end product of the second Tier. The iPCR is, as any chain reaction, potentially an exponential process provided that the entire output of every cycle is utilized in the next cycle of a chain reaction. However, if all sense strand products in every cycle were withdrawn because of the recruitment for protein synthesis, this Tier of amplification would be linear, akin to the chimeric pathway. In all probability, some sense iPCR-produced molecules are removed for translation and some are utilized as templates for further amplification, making the process faster than linear but slower than exponential. As for the stability of the iPCR templates, it could be regulated, at least in part, by nucleotide modifications present in both RNA strands. Indeed, when cells undergoing erythroid differentiation progress from the stage of rapid hemoglobin accumulation to the stage of the maintenance of steady hemoglobin levels (erythroblast/reticulocyte transition), modified RdRp-amplified globin mRNA accrues to extremely high levels during the former but rapidly disappears during the latter stage. This suggests that the amplification-associated RNA modification could mark the molecule for degradation when it is no longer needed.

Considering that the RNA-dependent mRNA amplification process proceeds through the double-stranded RNA stages, an obvious question is how it evades the dsRNA response? One possibility is that the amplification process is compartmentalized and the components of the dsRNA response have no access to it. Another possibility is the involvement of factors protecting from the dsRNA response. What is known experimentally, is that double-stranded RNA components associated with RNA-dependent mRNA amplification appear to be extremely short-lived as reflected in their steady-state levels. Thus, the frequency of reads for chimeric RNA intermediates for both chains of globin are one in  $10^6$  (after ribosomal RNA depletion). This means that strand separation occurs very rapidly and double stranded molecules have very limited exposure. Moreover, *provided that one of the functions of modifications is strand separation, when already modified RNA is used as a template for RdRp, as in the re-use of conventional mRNA as a template or in the iPCR process, strand separation would occur concurrently with the progression of the polymerase complex and no double-stranded RNA exposure would result.*

In the iPCR Tier of mRNA amplification, because the helicase/modifying activity requires the presence of the 3'-terminal poly(A) in a double-stranded configuration to commence strand separation, the 5' poly(U)-containing mRNA end product remains unmodified until its synthesis is completed. Even after the sense strand RNA product of the iPCR acquires its 3' poly(A) segment, transcribed from the 5' poly(U) of the antisense strand, its 5' portion remains unmodified

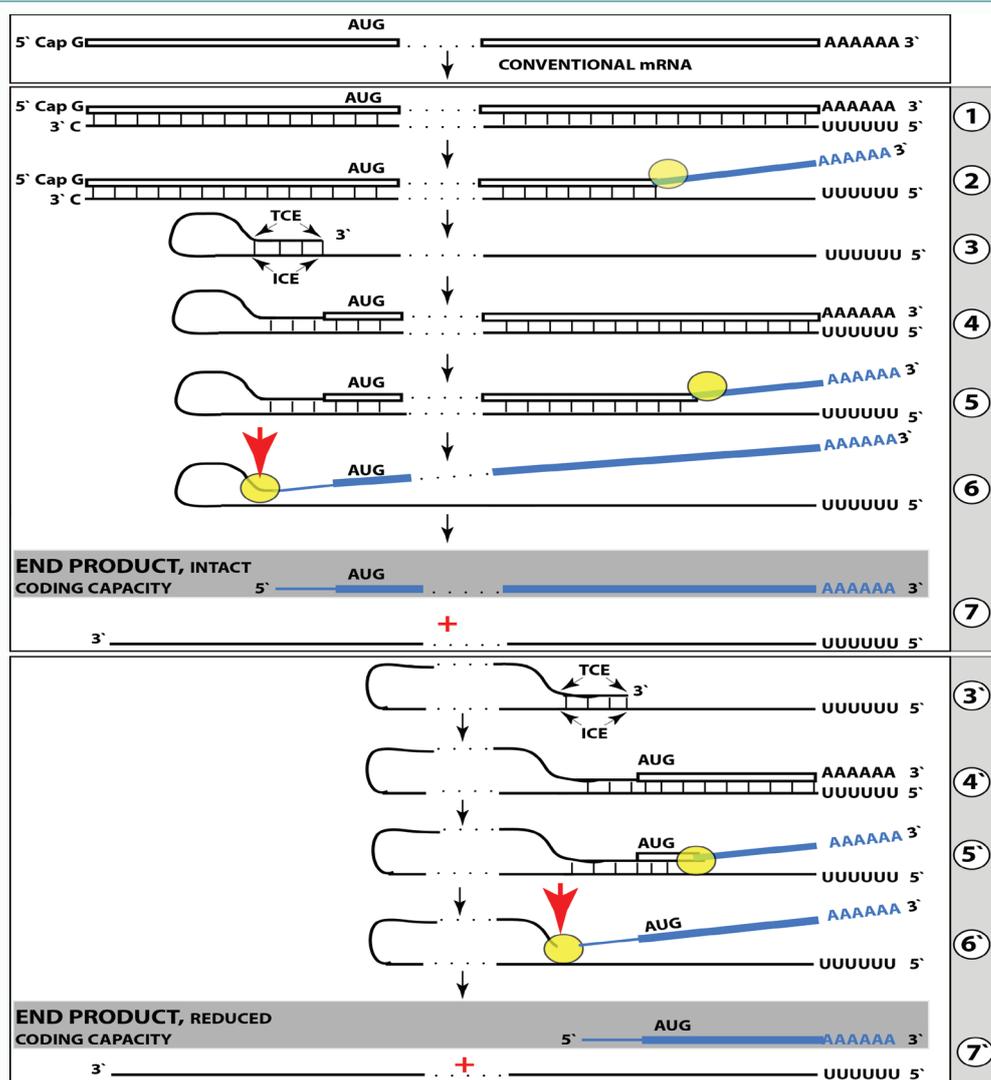
for the limited duration of the strand-separating/modifying activity traversing the length of the molecule. These two processes, the completion of the synthesis and the strand-separation/modification downstream from the region of interest, provide a temporal window of opportunity that, apparently, enabled the detection of the yet unmodified 5'-terminal poly(U)-containing segment of the mRNA end product of the iPCR Tier of the mRNA amplification process (Figure 2).

A temporal window of opportunity is much smaller with respect to the detection of yet unmodified 3'-terminal poly(A)-containing segment of the antisense strand. At the conclusion of the chimeric pathway, it remains unmodified for probably only a short duration between the cleavage/3' polyadenylation of the chimeric intermediate and its separation from a probably still nascent 5' poly(U)-containing sense strand being transcribed from it. During the iPCR process this duration could be even shorter, comprised only of the period between completion of the 3' polyadenylated antisense synthesis and the commencement of the modification starting at its newly added poly(A) component as soon as poly(U) is transcribed. In such a case the modifying activity could be lagging closely behind the polymerase complex. It should be emphasized that *the detected, not yet modified, 5'-terminal poly(U)-containing segment of the sense strand product of Tier Two represents, probably, a very minor subpopulation, with substantial populations of modified molecules remaining for now, alongside the highly abundant end product of the chimeric pathway, an RNA "Dark Matter" [58].*

### **The chimeric Tier of mammalian RNA-dependent mRNA amplification may result in multiple translational outcomes**

In the chimeric Tier of globin mRNA amplification, TCE and ICE are located within the antisense segment corresponding to the 5'UTR of a conventional genome-encoded mRNA. In such a situation, depicted in steps 3 through 7 of Figure 11, the chimeric RNA end product contains the entire protein-coding region of a conventional mRNA and can be translated into the original, conventional mRNA-encoded, polypeptide. In the chimeric mRNA amplification pathway, the position of the TCE within the antisense molecule is fixed; it is strictly 3'-terminal. In contrast, the intramolecular location of the ICE is variable, and potentially it can be positioned within a segment of the antisense strand corresponding to the coding portion of an mRNA, a scenario diagrammed in steps 3' through 7' of Figure 11. In this scenario, the chimeric RNA end product would consist of a 3'-terminal segment of the antisense strand (the TCE or its fraction) and a 3' portion of a conventional mRNA progenitor with a 5'-truncated coding region. In such a case, *the translational outcome would be decided by the position of the first functional (capable of initiation of translation) AUG or other translation initiation-competent codon. If it were in-frame with the protein-encoding information content of conventional mRNA, translation would result in a C-terminal fragment, CTF, of a conventionally encoded polypeptide.*

