

Folded RNA from an intron of one gene might inhibit expression of a counteracting gene

Clark Jeffries^{1,2,3*}, Michael Jarstfer² and Diana O. Perkins¹

¹ Department of Psychiatry, University of North Carolina at Chapel Hill, USA

² School of Pharmacy, University of North Carolina at Chapel Hill, USA

³ Renaissance Computing Institute, USA

E-mail: clark_jeffries@med.unc.edu; diana_perkins@med.unc.edu; jarstfer@email.unc.edu

* Corresponding author

Edited by E. Wingender; received April 27, 2005; revised June 15; accepted June 18, 2005; published July 03, 2005

Abstract

Homeostatic maintenance of mRNA levels including prompt availability of mRNAs for translation in response to changing protein demands might be partly enabled by a system of combinatorial controls involving noncoding RNA blocking agents. This article proposes a specific version of that control mechanism, namely, a double-stranded RNA folding from transcription of an intron of one gene might and leading to an agent that inhibits mRNA of a counteracting gene. Thus transcription of the first gene would automatically repress translation of the second.

On the basis of a bioinformatics search, we suggest a possible example, namely, that pro-apoptosis gene *PAR4* might inhibit anti-apoptosis gene *XIAP*. Part of an intron from *PAR4* folds to form a large, stable hairpin, and reverse complement of the hairpin stem (with ~280 nucleotides) matches a sub-sequence of an exon in *XIAP*. This would be part of an efficient system to drive initiated apoptosis to its conclusion. Figuratively speaking, it replaces two control knobs with one.

Since repeats, some with many thousands of copies, occur throughout the genome with complex distributions, care must be taken before asserting that the presence of any repeat in any gene is significant. Our apoptosis example involves repeats, so experimental verification is needed and planned. However, if it is found that the noncoding RNA by-product of one gene folds into a hairpin that is processed into an agent that inhibits a counteracting gene, then the same type of control unit might be found extensively among counteracting families of genes of many types.

Keywords: counteracting genes, RNA hairpins, noncoding RNA, miRNA, Alu repeat

Introduction and related work

Noncoding RNA (ncRNA) is RNA in a product of transcription that does not subsequently appear in mRNA that is eligible for translation. Many recent publications have greatly expanded the proposed and sometimes observed roles of ncRNA in regulation of gene expression. Seminal papers including [1, 2], and many others have led to scrutiny of dsRNAs from ncRNAs as translation-blocking or mRNA-cleaving agents supported in RNA Induced Silencing Complexes (RISCs).

One proposed mechanisms for ncRNA control of gene expression involves a transcribed sequence of perhaps ~100 nts that contains a sub-sequence and an approximation of the reverse complement (revcom) of the sub-sequence. Such a

microRNA (miRNA) strand folds into a hairpin, hammerhead, or more complex but stable molecule of dsRNA that can then be exported into the cytosol. The dsRNA stem is recognised, trimmed, and unzipped, and then from one side, a short "mature" sub-sequence (~20 nt) is selected as a revcom pattern for binding to a suitable portion of a target mRNA. The RISC might cause cleavage of the mRNA or might inhibit translation (see [3] and companion papers in the same issue of Nature).

Regulatory ncRNAs that are miRNAs were first described in *C. elegans* [4]. About a quarter of the ~227 known miRNAs (<http://www.sanger.ac.uk/Software/Rfam/mirna/>) are actually excised from the introns of mRNAs. The distributions of miRNAs in the genome are suggestive of some kind of system of orchestrated control, but needed is an understanding of its organizational principles. In some cases, several different miRNAs target the same mRNA, and in some cases several mRNAs are targeted by the same miRNA; the complex graph of such relationships would seem to be part of a system of "combinatorial control" [5].

Another type of gene regulation might involve repeats. Alu repeats are ~300 nt in length and are characteristic of primates. Some particular families of Alu repeats occur in much greater numbers in humans versus chimpanzees (e. g. 2200 versus nine [6]). This has been advanced as a partial explanation for the biomedical, morphological, and cognitive differences that distinguish humans. That is, Alu repeat differences are much in contrast to the commonality (99%) of DNA inventories of humans and chimpanzees.

