Pseudogene function: regulation of gene expression

John Woodmorappe

The discovery of a functional nitric oxide synthase (NOS) pseudogene compels us to understand pseudogenes in a new light. It confirms earlier clues suggesting that seemingly nonfunctional pseudogenes can regulate the expression of paralogous genes by producing antisense RNA. Moreover, only a partial sequence complementarity between sense and antisense segments of the gene and pseudogene is compatible with this function. This confutes the common evolutionary belief that major differences in sequence between paralogous genes and pseudogenes imply that the latter is necessarily a nonfunctional gene copy in a state of mutational drift. A second pseudogene may regulate the NOS gene by producing a truncated protein that can bind with the normal protein to produce an inactive heterodimer. Finally, the world of noncoding RNA, whether sense or antisense, offers further large-scale possibilities for undiscovered pseudogene function.

More and more noncoding DNA, long considered 'junk DNA', has eventually been found to be functional.[1–3] Hardly more than a few months pass by and there is not another scientific paper demonstrating function for some form of junk DNA. As summarized in this article, there is also growing evidence that at least some pseudogenes are functional. It should be stressed that pseudogenes, unlike other so-called junk DNA, have long been burdened not only with the ingrained belief that they lack function,[4] but also the additional onus of having supposedly lost a function. In addition, consider the following preconception relative to protein-coding genes in general:

> 'Considerably less analysis of this type has been performed on coding regions, possibly because the bias present from the protein-encoding function represented as nucleotide triplets (codons) promotes the general assumption that secondary functionality is present infrequently in protein coding sequences.'[5]

Since pseudogenes are supposedly nothing more than inactivated copies of protein-coding genes, and the sole function of protein-coding genes was thought to synthesize peptides, it seemed self-evident that (the apparent) loss of normal protein-coding function, in any gene copy, was synonymous with the loss of all function.

Potential modes of pseudogene function

It has been demonstrated that pseudogenic features, notably seemingly absent or disabled promoters, premature stop codons, splicing errors, frameshift-causing deletions and insertions, etc., do not necessarily abolish gene expression.[6] In fact, it is astonishing to realize that so-called pseudogenic features, instead of being 'gene killing' mutational defects, can serve as regulators of gene activity.[6] Finally, tests of gene as well as pseudogene expression commonly encounter difficulties in properly reproducing the conditions for activity. This is especially the case for genes and pseudogenes that express themselves only under very restricted conditions and/or in particular tissues.[6,7]

As will become obvious to the reader of this work, we need to expand our thinking about genes beyond their canonical protein-coding function. There is the growing realization that there is a whole world of noncoding functions possible for what usually are regarded as strictly protein-coding genes. Consequently, the pseudogene orthologs* (see Glossary) and paralogs* of protein-coding genes can no longer automatically be deemed nonfunctional just because the pseudogene is incapable of directing the synthesis of the customary peptide* (or any peptide). In fact, one set of functions involves the exclusive production of RNA, including antisense* RNA. As shown in Figure 1, and elaborated below, normal (sense) RNA can be modulated by its antisense counterpart as an important form of gene regulation.

Voices crying out in the wilderness

Against the backdrop of the customary negative opinion of pseudogenes, there have always been a few individuals who anticipated their functional potential. McCarrey *et al*.[8] were probably the first to suggest that pseudogenes can be functional in terms of the regulation of the expression of its paralogous genes. They noted that the sense RNA transcribed* by a gene could be effectively removed by hybridizing (forming a duplex) with the antisense RNA produced by the paralagous pseudogene. In addition, an otherwise nonfunctional peptide unit translated* by the pseudogene could inhibit the peptide translated by the gene. They likened these processes to a buffered acid-base titration. As described below, their ideas proved prophetic.