The scenario resulting in a CTF of conventionally encoded polypeptide as the translational outcome of mRNA amplification can be illustrated with the case of beta-amyloid overproduction in sporadic Alzheimer's disease. Beta-amyloid, A $\beta$ , the peptide associated with and, when overproduced, widely believed to have a pivotal early role in the etiology of Alzheimer's Disease (AD), is normally generated by proteolytic cleavages of a much larger molecule,  $\beta$  amyloid precursor protein,  $\beta$ APP [33-41]. Two sequential cleavages of  $\beta$ APP are involved in the production of A $\beta$ . The first is a cleavage of  $\beta$ APP by the  $\beta$



**Figure 11: RNA-dependent mRNA amplification can result in a 5'-truncated molecule encoding C-terminal fragment of a conventionally encoded polypeptide.** Boxed line: Sense strand RNA. Single line: Antisense strand RNA. "AUG"-functional translation initiation codon (could be other than AUG). "TCE"-3'-terminal complementary element; "ICE"-internal complementary element, both on the antisense RNA strand. Yellow circle – helicase/modifying activity complex. Blue lines (both single and boxed) – RNA strand modified and separated from its complement by a helicase complex. Red arrow – position of cleavage of the chimeric intermediate. Step 1: Synthesis of antisense strand; step 2: Strand separation; step 3: Folding of antisense strand into self-priming configuration; step 4: Extension of self-primed antisense RNA; step 5: Strand separation; step 6: Cleavage of the chimeric intermediate; step 7: End-products of RNA amplification. Steps 3'-7' correspond to steps 3-7. *Top panel:* Conventional, genome-transcribed mRNA molecule. *Middle panel:* Projected stages of RNA-dependent mRNA amplification. "ICE" is located within a segment of antisense RNA corresponding to the 5'UTR of conventional mRNA; the chimeric RNA end product contains the entire coding content of conventional mRNA. *Bottom panel:* "ICE" is located within a segment of antisense RNA corresponding to the coding region of conventional mRNA. The amplified chimeric end product contains a 5'-truncated coding region of conventional mRNA. The translational outcome is decided by position of the first functional translation initiation codon; if in-frame, a CTF of conventional polypeptide is produced.

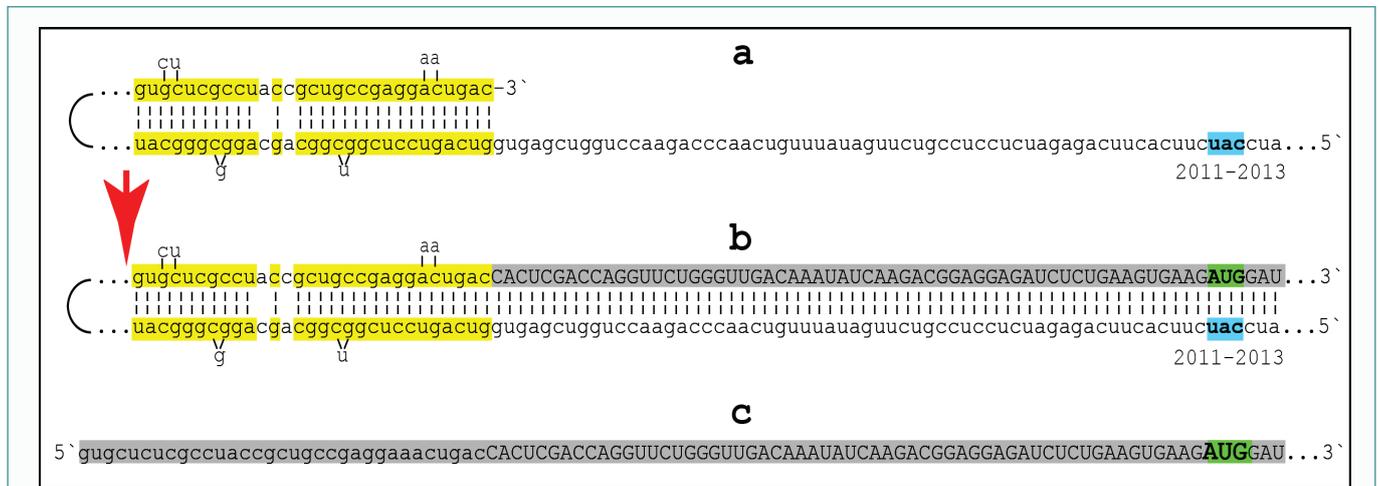
secretase enzyme. It occurs between residues 671 and 672 of the  $\beta$ APP molecule, generating the N-terminus of  $A\beta$ . The second cleavage, by  $\gamma$  secretase activity, generates the C-terminus of  $A\beta$  [34-36]. In inherited familial cases, constituting about 1% of the AD burden (the rest are sporadic cases) it was shown that the overproduction of  $A\beta$  is due to mutations around cleavage sites resulting in an abnormal proteolysis [42,43]. Since the pathological lesions and symptoms in the non-hereditary form of the disease are analogous to those seen in the familial forms, it has been assumed that abnormal proteolytic processing of  $\beta$ APP also underlies the pathogenesis of sporadic AD [37-39]. It was further assumed, therefore, that the suppression of APP proteolysis might alleviate the disease [44-46]. And, indeed,

inhibition of beta-secretase rescued functional impairments and actually reversed the disease in animal models bioengineered to imitate familial AD [47-50]. However, in several large human clinical trials such inhibition was completely inefficient in sporadic AD [27]. This suggested that, *in sporadic AD, beta-amyloid could be produced in the amyloid precursor protein-independent manner.* But how?

In  $\beta$ APP mRNA, the  $A\beta$ -coding segment is preceded immediately, contiguously and in-frame by the AUG codon normally encoding methionine in position 671 of the APP. If translation were initiated at this position, it would produce  $A\beta$ -containing CTF independently of  $\beta$ APP. Interestingly, the AUG in question is situated within a

nucleotide context optimal for the initiation of translation (an “A” in position –3 and a “G” in position +4 relative to the “A” of the AUG codon). In fact, of the twenty AUG codons encoding methionine residues in the  $\beta$ APP mRNA, only the AUG encoding Met671, not

[54], indicated the occurrence of topologically compatible TCE and ICE within the antisense strand of  $\beta$ APP mRNA and defined their sequences as well as the position of self-priming. Based on these results, the TCE/ICE-guided folding of the antisense strand of  $\beta$ APP mRNA [27,55-57] can be depicted as shown in Figure 12.



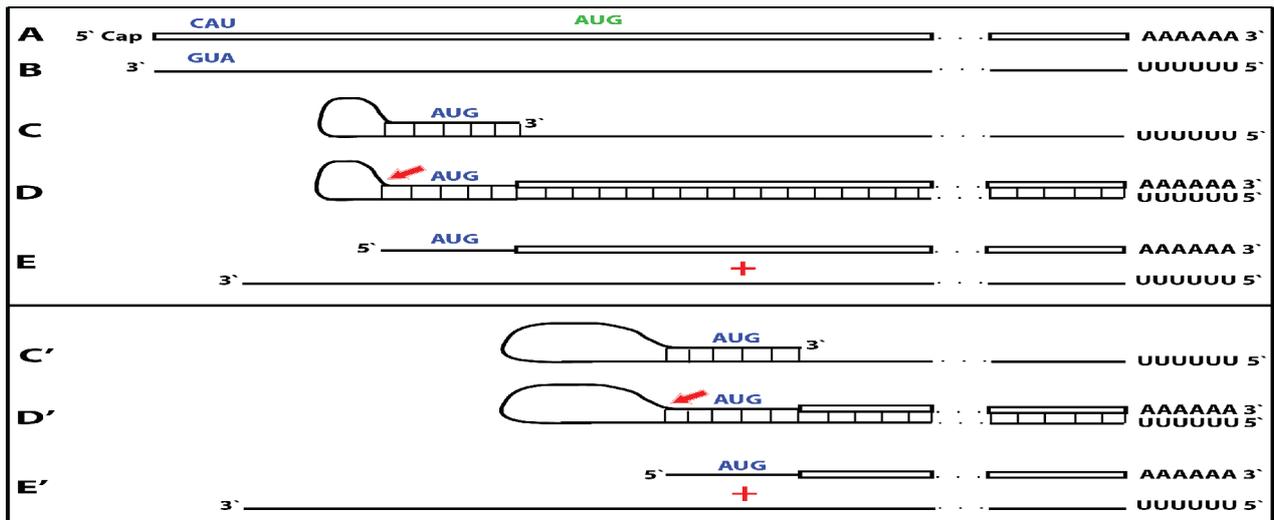
**Figure 12: Projected topology of RNA-dependent generation of 5'-truncated mRNA encoding beta-amyloid-containing C-terminal fragment of human amyloid precursor protein.** Lowercase letters-nucleotide sequence of the antisense RNA; uppercase letters-nucleotide sequence of the sense RNA. Double-stranded portions highlighted in yellow: TCE (top) and ICE (bottom) elements of the antisense RNA. Note that the TCE and the ICE are separated by about 2000 nucleotides. “2011-2013”: Nucleotide positions on the antisense RNA (starting from the complement of the AUG encoding Met1 of the APP) of the “uac” (highlighted in blue) corresponding to the “AUG” (highlighted in green) encoding Met671 in the APP mRNA. **a:** TCE/ICE-guided folding of the antisense APP RNA; 3'-terminal “c” corresponds to one of multiple transcription start sites of APP mRNA 149 nucleotides upstream from its AUG initiation codon [27]. **b:** Extension of self-primed antisense RNA into sense RNA and cleavage (red arrow; may also occur at one of the TCE/ICE mismatches), after strand separation, of the chimeric intermediate. **c:** Chimeric RNA end product contains 5' terminal antisense segment extending into severely 5'-truncated APP mRNA. Its translation initiates from the “AUG” (highlighted in green and encoding Met671 in conventional APP mRNA) immediately preceding the beta-amyloid-encoding region.