Shankar *et al.* [7] suggest some regulatory importance of one type of Alu repeat that, through retrotransposition, could distribute functional and regulatable promoter elements. If the mutation endowed the organism with an advantage, then it might be stabilized in the genome. Exaptation of regulatory elements in preexisting genes through Alu insertions could accelerate the evolution of primate genomes.

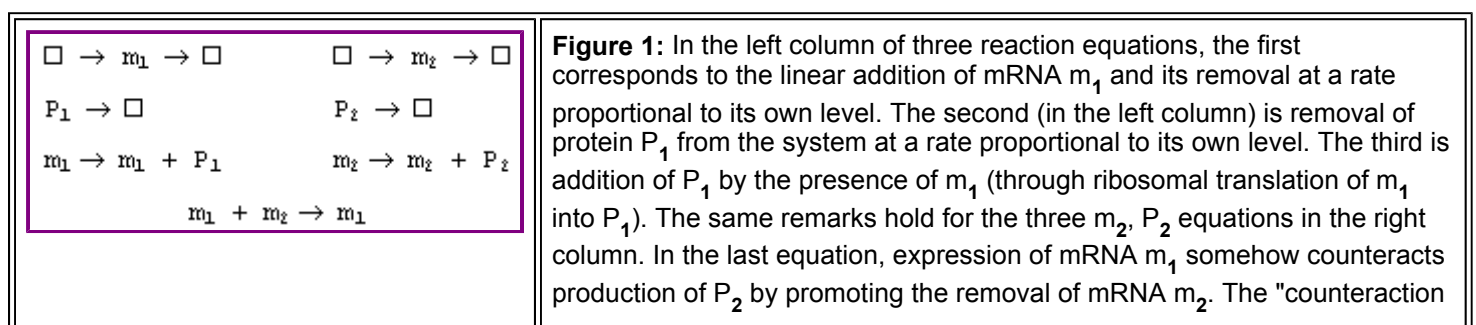
Somewhat related instances of intron folding appear in Group II introns of some fungi, plants, protists, and bacteria, but not in humans [8]. However, Group II editing of introns (including "twintrons" or introns within introns) is suggestive of the present hypothetical mechanism.

Supplementing the above instances of regulation involving repeats and revcoms, we suggest the folding of a ~600 nt sub-sequence from an transcriptional product intron to become eventually an inhibitory agent relative to a counteracting gene. That is, we consider how in one intron a repeat of the size of an Alu and its nearby revcom could inhibit expression of a counteracting gene. As a particular example, we offer a bioinformatics investigation of such a structure in an intron of the pro-apoptosis gene *PAR4* relative to the anti-apoptosis gene *XIAP*.

Reaction network

Miklos and Maleszka [9] have suggested that complex diseases might be understood through evaluation of network fluxes, network robustness, network topology, and other concepts of networks and dynamical systems.

In that spirit, let us suppose a first transcript encodes a first protein and a second transcript encodes a second protein. As depicted in Figure 1, a model reaction network can be devised that represents the two, initially independent systems. The square symbol denotes a source or a sink outside the reaction network.



"mechanism" is represented by the last reaction and is the subject of this paper.

Such a counteraction network would only be regarded as an approximation of a portion of a real network of biochemical pathways. Nonetheless, already interesting model dynamics can be observed. The reaction equations can be translated into a stoichiometric dynamical system as follows.

$$\begin{aligned} dP_1/dt &= c_1 * m_1 - e_1 * P_1 \\ dm_1/dt &= i_1 - s_1 * m_1 \\ dP_2/dt &= c_2 * m_2 - e_2 * P_2 \\ dm_2/dt &= i_2 - s_2 * m_2 - r * m_1 * m_2 \end{aligned}$$

The real number variables P_1 and m_1 are the concentrations in a cell of a first protein P_1 synthesized from a first mRNA m_1 at the rate $c_1 * m_1$ (so c_1 is a rate constant in appropriate units). P_1 is also assumed to be eliminated by removal from the system at the rate $-e_1 * P_1$, so e_1 is another rate constant. The induction rate of m_1 is i_1 , and m_1 is also self-regulating by the term, $-s_1 * m_1$. All the analogous definitions hold for variables with subscript 2, and in addition m_1 (as a transcription by-product) also catalyzes the removal of m_2 at a rate $-r * m_1 * m_2$ (the intron-hairpin counteraction).