Inouye[9] apparently independently realized the same possibility for pseudogenes. He pointed out that a processed pseudogene located near a suitable promoter could produce antisense RNA, thus potentially regulating its parent gene. He also warned that antisense RNA genes might be com-

5'--…C C U G U C U C G C G A G G C U…--3' (Sense mRNA)

| | | | | | | | | | | | |

3'--…G G U C A G A G C G C A C C G U…--5' (Antisense RNA)

Figure 1. Illustration of one mechanism by which antisense RNA can occur, and the actual mechanism by which a sense-antisense complex forms between the nNOS (Lym-nNOS) gene, and a NOS pseudogene (now called the antiNOS-1 gene). Shortened and modified from Korneev et. al.[28] A segment of the DNA chain (not shown) is postulated to have been emplaced in a backwards (i.e. tail to head, or 3' to 5', direction). The resulting transcript (shown) is thereby antisense*. It forms a duplex by combining with the corresponding sense transcript (shown). Complementary* pairing of bases (A,C,G, and U) is indicated by dotted lines. Where the dotted lines are absent, complementary pairing does not occur across the juxtaposed nucleotides. Note that, overall, only a partial complementary pairing suffices for a stable sense-antisense RNA-RNA duplex to form.

paratively difficult to detect in those organisms that have larger genome sizes.

Since that time, evidence has steadily accumulated that shows that many pseudogenes are not inert (see Weil et al.[10] and citations). In addition, it has been discovered that antisense transcription, fairly common for viruses and prokaryotes, not only occurs in eukaryotes, but also does so more commonly than previously supposed. Ever so gradually, cracks began to appear in the seemingly impregnable 'pseudogenes are useless' fortress. Consider, first of all, the discovery of antisense RNA transcripts from a human DNA topoisomerase 1 pseudogene:

'While the function of these TOP1 antisense transcripts remains unknown, recent studies of naturally occurring antisense RNA have demonstrated several potential regulatory roles. The production of antisense transcripts from a TOP1 pseudogene was the first example of a naturally occurring antisense RNA transcript produced from a pseudogene…This study serves to emphasize not only the need to examine pseudogenes as potential active or regulatory sites, but also the need to remain aware of orientation specific regulation within the cell.'[11]

Pointedly, the TOP 1 pseudogene is transcribed even though it has many seemingly disabling 'lesions'. Moreover, the pseudogenic antisense RNA is an unlinked transcript. This means that it would therefore be more versatile, in terms of gene regulation, than the antisense RNA that comes from the same gene as the sense RNA. This owes to the fact that the regulator itself could be regulated.[11] Note that the TOP 1 pseudogene is not only active, but its activity is quite different from its paralogous genes, whose function it is to encode a peptide that helps uncoil the DNA molecule.

Comparable subsequent discoveries, relative to antisense RNA-producing pseudogenes, facilitated a slight 'thaw' in opinions concerning pseudogene function:

'Although the function of pseudogenes is generally considered in an evolutionary context where they provide for genetic diversity, it is an emerging view that some pseudogene transcripts may also serve a regulatory role by mechanisms such as antisense RNA control. Unfortunately, functional studies are presently lacking.'[12]

More recently, Weil et al.[10] discovered that the murine* FGFR-3 pseudogene is transcribed in fetal tissues in an antisense direction. This prompted the following consideration:

'As the regions of exact identity between FGFR-3 and its pseudogene can be up to 60 nt long, it may be envisioned that ΨFGFR-3 transcripts could play a regulatory role in FGFR-3 expression. If these antisense transcripts could hybridize to sense FGFR-3 transcripts inside the cells, this may lead to either rapid degradation or inhibition of translation.'[13]

Although Weil et al. thought of antisense regulation as being one that required exact complementarity* between sense and antisense RNA, it turns out that partial complementarity is sufficient (Figure 1), as elaborated below. Once again, there is a substantial difference in the behavior of this pseudogene and its peptide-encoding gene paralogs. These FGFR genes direct the synthesis of a group of structurally related growth factors that signal the cell.