even Met1, is located within an optimal translation initiation context. Such favorable positioning of the AUG encoding Met671 was the basis for an early proposal that in AD,  $A\beta$  may be generated independently from  $\beta$ APP by the internal initiation of translation of the intact  $\beta$ APP mRNA [51]. The possibility of the internal initiation of translation has been subsequently ruled out by the elegant experiments of Citron and co-investigators [52]. There is, however, another possibility of utilization of the AUG in question as a translation initiation codon, namely the generation of a severely 5'-truncated  $\beta$ APP mRNA in which the AUG encoding Met671 in the intact mRNA becomes the first AUG codon.

To determine if an mRNA species of interest can potentially be a subject of RNA-dependent mRNA amplification, one needs to assess whether its antisense complement contains both TCE and ICE and is capable of folding into a self-priming configuration. If it were, the position of the ICE would indicate the possible translational outcome. Such an assessment can be conducted in a model experiment where an mRNA of interest serves as a template for synthesis of cDNA initiating at the 3'-terminal poly (A) and is subsequently removed by RNase H activity present in a preparation of reverse transcriptase used. If an mRNA were fully transcribed, if complementary elements were present within the antisense strand (cDNA), if one of them were 3'-terminal, and if they were topologically compatible, self-priming and the extension synthesis of a segment of the sense strand would occur. The junction between the antisense and sense components would define the site of self-priming and facilitate identification of the TCE and ICE. Just such an experiment was inadvertently carried out with human  $\beta$ APP mRNA [53]. The results of this experiment, misinterpreted and eventually dismissed by the authors as an artifact

Approximately a 30 nucleotide-long 3'-terminal sequence of the antisense strand of APP mRNA constitutes the TCE. Its counterpart, the ICE, is separated by nearly 2000 nucleotides, yet these elements are topologically compatible and the folding of the antisense molecule results in a self-priming configuration (Figure 12a). The TCE serves as a primer and is extended; thus, generating the sense strand as shown in the Figure 12b. Strands are then separated as illustrated in Steps 5' and 6' of Figure 11, and cleavage occurs either at the mismatches within the TCE or immediately upstream from it as indicated by the arrow in Figure 12b. The resulting chimeric RNA end product, shown in Figure 12c, consists of an antisense segment (TCE or its portion) continued into a sense-orientation molecule. The translational outcome is decided by the first, 5'-most, initiation-competent AUG codon. As shown in Figure 12b and c, the first AUG codon is located 58 nucleotides downstream from the TCE portion of the chimeric RNA end-product and it is, in fact, the AUG encoding Met671 in the intact APP mRNA! Translation from this position would produce beta amyloid-containing CTF in the APP-independent manner. The major prediction of such a mechanism is a complete inefficiency of beta secretase inhibition in sporadic Alzheimer's disease. This prediction was, in fact, borne out in several massive stage III clinical trials [27].

The above discussion considered RNA-dependent mRNA amplification scenarios where protein coding information is contained only in the sense RNA strand. There is, however, another possibility. If the TCE of an antisense RNA contains a functional translation initiation codon and if, after the cleavage of the chimeric intermediate, it was retained in the chimeric RNA end product, translation would be initiated from the initiation codon within the antisense portion of the chimeric RNA end product. In such a

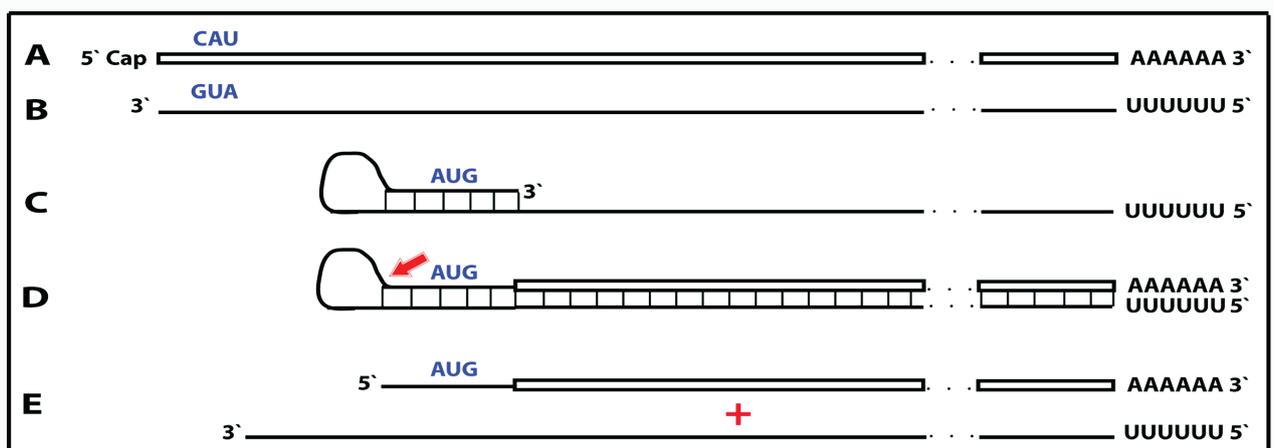


**Figure 13: RNA-dependent mRNA amplification process may enhance protein-encoding information content of a conventional mRNA and generate polypeptides non-contiguously encoded in the genome.** Boxed line: Sense strand RNA. Single line: Antisense strand RNA. "AUG": Functional translation initiation codon (could be other than AUG). 5'-CAU-3' on the sense RNA: Complement of 5'-AUG-3' (both in blue) on the antisense RNA. Red arrow: Position of cleavage of the chimeric intermediate. **A:** Conventional mRNA with translation initiation "AUG" in green. **B:** Antisense complement of conventional mRNA; 3'-GUA-5' (in blue) is the 5'-AUG-3'. **C:** Folding of the antisense strand into self-priming configuration. **D:** Extension of self-primed antisense strand into sense-oriented sequence followed by strand separation and cleavage of the chimeric intermediate. **E:** Chimeric RNA and 3'-truncated antisense RNA end products of RNA-dependent mRNA amplification. Note that translation of the chimeric end product starts from the antisense RNA- originated initiation codon ("AUG", in blue) and produces a chimeric polypeptide non-contiguously encoded in the genome. Steps **C'**, **D'**, and **E'** correspond to **C**, **D**, and **E**. *Top panel:* Antisense folding/self priming occurs within its segment corresponding to the 5'UTR of conventional mRNA. *Bottom panel:* Antisense folding/self priming occurs within its segment corresponding to the coding region of conventional mRNA.

case, if the folding/self-priming of the antisense RNA occurs within its segment corresponding to the 5'UTR of conventional mRNA (Figure 13, *top panel*) and if the initiation codon within the antisense portion of the chimeric RNA end product is in-frame with the mRNA coding sequence, translation will result in a *chimeric protein*. It will contain, in its entirety, the polypeptide encoded by a conventional mRNA, enhanced at its N-end by additional amino acids encoded by the antisense portion of the chimeric RNA end product and by a segment of the 5'UTR of conventional mRNA. Thus, interestingly, the protein-encoding information content of the chimeric RNA end product of amplification would be *non-contiguously* encoded in the genome. If, however, the folding/self-priming of the antisense strand

occurs within the segment corresponding to the coding region of conventional mRNA (Figure 13, *bottom panel*), and if the initiation codon within the antisense portion of the chimeric RNA end product is in-frame with the mRNA coding sequence, the translational outcome would be a CTF of the conventional polypeptide, enhanced at its N-end by additional amino acids encoded by the antisense portion of the chimeric RNA end product, also a chimeric protein non-contiguously encoded in the genome.