Suppose initially $i_1 = 0$ and that there is no m_1 or P_1 in the system. The system with only P_2 and m_2 exists in equilibrium. Then suppose i_1 becomes a positive constant comparable to i_2 . Qualitatively the corresponding dynamics appear as in [Figure 2](#).

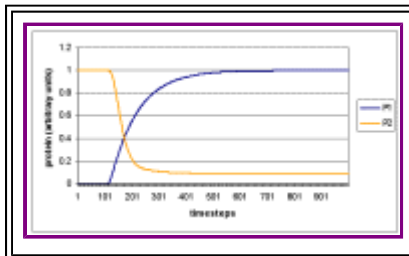


Figure 2: A model of the effect of assuming the transcription of one gene includes an intron product that automatically inhibits transcription of a counteracting gene. As expected, the introduction of transcript of the first gene not only causes the production of protein P_1 but drives the production of protein P_2 to a lower level. If the reaction rate of inhibition is increased, then the final state of P_2 will be decreased. See text for details of numerical simulation.

The introduction of m_1 removes some m_2 and so m_2 and P_2 are driven from high to low levels. In the figure the horizontal axis is the number of time steps with a time increment of 0.1 units. The reaction rates used with $i = 1$ or 2 are $c_i = e_i = r = 1$ and $i_2 = s_i = 0.1$. For the first 100 timesteps $i_1 = 0$, then $i_1 = 0.1$. The vertical axis shows the relative values of P_1 and P_2 in arbitrary units.

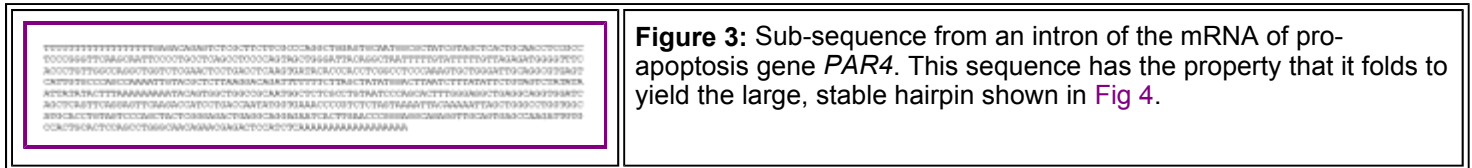
With a suitable choice of reaction rates, the new level of P_2 can be arbitrarily low; namely, the larger the counteraction rate r , the more dramatic the reduction in final P_2 .

A possible example based on bioinformatics investigation

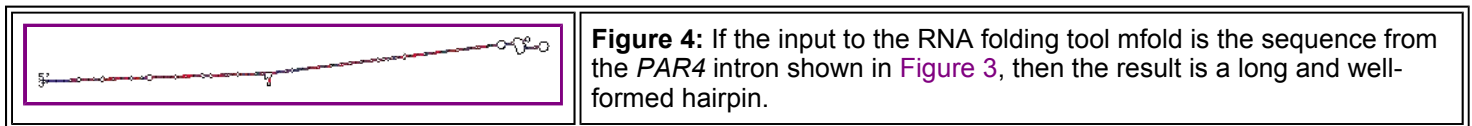
Apoptosis is a critical, highly regulated event, essential to organism development and homeostasis. Both pro- and anti-apoptotic proteins are expressed in cells, and it is the relative abundance of multiple compounds that determines whether a cell lives or dies. One important pro-apoptotic gene is *PAR4* (alias *PAWR*, described in [Entrez Gene](#)). A member of the class of inhibitor of apoptosis (IAP) genes is *XIAP* (on chromosome X). IAP genes are said to be "probably the most important intrinsic regulators of apoptosis" [10].