A functional antisense RNA-producing pseudogene

In the snail *Lymnea stagnalis*, the neuronal enzyme NO synthase (nNOS) is encoded by the NOS gene (now called the *Lym*-nNOS gene). The enzyme induces the production of nitrogen oxide (NO), an intracellular signaling molecule in the snail's nervous system. One function of (NO) is the mediating of its feeding behavior.

Korneev, Park, and O'Shea[14] were probably the first to provide decisive evidence of sense RNA being regulated by the antisense RNA pseudogenic transcripts. They demonstrated that the neuronal nitric oxide synthesize (nNOS) gene is actively regulated by the antisense transcripts from a NOS pseudogene. The old hypothesis of McCarrey et al.[8] repeated in some form by subsequent investigators, had apparently found confirmation. Members of the Yale University group that is actively studying pseudogenes recognize the foregoing as an example of a possible functional pseudogene.[15] However, as a matter of semantics, a 'functional pseudogene' invariably becomes renamed a gene. In this instance, the erstwhile NOS pseudogene is now called

the antiNOS-1 gene.[16]

In their investigation, Korneev et al.[17] demonstrated that the pseudogenic antisense RNA and the nNOS (Lym-nNOS) mRNA can be independently expressed in a neuron-specific manner. They also verified that stable RNA-RNA duplexes do form in vivo. Next they studied the modulating effects of the antisense RNA on actual protein synthesis, first in vitro and then in vivo. As for the latter:

> 'To summarize, an identified neuron that contains the nNOS mRNA but not the pseudo-NOS RNA consistently expresses a functional NOS protein. In contrast, a neuron in which both transcripts are colocalized, NOS enzyme activity is practically undetectable. These in vivo observations support the view that the antisense pseudo-NOS transcript suppresses the translation of functional nNOS mRNA in neurons in which the two transcripts are colocalized.'[18]

The actual pseudogene function that regulates NOS gene expression is believed to operate as follows:

> 'Specifically, the active transcription of the pseudogene will lead to the suppression of nNOS protein synthesis, and on the other hand, the inhibition of pseudogene transcription will permit nNOS production. Importantly, a switch from the "off" to the "on" mode of nNOS expression would be achieved rapidly because the functional nNOS gene is already active in both modes and nNOS mRNA could be available immediately for translation once the suppressive effect of the NOS pseudogene is removed. We therefore propose that in the CGC [cerebral giant cell] antisense-mediated translational control, supplemented by transcriptional regulation of the NOS pseudogene, provides an effective molecular mechanism for achieving rapid changes in nNOS protein production in response to some internal or external signals.'[19]

Protein suppression by a second functional antisense pseudogene?

A second functional (pseudo)gene, occurring in tandem with the first one (see Figure 2) is now known to exist:

> 'Note that in antiNOS-1 the antisense region is located at the 5' end of the molecule, whereas in antiNOS-2 it is located at the 3' end. Another important difference is that although antiNOS-1 cannot be translated into a protein because all three reading frames contain multiple stop codons, the antiNOS-2 transcript contains an open reading frame encoding a truncated nNOS-homologous protein of 397 amino acids.'[20]

This peptide, however, lacks certain functional domains.[21] Ordinarily, this would be taken as an obvious indicator of the fact that the protein is fatally defective and thus devoid of function, as is the pseudogene that directs

its synthesis. Counter intuitively, however, this protein may have its own function of regulating gene expression, and to do so at a level that differs from that of the other pseudogene:

> 'One intriguing possibility, therefore, is that the antiNOS-2 protein functions as a natural dominant negative regulator of nNOS activity through binding to the normal nNOS monomer, forming a nonfunctional heterodimer.'[22]

Instead of removing mRNA 'out of circulation' by hybridizing with antisense RNA, as antiNOS-1 does, the regulatory function of antiNOS-2 apparently consists of its shortened 'nonfunctional' protein forming a complex with the protein synthesized by the NOS (Lym-nNOS) gene, thus removing it 'out of circulation'. Without this modulation, this gene would freely encode a homodimer that contains two major functional domains.