An interesting variant of the above scenario is one where an mRNA is programmably translated only under certain circumstances, for example a particular stress that activates the RNA-dependent



**Figure 14: RNA-dependent mRNA amplification process may activate a dormant protein-encoding information and generate polypeptides non-contiguously encoded in the genome.** Boxed line: Sense strand RNA. Single line: Antisense strand RNA. "AUG" (in blue): functional translation initiation codon (could be other than "AUG"). 5'-CAU-3' (in blue) on the sense RNA: Complement of 5'-AUG-3' on the antisense RNA. Red arrow: Position of cleavage of the chimeric intermediate. **A:** Genome-originated "dormant" mRNA containing protein-encoding information but lacking in-frame translation initiation codon. **B:** Antisense complement of conventional mRNA. Note that 3'-GUA-5' (in blue) is the 5'-AUG-3'. **C:** Folding of the antisense strand into self-priming configuration. **D:** Extension of self-primed antisense strand into sense-oriented sequence followed by strand separation and cleavage of the chimeric intermediate. **E:** End products of RNA-dependent mRNA amplification. Note that translation of the chimeric RNA end product starts from the antisense RNA-originated initiation codon ("AUG", in blue) and produces a chimeric polypeptide non-contiguously encoded in the genome.

mRNA amplification process. Under normal conditions, such RNA encodes a polypeptide but lacks a functional translation initiation codon in-frame with the encoded information. Since its potential to produce a protein is present but unrealized under regular circumstances, it can be considered a silent, or “dormant” mRNA. The potential to direct synthesis of a polypeptide can be realized, and a dormant mRNA activated, by the mRNA amplification process. In the chimeric pathway of mRNA amplification, the chimeric RNA end product contains a 5'-terminal segment contributed by the antisense RNA and not present in the conventional genome-encoded RNA molecule. If this additional segment contains a functional translation initiation codon (Figure 14) and if this codon were in-frame with the encoded information, the protein-encoding content of a conventional “dormant” mRNA would be activated. *The resulting translational outcome would be a chimeric protein consisting of a conventionally encoded polypeptide with the N-end encoded by the antisense RNA. Such a chimeric protein would be, therefore, non-contiguously encoded in the genome.*

### **The only translational outcome of the iPCR Tier of mammalian mRNA amplification is a polypeptide identical to that encoded by conventional mRNA**

As was mentioned above, whether or not the second Tier of amplification, iPCR, occurs, does not affect in any way the first Tier, a chimeric pathway. In turn, the translational outcome of the first, chimeric, Tier does not affect in any way the translational outcome of the second, iPCR, Tier. This is because, in sharp contrast to and irrespective of any of the potentially multiple translational outcomes of the chimeric pathway Tier, fully reviewed in [58], *there is only one possible translational outcome of the iPCR Tier of mRNA amplification – an intact conventional polypeptide, a phenomenon we refer to as “Two-Tier Paradox” [58].* This disparity reflects the asymmetry in the truncation of sense and antisense components of the chimeric intermediate during its cleavage at the conclusion of the Tier One. The source of the asymmetry is that the truncation of molecules in question is determined by two different, spatially independent, elements. The truncation of the sense strand component of the chimeric RNA end product, or, rather, the position of the initiation of its transcription from the antisense RNA template, is defined by the location of the ICE whereas the TCE determines the truncation of the antisense strand. Because the ICE can be anywhere in the antisense RNA molecule, so can the sense strand's 5' truncation be, and because the TCE, by definition, can be only 3'-terminal, the antisense always loses only a short segment at its 3'end.

This phenomenon is demonstrated in Figure 11. Whereas the sizes of chimeric RNA end products are drastically different in steps 7 and 7', the sizes of antisense RNA end products in the same steps are identical. The 3'-truncated antisense RNA, if polyadenylated at the conclusion of Tier One, gives rise in Tier Two to poly(U)-containing sense strand truncated at the 5'end by exactly the same extent as the 3' antisense truncation, which in most, if not all, instances is substantially shorter than a typical 5'UTR of a conventional mammalian mRNA [59]. Consequently, because the 5' poly(U)-containing sense RNA molecule, derived from the 3' poly(A)-containing antisense protagonist by the iPCR process, retains the bulk of its 5'UTR and the intact protein-encoding portion of a conventional mRNA, regardless of the translational outcome of the Tier One, in the Tier Two the translational outcome would always be the intact polypeptide, identical to that attained with a conventional progenitor mRNA. *The translational outcome of the Tier Two would match that of the Tier*

*One only when the ICE is positioned within the antisense RNA segment corresponding to the 5'UTR of conventional mRNA.* In other cases, the situation can be quite remarkable. It is conceivable, for example, that the outcome of Tier One is a CTF of a conventional protein whereas the Tier Two would produce, *in parallel*, the intact conventionally encoded polypeptide. The possible beta-APP-independent generation of beta-amyloid in sporadic Alzheimer's disease discussed above provides a good illustration of such a phenomenon. If the antisense beta-APP RNA strand, resulting from the cleavage of the chimeric intermediate and lacking only its TCE (or a fraction thereof), were 3'-polyadenylated in a cleavage-coupled manner and utilized as an iPCR template in the second Tier of the mRNA amplification process, it would give rise to a 5'-truncated and poly(U)-containing sense strand retaining the bulk of the 5'UTR and the complete coding region of the conventional beta-APP mRNA. Following its translation, an intact beta-APP polypeptide would be generated *alongside* its beta-amyloid-carrying CTF, produced from the 5'-truncated sense-strand in the proteolysis-independent manner following the first, chimeric pathway Tier of the mRNA amplification process.

### **Novel experimental strategies**

The detection of a full-size antisense RNA initiated at the 3'poly(A) of its conventional mRNA counterpart [6], and the identification of the putative end-product that accumulates to massive cellular levels, unprecedented for conventionally expressed mRNA, and matches the expected properties of a projected chimeric RNA end product, strongly support the mechanism for RdRp-mediated amplification of mRNA described above. The decisive evidence in support of the occurrence of this process is the detection, in the present study, of uniquely defined chimeric RNA intermediates containing antisense and sense sequences covalently joined in a rigorously predicted manner. Experimental approaches used for the detection of these sequences included multiple procedures that were safeguarded by numerous checkpoints. Yet, very strong arguments for the cellular origin of these chimeric RNA molecules notwithstanding, their possible inadvertent generation during experimental manipulations cannot be ruled out with absolute certainty. *A definitive proof for the occurrence of RNA-dependent amplification of mammalian mRNA, it appears, requires a direct measurement, without any outside experimental manipulations whatever, of a key cellular product of this process.*

A novel experimental strategy that would enable such direct measurement is suggested by a scenario for RNA-dependent mRNA amplification diagrammed in Figure 14 above. In this scenario, a latent mRNA contains “dormant” protein-coding information that cannot be translated for the lack of a functional in-frame initiation codon. It can, however, be translationally activated by RNA-dependent mRNA amplification, if the antisense portion of the chimeric RNA end product generated in the amplification process contains a functional translation initiation codon in-frame with the mRNA-encoded content. In such a case, *not only would a “dormant” mRNA be translated, but a distinctively defined chimeric polypeptide non-contiguously encoded in the genome would also be produced*, with the N-end portion encoded by the antisense RNA. The detection, by direct measurement in cell lysate, of such a molecule, the ultimate end product, would constitute a definitive and indisputable proof for the occurrence of RNA-dependent mRNA amplification. *This scenario can be enacted experimentally, as follows.*