We have found a ~280 nucleotide (nt) repeat followed closely by a very good approximation of its complementary sequence in an intron of pro-apoptosis gene *PAR4*. It is Alu-type as shown in the repeat database [GIRI Replibase](#) [11, 12]. The hairpin

has 683 characters and its bases are as shown in [Figure 3](#).



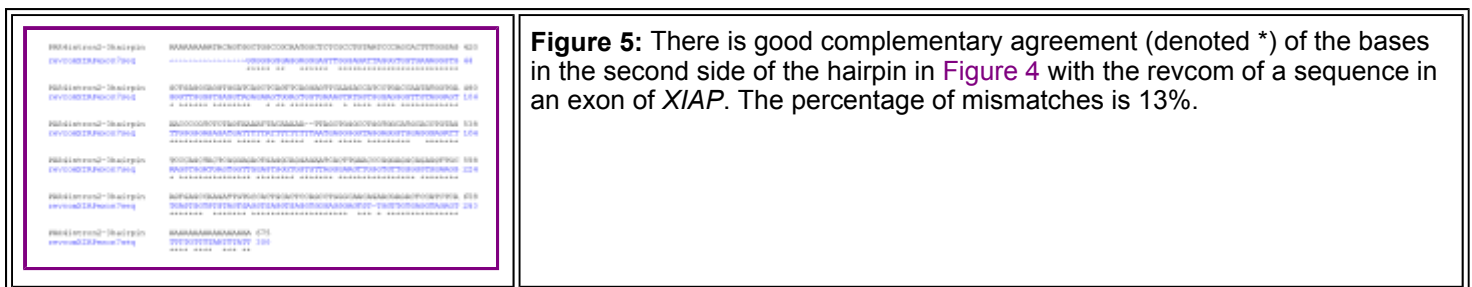
Folding the hairpin sequence in [Figure 3](#) with [mfold](#) [13] and using default settings [14] yields the image shown in [Figure 4](#).



Ordering the hairpin nt from 5' to 3', the second side of the stem is a good approximation of the revcom of a sequence in the seventh exon of the inhibitor of apoptosis (IAP) gene *XIAP*. The sequence is

```
>XIAPexon7seq
TTATTTAATTTTCTTTTGGAGATGGAGTCTTGCTTGTCAACCAGGCTGGAGTGCAGTGGAGTGATCTCTGCTCACTGCAACCTCCGCCT
TCTGGGTCAAGCGATTCTCGTGCCCTCAGCTTCCTGAGTAGCTGGAATTACAGGCAGGTGCCACCATGCCCGACTAATTTTTTTTTTATTT
TTTAGTAGAGACGGGGTTTACCATGTTGGCCAGGCTGGTATCAAACCTCCTGACCTCAAGAGATCCACTCGCCTTGCCCTCCCAAAGTG
CTGGGATTACAGGCTTGAGCCACCACGCCCCGGC
```

[Figure 5](#) shows the complementary alignments of the second side of the hairpin stem to the revcom of the exon (in mRNAs, T becomes U).



We have looked at flanking nt the above sequence in *XIAP*, but apparently it is not itself in such a large hairpin. Interestingly, all eight of the anti-apoptosis genes *BIRC1-BIRC8* have substantial sections that are within the Alu-like pattern and that align one or the other side of the above *PAR4* hairpin. We thus propose that transcription of the pro-apoptotic *PAR4* mRNA may produce regulatory elements that serve to rapidly inhibit translation of the IAP genes *BIRC1-BIRC8*. Of special interest would be alignments with sub-sequences of the 3' UTR (see note on MnSOD below). However, as repeats, the sides of the stems of the *PAR4* hairpin also appear approximately in thousands of positions in the genome. Hence biochemical tests are needed to determine whether or not any of the *PAR4-BIRC* (including *PAR4-XIAP*) relationships actually include the proposed counteraction.

There are doubtless many instances of large hairpins that could be folded from introns. In fact the same intron in *PAR4*, intron 2-3, contains two others known to us. They are shown in [Figure 6](#).