Functional mildly-conserved pseudogene nucleotide sequences

A long held ostensible support for the absence of pseudogene function has been their usual apparent lack of sequence conservation. Protein-coding genes typically vary only slightly among orthologs and paralogs as a result of purifying selection*. This is a result of the fact that most proteins cannot tolerate more than a few alterations without a marked detriment to their functional performance. The usually high nucleotide sequence variance of pseudogene copies, relative to each other and to their protein-coding gene orthologs and paralogs, is conventionally ascribed to random mutational drift, a 'sure' hallmark of nonfunction. This attitude is a consequence of the previously discussed ingrained belief that the function of protein-coding genes begins and ends with the encoding of a (usually highly conserved) protein.

In rebuttal to such reasoning, Zuckerkandl[2] points out that an apparent lack of sequence conservation in junk DNA, of whatever type, may only imply a function that does not require a conserved sequence (or, in the present case, at least a highly conserved sequence), not absence of all function. The recent cited studies unequivocally bear this fact out for pseudogenes. Note that, in terms of both overall similarity and sequence, the antiNOS-2 sequence is translated into a peptide that differs considerably from that of its Lym-nNOS paralogous gene. Yet it is probably functional. The antiNOS-1 paralog of antiNOS1 cannot even be translated into any peptide at all. According to standard evolutionary thinking, a pseudogene that is paralogous to a protein-coding gene but incapable of translation into a peptide unquestionably lacks function. The functional antiNOS-1 (pseudo)gene soundly refutes this long-held dogma.

In terms of their respective nucleotide sequences, the functional genes and pseudogenes under consideration are not strongly conserved relative to each other. Large sections of the sense DNA strands of antiNOS-1 and antiNOS-2

show only a 76%–80% correspondence to the paralogous *Lym*-nNOS gene[20]. Much the same holds for the biologically active antisense segments themselves. In fact, Korneev *et al.*[14] show that the sense-antisense RNA need only have about 80% complementary, and that over a sequence of only ~150 nucleotides, for stable duplex formation, and ensuing gene regulation, to occur (Figure 1).

Clearly, pseudogene sequences can have appreciable variability relative to their protein-coding gene paralogs, and yet remain functional. Furthermore, relative to the antisense strands themselves, we now realize that only partial blockage of target RNA by its antisense complement is sufficient to disrupt the function of the former. Sense-antisense RNA can interact at several different levels. This further increases the potential variability of pseudogene sequences that are compatible with function.

Multiple modes of pseudogenic antisense RNA production

In conventional protein synthesis, the process begins with the transcription of the DNA molecule in a 5' to 3' (left to right) direction, and the complementary DNA strand is not used. When antisense-mediated pseudogenic regulation of genes, by pseudogenes, was first proposed, the antisense RNA was envisioned as originating from a 5' to 3' transcription of the anticoding strand (not shown) of DNA.[23] However, the TOP 1 pseudogene produces its antisense transcript through a backwards (3' to 5' direction) reading of the coding strand (for illustration, see Fig. 1 of Zhou et al.[24]). In the antisense-transcribed FGFR-3 pseudogene, backwards transcription of the sense strand also takes place, and is probably caused by a heterologous promoter.[13]

The NOS pseudogenes (now called the antiNos-1 gene and anti-Nos-2 gene) display yet another manner of antisense RNA production.[15] The transcription proceeds in the conventional 5' to 3' direction, and most of it is in the expected sense direction. However, a segment of the DNA, now shared by the two (pseudo)genes, is transcribed in an antisense direction. As shown in Figure 1, this owes to the fact that the DNA segment itself is backward in orientation within the overall DNA sequence. This forces it to be transcribed in the 3' to 5' direction despite the 5' to 3' direction of transcription. It is believed that one copy of the ancestral gene underwent a small internal inversion, thus half-somersaulting a DNA segment into the antisense (tail to head) orientation