The experimental design, outlined in Figure 15, starts with the elucidated chimeric pathway of beta-globin mRNA amplification shown in the bottom panel of Figure 2 and modifies it, by *in vivo*



of the antisense RNA (stage C) and its extension into chimeric intermediate followed by strand separation and cleavage (red arrow, stage D). Line E shows the projected edited chimeric RNA end product. In its antisense portion it contains an AUG codon (“aug” highlighted in green) in an optimal translation initiation context (purine in position -3 and “G” in position +4 relative to the “A” of the AUG) and in-frame with the INS/globin coding sequence. Translation of the edited chimeric RNA end product would result in a chimeric polypeptide shown in line F of the *middle panel* and comprised of a *globin portion* (highlighted in blue and identical to that shown in the *top panel*, line C), a *Tag peptide* (highlighted in grey), and the *N-end portion* encoded by the antisense RNA component of the chimeric RNA end product (highlighted in grey and underlined); it would be non-contiguously encoded in the genome. The detection of a Tag peptide would indicate that a chimeric polypeptide has been produced and the isolation and sequencing of this polypeptide would provide a definitive and irrefutable proof of the occurrence of the RNA-dependent mRNA amplification process.

Identification/sequencing of the projected chimeric polypeptide would be sufficient to make a determination of the occurrence of RNA-dependent mRNA amplification, but the above experimental design allows a number of additional controls. One of these is illustrated in the *bottom panel* of Figure 15. The AUG codon of the antisense strand can be mutated into a codon (ACA) not known to initiate translation, without changing the complementary relationship between the TCE and the ICE components of the antisense RNA. In such a case, the amplification would proceed similarly to that depicted in the *middle panel*, but the resulting chimeric RNA end product (*bottom panel*, line E) would lack a functional in-frame initiation codon and no translation of INS/globin sequence would occur. Other possible controls are frame-shifts of the edited RNA shown in the *middle panel* of Figure 15. Effecting a frame-shift of one, two but not three nucleotides in the region between the AUG of the antisense portion and the “INS” of the chimeric RNA end product (*middle panel*, line E) should abolish translation of the INS/globin whereas a frame-shift of either one or two or three nucleotides upstream from the AUG of the antisense RNA segment within the same molecule should have no such effect.

In the above experimental design, the edited gene would produce translationally “silent” mRNA. The protein-coding content could be activated in a rigorously predicted “enhanced” manner only when mRNA amplification process is induced during erythroid differentiation. This can be conveniently achieved within the framework of *in vivo* murine phenylhydrazine-driven/hemolytic anemia-induced splenic erythropoiesis. It takes place in a large, nearly homogeneous, organ, and its features and dynamics are well understood both in basic terms and in terms of RNA-dependent globin mRNA amplification. Beta-globin mRNA is much better suited for such an experiment than alpha-globin RNA because it has a longer 5'UTR and its antisense strand contains a conveniently positioned AUG codon (Figure 2, *bottom panel*, stages C and D) that can be incorporated in experimental design. Only one gene copy should be edited, making mice “heterogeneous” so as to interfere minimally with the physiological functions of erythroid cells. Circulating spleen-originated reticulocytes (over 90% of circulating red blood cells in this model system are reticulocytes) can be repeatedly collected from the same animals for time-points analyses prior to the collection and analysis of spleen cells. A possibility of non-specific side effects resulting from the administration of phenylhydrazine can

be addressed by parallel measurements in “edited” animals following the induction of hemorrhagic anemia shown to trigger, similarly to hemolytic anemia, globin mRNA amplification processes as described previously [6]. Moreover, whereas induced anemia-driven splenic erythropoiesis is indispensable for the analysis of RNA components and enzymatic activities involved in the RNA-dependent mRNA amplification process, as described below, the projected chimeric polypeptide non-contiguously encoded by the edited beta-globin gene, as discussed above (Figure 15), should be detectable, albeit at lower abundance, in erythroid cells of “edited” animals under physiologically normal, i.e. no induced anemia, conditions.

### Scope, significance and directions of future investigations

The results of the present study, which substantiate mammalian RNA-dependent mRNA amplification, were obtained *in vivo* with cells undergoing terminal erythroid differentiation and programmed for only a very short survival span. This raises a question of whether mammalian RNA-dependent mRNA amplification is a specialized occurrence limited to extreme circumstances of terminal differentiation or a general physiological phenomenon. This question was addressed in a companion study [26] by testing for the occurrence of RNA-dependent amplification of mRNA encoding extracellular matrix proteins abundantly produced throughout the tissue and organ development and homeostasis, an exceptionally revealing indicator of the range and scope of this phenomenon. The study [26] resulted in the detection of major identifiers of RNA-dependent amplification of mRNA encoding  $\alpha 1$ ,  $\beta 1$ , and  $\gamma 1$  chains of laminin in mouse tissues producing large quantities of extracellular matrix proteins, thus confirming the occurrence of mammalian RNA-dependent mRNA amplification and establishing it as a general physiological phenomenon.

The potential physiological, therapeutic and bioengineering significance of RNA-dependent mammalian mRNA amplification, with every genome-originated mRNA molecule acting as a possible template, can hardly be overstated. Malfunctions of this process may be involved in pathologies associated either with the deficiency of a protein normally produced by this mechanism or with the overproduction of a protein normally not involved in such a process. One of the experimental approaches to assess its involvement in such pathologies would be to interfere *in vivo* with the extent of complementarity between the two elements involved in the antisense RNA self-priming. In fact, multiple experiments of this sort have been carried out by nature. Thus, at least four different types of familial beta thalassemia, characterized by the reduced production of beta-globin chains, are associated with different point mutations in the 5'UTR of human beta-globin mRNA that clearly impede the amplification-associated complementary TCE/ICE relationship within the antisense beta-globin RNA and appear to solely account for the disease [64]. A possible example of an opposite impact, the mRNA amplification-driven pathology that is due to the abnormal overproduction of a polypeptide, is the increased generation of beta-amyloid in sporadic Alzheimer's disease. As discussed above, this can be a consequence of RNA-dependent amplification of a 3'-portion of the beta-APP mRNA encoding beta-amyloid-containing CTF of the conventional protein that would enable amyloid precursor protein-independent generation of beta-amyloid peptide [27,55-57]. The increased understanding of components and mechanisms involved in RNA-dependent mammalian mRNA amplification could open up new

ways and approaches not only for therapeutic interference in multiple pathologies but also for novel and powerful forms of bioengineering.

Multiple directions of investigation would be instrumental in bringing about this increased understanding. Virtually every step of RNA-dependent mRNA amplification discussed above, from the nature of RdRp activity and its postulated processivity conferring co-factor to strand-separating/modifying activity and the nature of nucleotide modifications, to chimeric intermediate-cleaving activity and the 3' polyadenylating activity coupled with it, needs to be further elucidated. Importantly, both genetic/molecular biology research and biochemical studies should be involved in these investigations. In the first category, a well targeted DNA editing and the assessment of its consequences in various experimental models could be crucial in eliciting additional information. In the second category, the choice of the starting material could be decisive in studying enzymatic activities involved in the amplification process and their networking and regulation. The first identification of mammalian RdRp activity capable of generating long antisense transcripts was made in and the activity was isolated from circulating reticulocytes produced during the hemolytic anemia-induced splenic erythropoiesis [2]. But now we understand that it was only a residual, remnant, activity carried over from the previous differentiation stage. We know this from the kinetics of the accumulation and subsequent loss of the putative chimeric RNA end product of globin mRNA amplification; its levels increase massively through the erythroblast stage, while still in spleen, but decline drastically when cells reach the reticulocyte stage and are released into circulation. Therefore, cells undergoing anemia-induced splenic erythropoiesis, in a nearly homogeneous large organ with over 90% of cells differentiating into erythroblasts, appear to be the most suitable experimental model for this line of inquiry and by far the best choice of starting material. Moreover, different stages of splenic erythropoiesis could be best suited to address diverse specific aspects of RNA-dependent mRNA amplification. For example, RdRp activity and its components could be best evaluated during later stages of splenic erythropoiesis whereas certain activities, such as 3' polyadenylating activity coupled with the cleavage of the chimeric intermediate and essential for the iPCR process, should be, for the reasons discussed above, investigated at the earlier stages of splenic erythroid differentiation.