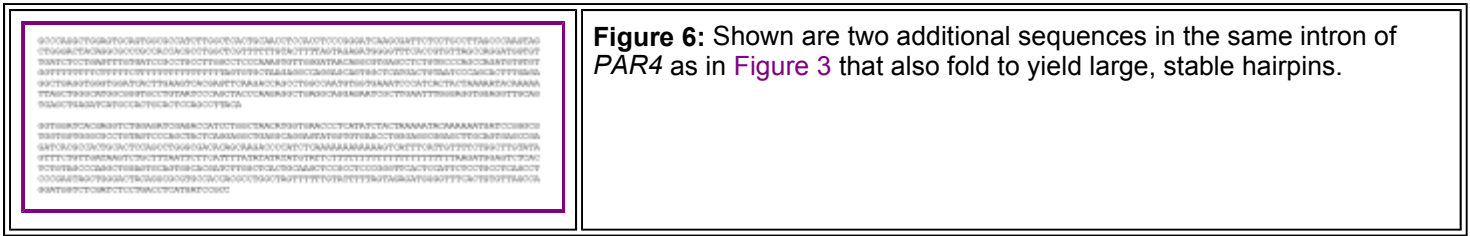


Figure 6: Shown are two additional sequences in the same intron of *PAR4* as in Figure 3 that also fold to yield large, stable hairpins.

Large hairpins are somewhat difficult to find because the number of windows to test in n bases is of the order n^2 and because how well the sides of the stem must complement each other is unclear. Interestingly, among the powerful [EMBOSS](#) tools, [einverted](#) with default settings seems not to find the *PAR4* intron 2-3 hairpin in Figure 3 and only the second of the two hairpins in Figure 6. Our algorithm, presently still under development, actually looks first for ~30 nt repeats and their nearby revcoms (up to a few hundreds of nucleotides away). Then it attempts to find a window that includes the pair embedded in a convincing, large fold. As folds are discovered, it considers the absolute value of the fold energy (kcal/mol) divided by the number of nt in the window to be an index of fold stability. The value of this index for the hairpin in Figure 4 is $463/683 = 0.68$, higher than the same ratio for miRNAs (typically ~0.35 to ~0.45).

At any rate, there was an element of serendipity in finding the hairpins in Figures 4 and 6. We were studying apoptosis and related phenomena that might be misregulated and become factors in the etiology of disease, as suggested in [15]. We hypothesized intron hairpins in regulation as a mechanism that would be both a natural control shortcut and a mechanism that could be maladjusted due to regulatory mutation or other factors affecting regulation (not affecting the gene expression products *per se*). Starting with the well-known pro-apoptosis gene *PAR4*, we were fortunate to find an instance by looking for short repeats and their nearby revcoms, then building a surrounding window for mfold testing.

Aside from whether such large intron secondary structures occur, another uncertainty is the mechanism by which small dsRNA molecules are selected for excision from the stem by an enzyme like Dicer. *In vitro*, Dicer will produce many species of small dsRNA molecules from a large one, but the output mixture is known by Ambion TechNotes [16] only as a "cocktail". Different small dsRNAs in the cocktail typically have different potencies as antisense inhibitors of targeted mRNA.

Biochemical experiments of various kinds are needed; we are planning such experiments using cells that can be forced by stress to undergo apoptosis. The cells must express as a default the anti-apoptosis gene and then the pro-apoptosis gene during the transition. If the above mechanism actually occurs in nature, then experiments must demonstrate inhibition of apoptosis by careful inhibition of selected parts of the above mechanism.

There appears to be a related instance of inhibition of gene expression in the case of human manganese superoxide dismutase (MnSOD) as described in the work of Stuart *et al.* [17]. MnSOD is important in detoxifying superoxide radicals generated in cellular respiration and in some types of cancer treatment. A 280 nt sub-sequence of its 3' UTR is Alu-like and apparently hybridizes with a specific cytoplasmic RNA of about 300 nt. The repression process might involve translational inhibition.