## Conclusion

In conclusion, the results of the present study demonstrate the occurrence of RNA-dependent mRNA amplification and establish it as a new mode of genomic protein-encoding information transfer in mammalian cells. Its potential physiological impact is substantial and it appears relevant to multiple pathologies. Its understanding opens new venues of therapeutic interference, it suggests powerful novel experimental and bioengineering approaches, and its further rigorous investigations are highly merited.

## Materials and Methods

Balb/C mice, 2-4 month old, were rendered anemic by daily intraperitoneal injections of 0.1 ml of 0.8% neutralized phenylhydrazine as described [11]. After 6 days, 80-90% of the cells in circulating peripheral blood were reticulocytes and a similar proportion of spleen cells were erythroid, most having the characteristics of basophilic erythroblasts, consistent with previous observations [10,11]. Spleen and blood were collected on day seven as described [65]. For RNA preparation, minced spleen was passed sequentially through sieves of mesh 100, 150 and 200 or through 100

µm cell strainer (BD Falcon). Cells were collected by centrifugation and lysed in a buffer containing 30 mM Hepes, pH8.2; 50 mM NaCl; 5 mM MgCl<sub>2</sub>; 1% NP40; 10% sucrose; 5 mM DTT and 0.5 units/µl RNase inhibitor (Promega, NEB). After centrifugation, SDS and EDTA were added to 0.2% and 10 mM, respectively, and the lysate was extracted with phenol/chloroform with or without prior treatment with proteinase K. Precipitated RNA was dissolved, treated with DNase, phenol extracted and re-precipitated. Reticulocytes were lysed by incubation and occasional pipetting in 3 mM MgCl<sub>2</sub>; lysis was indicated by a change of coloration, at which point NaCl was added to 100 mM. After centrifugation, the lysate was adjusted to 100 mM Tris/HCl pH 7.5; 50 mM NaCl; 0.2% SDS and 10 mM EDTA and extracted as described above. Reticulocyte globin RNA was purified by oligo(dT) chromatography followed by high stringency hybridization with a mixture of immobilized α and β globin-specific probes (probes α3, α7, α10, β3, β7 and β10 listed below). For immobilization and use of oligonucleotides, an oxidizable solid support (Molecular Biosystems, San Diego) was employed following manufacturer's instructions [14,15]. RNA was resolved by electrophoresis on agarose-methylmercury [66,67] and the major RNA components were visualized by staining with ethidium bromide. The RNA of interest was electroeluted from slices of gel in electrophoresis running buffer in dialysis bags (Spectrapore CE, MWCO 3500). Slices of agarose-methylmercury gels were soaked in 0.1 M DTT for 30 min prior to electroelution [68]. Eluates were passed through Spin-X centrifugation units (Costar) and dialyzed against 1 mM EDTA, pH 6.2. To analyze nucleotide composition, 20 µg of gel-eluted pepRNA and of poly(A)-containing RNA eluted from a slice of the same gel, located between rRNA and pepRNA bands (control RNA), were dissolved in 20 µl of water, denatured by 3 min incubation at 100°C, adjusted to 10 mM ammonium acetate pH 5.3 and treated with nuclease P1 (EC 3.1.30.1; Sigma) for 2 hrs at 45°C [69]. Following P1 treatment, the reaction was adjusted to 0.1 M ammonium bicarbonate pH 7.9 and incubated with 2 milliunits of snake venom phosphodiesterase (EC 3.1.4.1; Sigma) at 37°C for 2 hrs [68]. Finally, 0.5 units of bacterial alkaline phosphatase (EC 3.1.3.1; Sigma) was added and incubation continued for 1 hr at 37°C [69]. The resulting digest was analyzed by reverse phase HPLC on a 2.1 x 250 mm C18 column (Vydec) with a 2.1 x 150 mm silica gel guard column at a flow rate of 1.5 ml/min. Gradient conditions were as described [70]. Further analysis of the eluted fractions was carried out on ABI Voyager MALDI-tof mass spectrometer using 2,5-dihydroxybenzoic acid as the matrix. HPLC and mass-spectrometric analyses were performed by Analytical Biotechnology Services, Boston. For oligonucleotide-mediated RNase H cleavage (RNase H mapping), pepRNA and reticulocyte globin RNA were labeled at either 5' or 3' termini. 5' labeling was carried out with polynucleotide kinase and γ<sup>32</sup>P ATP. Prior to phosphorylation, reticulocyte globin RNA was treated with tobacco acid pyrophosphatase to de-cap RNA molecules and with calf intestinal phosphatase to remove the remaining phosphate. No preliminary treatment was required for 5'-labeling of pepRNA; labeling reactions with both RNA preparations proceeded with similar efficiency. 3'-labeling was carried out by ligation with <sup>32</sup>P-labeled pCp and, unlike 5' labeling, proceeded much slower with pepRNA than with reticulocyte globin RNA or with gel-eluted conventional globin RNA from anemic spleen. Labeled RNA was treated with proteinase K, phenol extracted, precipitated and subjected to electrophoresis on 8% polyacrylamide-urea gels. RNA was visualized by autoradiography, eluted by overnight incubation in ammonium acetate 0.5M, pH 7.2; SDS 0.1%; EDTA 1 mM; tRNA 20 µg/ml, extracted with phenol/chloroform and precipitated with

ethanol. For RNase H cleavage, 10-50 ng aliquots of RNA were mixed with 100-fold excess of oligonucleotide complementary to different segments of either  $\alpha$  or  $\beta$  globin RNA and with 2  $\mu$ g of unrelated carrier RNA, incubated in 1mM EDTA pH 6.2 for 3 min at 95°C, cooled on ice, adjusted to 40 mM Tris pH 8.0, 7 mM MgCl<sub>2</sub>, 150 mM NaCl, and prehybridized for 30 min at 37°C. Reactions were started by addition of BSA to 50  $\mu$ g/ml, DTT to 5 mM, RNase inhibitor to 1 unit/ $\mu$ l and RNase H and were continued for 60 min at 37°C. RNase H was used at 10-30 milliunits per reaction with globin RNA and 100-500 milliunits per reaction with pepRNA. Following cleavage, the RNA was treated with proteinase K, phenol extracted, precipitated and analyzed by electrophoresis on 8% polyacrylamide-urea gels. When cleavage resulted in a short fragment, samples were analyzed directly, without proteinase K treatment and extraction. The following oligonucleotides were used in RNase H mapping (nucleotide numbers in parenthesis indicate location of oligonucleotide termini on corresponding mRNA):

- $\alpha$ 1: GAAGAAGGGCATGGCCAGAAGG (506-485);
- $\alpha$ 2: ACGGTACTTGGAGGTCAGCACG (461-440);
- $\alpha$ 3: GGCATGTACCGCGGGGGTG (407-389);
- $\alpha$ 4: GTGGTGGCTAGCCAAGGTCAC (377-357);
- $\alpha$ 5: AAGGTCACCAGCAGGCAGTGG (364-344);
- $\alpha$ 6: CCAGAGAGCACCATGGTTTCTTCCT (49-25);
- $\alpha$ 7: GGCCACCAATCTTCCCCCAG (96-77);
- $\alpha$ 8: AGCTCCATATTCAGCACCATGGCC (116-93);
- $\alpha$ 9: GGTCTTGGTGGTGGGAAGCT (161-141);
- $\alpha$ 10: GATCCACACGCAGCTTGTGGG (321-301);
- $\alpha$ 11: TCGGCAGGTTGGTGGCTAGCC (385-365);

- $\beta$ 1: AACAACTGACAGATGCTCTCTTGGGAA (565- 539);
- $\beta$ 2: GGGGGTTTGTAGTGGTACTTGTGAGCC (504-480);
- $\beta$ 3: GGCAGCCTGTGCAGCGGG (443-426);
- $\beta$ 4: TCCTTGCCAAGGTGGTGGC (419-401);
- $\beta$ 5: GCTGTCCAAGTGATTCAGGCCATC (297-274);
- $\beta$ 6: ATGATGTCTGTTTCTGGGGTTGTGAGT (56-30);
- $\beta$ 7: GGTGCACCATGATGTCTGTTTCTGG (64-40);
- $\beta$ 8: CCTTCCCCACAGGCAAGAGACAG (109-86);
- $\beta$ 9: CCAGGGCCTCACCACCAAC (142-124);
- $\beta$ 10: GGCTGGCAAAGGTGCCCTTGAGG (322-300).

For cell-free translation, pepRNA and regular globin RNA from anemic spleen were eluted from agarose-methylmercury gels as described above. Regular globin RNA was further purified and its traces were removed from pepRNA by high stringency hybridization of both preparations with a mixture of immobilized  $\alpha$  and  $\beta$  globin-specific probes ( $\alpha$ 3,  $\alpha$ 7,  $\alpha$ 10,  $\beta$ 3,  $\beta$ 7 and  $\beta$ 10). Reticulocyte globin RNA, similarly purified, was also used in this assay. The translation was carried out using wheat germ extract (Promega) following the manufacturer's instructions in the presence of methylmercury hydroxide as described [71]. The translation products were analyzed on 15% polyacrylamide-SDS gels (Bio-Rad). Immunoprecipitation of translation products was carried out as described [72] using sheep

antisera against adult murine globins (courtesy of Dr. Jeffrey Ross, Univ. of Wisconsin). For Next Generation Sequencing, cytoplasmic RNA from anemic spleen was used for preparation of sequencing libraries. Two library preparation kits were employed: TruSeq RNA/V2 (Illumina) and NEBNext Ultradirectional RNA (NEB); both kits were used following manufacturers' instructions with minor modifications. Libraries obtained were analyzed on a MiSeq sequencer (Illumina) as directed by manufacturer. The resulting reads were either aligned with references containing sequences of chimeric junctions or blasted against sequences of chimeric junctions and other sequences of interest at the stringency word 7. Following the analysis of reads, fragments of interest were extracted from raw data and analyzed.

## Authors Contributions

VV and SR conceived and/or developed all concepts described in the present study. SR and VV designed and carried out, with the help of individuals listed in the acknowledgements section, all experiments described above. VV wrote the manuscript; SR and BO participated in the writing of the manuscript. BO hosted the later stages of the study. SR and BO contributed to discussions, analysis and interpretation of the results.

## Funding

NIH R21 GM056179 to Vladimir Volloch and NIH R01 AR036820 to Bjorn R. Olsen.

## Acknowledgements

Authors are grateful to Philipp Kapranov for insightful and stimulating discussions, to Bruce Schweitzer and Paul Leavis for experimental help, to Matt Vivero, Ugur Ayturk and Nam Pho for help with the analysis of sequencing data, to Matt Warman for making his resources available, and to Jeff Ross for gift of antibodies.

## References

- Volloch V. Cytoplasmic Synthesis of Globin RNA in Differentiated Murine Erythroleukemia Cells: Possible Involvement of RNA-dependent RNA Polymerase. *Proc. Natl. Acad. Sci. USA*. 1986;83(5):1208-12.
- Downey KM, Byrnes JJ, Jurmark BS, So AG. Reticulocyte RNA-dependent RNA polymerase. *Proc Natl Acad Sci U S A*. 1973;70(12):3400-4.
- Volloch V, Schweitzer B, Rits S. Synthesis of globin RNA in enucleated differentiating murine erythroleukemia cells. *J Cell Biol*. 1987;105(1):137-43.
- Richards OC, Ehrenfeld E. Poliovirus RNA replication. *Curr Top Microbiol Immunol*. 1990;161:89-119.
- Volloch V, Schweitzer B, Rits S. Evolutionarily Conserved Elements in the 5'-untranslated Region of beta Globin mRNA Mediate Site-specific Priming of a Unique Hairpin Structure during cDNA Synthesis. *Nucleic Acids Res*. 1994;22(24):5302-9.
- Volloch V, Schweitzer B, Rits S. Antisense Globin RNA in Murine Erythroid Tissues: Structure, Origin and Possible Function. *Proc Natl Acad Sci USA*. 1996;93:2476-81.
- Volloch V. Determination and characterization of RNA termini by ligation-mediated RNA amplification. In: *PCR-Essential Techniques*. 1997; BIOS, Oxford, J. Burke, editor.
- Kapranov P, Ozsolak F, Kim SW, Foissac S, Lipson D, Hart C, et al. New class of gene-termini-associated human RNAs suggests a novel RNA copying mechanism. *Nature*. 2010;466:642-6.
- Jain SK, Subrahmanyam D. On the mechanism of phenylhydrazine-induced hemolytic anemia. *Biochem Biophys Res Commun*. 1978;82(4):1320-4.
- Conkie D, Kleiman L, Harrison PR, Paul J. Increase in the accumulation of globin mRNA in immature erythroblasts in response to erythropoietin *in vivo* or *in vitro*. *Exp Cell Res*. 1975;93(2):315-24.