Conclusions

The genome is much smaller than the transcriptome, which is much smaller than the proteome. Thus it is generally recognized that some system for combinatorial control of transcription, editing, and translation is needed, but achieving a complete understanding of that system remains a major challenge to molecular geneticists.

The purpose of our paper is to propose a seemingly efficient mechanism by which initiating production of one protein might automatically inhibit production of proteins with counteracting functions. In addition to microRNAs, intronic sections of one

transcriptional unit might contain sequences such as Alus and their nearby revcoms that fold into dsRNA hairpins. Reactions involving the hairpins might lead eventually to inhibition of processing of transcriptional units for genes with opposing functions. While cogent and potentially an important theme in gene regulation, such a mechanism will require experimental verification. If preliminary experiments were to indicate the mechanism between counteracting genes occurs at least *in vitro*, then refined algorithms for searching for such patterns might be devised. Of high importance would be additional lab work to find such introns functioning as proposed *in vivo*.

Acknowledgement

Many thanks are due to our colleague Dr. Patrick Sullivan for helpful discussions on the topics in this paper. CJ and DOP received support by USP#SMH01905-01 North Carolina Foundation of Hope, Anonymous donor administered through the San Francisco Foundation, MJ by Elsa U. Pardee Foundation and NSF MCB 0446019.

References

1. [Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E. and Mellow, C. C. \(1998\). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391**, 806-811.](#)

2. [Tuschl, T., Zamore, P. D., Lehmann, R., Bartel, D. P. and Sharp, P. A. \(1999\). Targeted mRNA degradation by double-stranded RNA in vitro. *Genes Dev.* **13**, 3191-3197.](#)

3. [Eccleston, A. and Eggleston, A. K. \(2004\). RNA Interference. *Nature* **431**, 337.](#)

4. [Lee, R. C., Feinbaum, R. L. and Ambros, V. \(1993\). The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* **75**, 843-854.](#)

5. [Lewis, B. P., Burge, C. B. and Bartel, D. P. \(2005\). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**, 15-20.](#)

6. [Gibbons, R., Dugaiczyk, L. J., Girke, T., Duistermars, B., Zielinski, R. and Dugaiczyk, A. \(2004\). Distinguishing humans from great apes with AluYb8 repeats. *J. Mol. Biol.* **339**, 721-729.](#)

7. [Shankar, R., Grover, D., Brahmachari, S. K. and Mukerji, M. \(2004\). Evolution and distribution of RNA polymerase II regulatory sites from RNA polymerase III dependent mobile Alu elements. *BMC Evol. Biol.* **4**, 37.](#)

8. [Bonen, L. and Vogel, J. \(2001\). The ins and outs of group II introns. *Trends Genet.* **17**, 322-331.](#)

9. [Miklos, G. L. and Maleszka, R. \(2004\). Microarray reality checks in the context of a complex disease. *Nat. Biotechnol.* **22**, 615-621.](#)

10. [Holeik, M. \(2004\). Targeting endogenous inhibitors of apoptosis for treatment of cancer, stroke and multiple sclerosis. *Expert Opin. Ther. Targets* **8**, 241-253.](#)

11. [Jurka, J. \(1998\). Repeats in genomic DNA: mining and meaning. *Curr. Opin. Struct. Biol.* **8**, 333-337.](#)

12. [Jurka, J. \(2000\). Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet.* **9**, 418-420.](#)

13. [Zuker, M. \(2003\). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415.](#)

14. [Mathews, D. H., Sabina, J., Zuker, M. and Turner, D. H. \(1999\). Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.* **288**, 911-940.](#)

-
15. Perkins, D. O., Jeffries, C. and Sullivan, P. (2005). Expanding the 'central dogma': the regulatory role of nonprotein coding genes and implications for the genetic liability to schizophrenia. *Mol. Psychiatry* **10**, 69-78.

 16. Ambion Inc. (2005). Generate siRNA Populations with Dicer. *Ambion TechNotes* **11**:4.

 17. Stuart, J. J., Egry, L. A., Wong, G. H. and Kaspar, R. L. (2000). The 3' UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. *Biochem. Biophys. Res. Commun.* **274**, 641-648.