11. Bastos RN, Volloch Z, Aviv H. Messenger RNA population analysis during erythroid differentiation: a kinetical approach. *J Mol Biol.* 1977;110(2):191-203.
12. Wojda U, Noel P, Miller JL. Fetal and adult hemoglobin production during adult erythropoiesis: coordinate expression correlates with cell proliferation. *Blood.* 2002;99(8):3005-13.
13. Hundley HA, Bass BL. ADAR editing in double-stranded UTRs and other noncoding RNA sequences. *Trends Biochem Sci.* 2010;35(7):377-83.
14. Volloch V, Schweitzer B, Rits S. Uncoupling of the synthesis of edited and unedited COIII RNA in *Trypanosoma brucei*. *Nature.* 1990;343(6257):482-4.
15. Volloch V, Schweitzer B, Rits S, Zhang X. Identification of negative-strand complements to cytochrome oxidase subunit III RNA in *Trypanosoma brucei*. *Proc Natl Acad Sci U S A.* 1991;88(23):10671-5.
16. Chomczynsky P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem.* 1987;162(1):156-9.
17. Chomczynsky P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction: twenty something years on. *Nat Protoc.* 2006;1(2):581-5.
18. Lai M. RNA replication without RNA-dependent RNA polymerase: surprises from hepatitis delta virus. *J Virol.* 2005;79(13):7951-8.
19. Taylor J. Replication of human hepatitis delta virus: recent developments. *Trends Microbiol.* 2003;11:185-90.
20. Lehmann E, Brueckner F, Cramer P. Molecular basis of RNA-dependent RNA polymerase II activity. *Nature.* 2007;450(7168):445-9.
21. Wagner SD, Yakovchuk P, Gilman B, Ponicsan SL, Drullinger LF, Kugel JF, et al. RNA polymerase II acts as an RNA-dependent RNA polymerase to extend and destabilize a non-coding RNA. *EMBO J.* 2013;32(6):781-90.
22. Maida Y, Yasukawa M, Masutomi K. *De novo* RNA Synthesis by RNA-Dependent RNA Polymerase Activity of Telomerase Reverse Transcriptase. *Mol Cell Biol.* 2016;36(8):1248-59.
23. Tseng CH, Lai MM. Hepatitis delta virus RNA replication. *Viruses.* 2009;1(3):818-31.
24. Brazas R, Ganem D. A cellular homolog of hepatitis delta antigen: implications for viral replication and evolution. *Science.* 1996;272(5284):90-4.
25. Huang CR, Lo SJ. Evolution and diversity of the human hepatitis d virus genome. *Adv Bioinformatics.* 2010.
26. Volloch V, Rits S, Olsen B. RNA-dependent amplification of mammalian mRNA encoding extracellular matrix components: identification of chimeric RNA intermediates for  $\alpha 1$ ,  $\beta 1$ , and  $\gamma 1$  chains of laminin. *Ann Integr Mol Med.* 2019;1(1):1004.
27. Volloch V, Rits S. Results of beta secretase-inhibitor clinical trials support amyloid precursor protein-independent generation of beta amyloid in sporadic Alzheimer's disease. *Med Sci (Basel).* 2018;6(2):pii: E45.
28. Clissold PM, Arnstein HR, Chesterton CJ. Quantitation of globin mRNA levels during erythroid differentiation in the rabbit. *Cell.* 1977;11(2):353-61.
29. <https://bscb.org/learning-resources/softcell-e-learning/ribosome/>
30. Rabouille C. Pathways of Unconventional Protein Secretion. *Trends Cell Biol.* 2017;27(3):230-40.
31. Yoshida H, Matsui T, Yamamoto A, Okada T, Mori K, et al. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell.* 2001;107(7):881-91.
32. Jiang S, Zhang E, Zhang R, Li X. Altered activity patterns of transcription factors induced by endoplasmic reticulum stress. *BMC Biochem.* 2016;17:8.
33. Iizuka T, Shoji M, Kawarabayashi T, Sato M, Kobayashi T, Tada N, et al. Intracellular generation of amyloid  $\beta$ -protein from amyloid  $\beta$ -protein precursor fragment by direct cleavage with  $\beta$ - and  $\gamma$ -secretase. *Biochem. Biophys. Res. Commun.* 1996;218(1):238-42.
34. De Strooper B, Annaert W. Proteolytic processing and cell biological functions of the amyloid precursor protein. *J Cell Sci.* 2000;113(11):1857-70.
35. Barber RC. The genetics of Alzheimer's disease. *Scientifica (Cairo).* 2012;2012:246210.
36. Beyreuther K, Masters CL. Amyloid precursor protein (APP) and B2A4 amyloid in the etiology of Alzheimer's disease: Precursor-product relationships in the derangement of neuronal function. *Brain Pathol.* 1991;1(4):241-51.
37. Hardy J, Allsop D. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. *Trends Pharmacol Sci.* 1991;12(10):383-8.
38. Selkoe DJ. The molecular pathology of Alzheimer's disease. *Neuron.* 1991;6(4):487-98.
39. Hardy JA, Higgins GA. Alzheimer's disease: the amyloid cascade hypothesis. *Science.* 1992;256(5054):184-5.
40. Hardy J, Selkoe DJ. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science.* 2002;297(5580):353-6.
41. Dyrks T, Dyrks E, Mönning U, Urmoneit B, Turner J, Beyreuther K. Generation of beta A4 from the amyloid protein precursor and fragments thereof. *FEBS Lett.* 1993;335(1):89-93.
42. Haass C, Lemere CA, Capell A, Citron M, Seubert P, Schenk D, et al. The Swedish mutation causes early-onset Alzheimer's disease by beta-secretase cleavage within the secretory pathway. *Nat Med.* 1995;1(12):1291-6.
43. Hunter S, Brayne C. Understanding the roles of mutations in the amyloid precursor protein in Alzheimer disease. *Mol Psychiatry.* 2018;23(1):81-93.
44. Vassar R, Bennett BD, Babu-Khan S, Kahn S, Mendiaz EA, Denis P, et al. Beta-secretase cleavage of Alzheimer's amyloid precursor protein by the transmembrane aspartic protease BACE. *Science.* 1999;286(5440):735-41.
45. Hussain I, Powell D, Howlett DR, Tew DG, Meek TD, Chapman C, et al. Identification of a novel aspartic protease (Asp 2) as beta-secretase. *Mol Cell Neurosci.* 1999;14(6):419-27.
46. Vassar R. BACE1 inhibitor drugs in clinical trials for Alzheimer's disease. *Alzheimers Res Ther.* 2014;6(9):89.
47. Stamford AW, Scott JD, Li SW, Babu S, Tadesse D, Hunter R, et al. Discovery of an orally available, brain penetrant BACE1 inhibitor that affords robust CNS A $\beta$  reduction. *ACS Med Chem Lett.* 2012;3(11):897-902.
48. Keskin AD, KekuÅ; M, Adelsberger H, Neumann U, Shimshek DR, Song B, et al. BACE inhibition-dependent repair of Alzheimer's pathophysiology. *Proc Natl Acad Sci U S A.* 2017;114(32):8631-6.
49. Neumann U, Rueeger H, Machauer R, Veenstra SJ, Lueoend RM, Tintelnot-Blomley M, et al. A novel BACE inhibitor NB-360 shows a superior pharmacological profile and robust reduction of amyloid- $\beta$  and neuroinflammation in APP transgenic mice. *Mol Neurodegener.* 2015;10:44.
50. Hu X, Das B, Hou H, He W, Yan R. BACE1 deletion in the adult mouse reverses preformed amyloid deposition and improves cognitive functions. *J Exp Med.* 2018;215(3):927-40.
51. Breimer LH, Denny P. Alzheimer amyloid aspects. *Nature.* 1987;326(6115):749-50.
52. Citron M, Haass C, Selkoe DJ. Production of amyloid-beta-peptide by cultured cells: no evidence for internal initiation of translation at Met596. *Neurobiol Aging.* 1993;14(6):571-3.
53. Mita S, Sadlock J, Herbert J, Schon EA. A cDNA specifying the human amyloid beta precursor protein (ABPP) encodes a 95-kDa polypeptide. *Nucleic Acids Res.* 1988;16(19):9351.
54. Mita S, Sadlock J, Herbert J, Schon EA. A cDNA specifying the human amyloid beta precursor protein (ABPP) encodes a 95-kDa polypeptide. *Nucleic Acids Res.* 1988;16:11402.
55. Volloch V. A mechanism for  $\beta$ -amyloid overproduction in Alzheimer's disease: Precursor-independent generation of  $\beta$ -amyloid via antisense RNA-primed mRNA synthesis. *FEBS Lett.* 1996;390:124-8.
56. Volloch V. Mechanism for  $\beta$ -amyloid overproduction in Alzheimer's Disease: Possible antisense RNA-mediated generation of a 5'-truncated  $\beta$ APP mRNA encoding 12

- kDa C-terminal fragment of  $\beta$ APP, the immediate precursor of A $\beta$ . In *Molecular Mechanisms of Dementia*. 1997; Wasco, W., Tanzi, R., Eds.; Humana Press.
57. Volloch V. Possible mechanism for resistance to Alzheimer's disease (AD) in mice suggests new approach to generate a mouse model for sporadic AD and may explain familial resistance to AD in man. *Exp. Neurobiol.* 1997;144: 214-8.
58. Volloch V. Protein-encoding RNA to RNA information transfer in mammalian cells: principles of RNA-dependent mRNA amplification. *Ann Integr Mol Med.* 2019;1(1):1002.
59. Lodish H, Berk A, Zipursky SL. In: *Molecular Cell Biology*. 2000, New York, Freeman and Co.
60. Juba A, Chaput J, Wellenslek B. Exploring the Role of AUG Triplets in Human Cap-Independent Translation Enhancing Elements. *Biochem.* 2018;57: 6308-18.
61. Kearse MG, Wilusz JE. Non-AUG translation: a new start for protein synthesis in eukaryotes. *Genes Dev.* 2017;31(17):1717-31.
62. Varshavsky A. The N-end rule pathway of protein degradation. *Genes Cells.* 1997;2(1):13-28.
63. Giglione C, Boularot A, Meinnel T. Protein N-terminal methionine excision. *Cell Mol Life Sci.* 2004;61(12):1455-74.
64. Cao A, Galanello R. Beta-thalassemia. *Genet Med.* 2010;12(2):61-76.
65. Aviv H, Volloch Z, Bastos R, Levy S. Biosynthesis and stability of globin mRNA in cultured erythroleukemia Friend cells. *Cell.* 1976;8(4):495-503.
66. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, et al. *Current protocols in molecular biology*. 1987; Wiley publishing.
67. Sambrook J, Fritsch E, Maniatis T. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory press. 1989.
68. Lemischka IR, Farmer S, Racaniello VR, Sharp PA. Nucleotide sequence and evolution of a mammalian alpha-tubulin messenger RNA. *J Mol Biol.* 1981;151(1):101-20.
69. Crain PF. Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol.* 1990;193:782-90.
70. Pomerantz SC, McCloskey JA. Analysis of RNA hydrolyzates by liquid chromatography-mass spectrometry. *Methods Enzymol.* 1990;193:796-824.
71. Payvar F, Schimke RT. Methylmercury hydroxide enhancement of translation and transcription of ovalbumin and conalbumin mRNAs. *J Biol Chem.* 1979;254(16):7636-42.
72. Harlow E. and Lane D. *Antibodies: a laboratory manual*. 1988, Cold Spring Harbor Laboratory press.
73. Rits S, Olsen B, Volloch V. RNA-dependent synthesis of mammalian mRNA: Identification of chimeric intermediate and putative end-product. *BioRxiv*, 2016.