(<<<<<<<<<< in Figure 2). New stop and start signals (not shown) generated a new intergeneric region, effectively splitting the original inversion-containing gene copy into the two new (pseudo)genes shown in Figure 2. The inverted DNA sequence itself was apportioned by this process into the new intergenic region as well as the two (pseudo)genes. In fact, the end segments of the original antisense DNA segment that find themselves situated within the boundaries of the two new (pseudo)genes (Figure 2) are the ones which perform the regulatory antisense functions. There was also a large-scale restructuring of the intron-exon structure, in antiNOS-1 and antiNOS-2, relative to the original gene (not shown).

The created origin of the functional pseudogenes

Let us evaluate the above-discussed evolutionary scenario (Figure 2) for the origin of these functional pseudogenes. To begin with, the relatively low degree of sequence similarity between the paralogous gene and pseudogenes weakens the argument that they necessarily arose from a common ancestral gene. Most re-arrangements of DNA segments within functional genes are harmful or fatal to gene function. It does not appear likely that an accidental occurrence (essentially a hopeful monster inversion mutation) would generate new functionality of such magnitude, even if the hypothesized millions of years of natural selection were available to remove all of the unsuccessful DNA rearrangements within genes every generation. Even the evolutionists
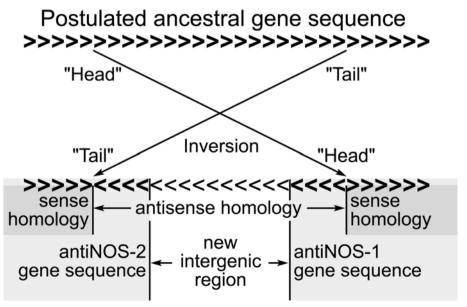

Figure 2. *The inferred evolutionary origin of the antisense sequence containing (pseudo)genes, antiNOS-1 and antiNOS-2. Explained in text. The DNA sequences are denoted by (>>>>>) and (<<<<<), in sense orientation and antisense orientation, respectively. Bold indicates coding DNA, no bold denotes intergenic DNA. Note that the antisense-oriented segment is shared by the face to face end parts of the antiNOS-1 and antiNOS-2 genes as well as the intergenic region between them.*

who propose this scenario hint at its improbability:

> 'Here, we show that the duplication of an ancestor of the *Lymnaea* nNOS gene was followed by the occurrence of an internal DNA inversion in one of the copies. *Remarkably*, this produced new regulatory elements required for the termination and the activation of transcription. Consequently, the gene was split, and simultaneously two new genes with *entirely new functions* were created (emphasis added).'[25]

Remarkable indeed. The special creation of these functional pseudogenes, in nearly their present state, seems at very least to be a much more parsimonious explanation for their origins than the accidental evolutionary scenario.

Conclusion

Exciting new evidence is now surfacing on the functionality of pseudogenes. Korneev *et al.* assess the significance of their discovery as follows:

> 'With respect to the evolution of regulatory functions of pseudogenes, we must now conclude that transcribed pseudogenes are not necessarily without function. Indeed, they would appear to be especially suited to roles involving the antisense regulation of the active genes to which they are related.'[26]

No longer can it be assumed (*contra* Max[4]) that pseudogenes are just useless evolutionary discards. In fact, Korneev *et al.*'s[21] discovery prompts them to suggest that theirs is but the first discovery of an entirely new class of regulatory gene. Of course, research on sense-antisense regulatory roles is only in its infancy. The fact that there are many different modes of antisense-RNA action bodes well for large-scale pseudogene function.

Most important of all, the challenge to the 'pseudogenes are dead genes' thinking goes far beyond antisense RNA. We now know that RNA-only genes are not only more common than previously supposed, but that their numbers may potentially dwarf those of protein-coding genes. In addition, there is actually a whole barely understood new world of noncoding RNA functions,[27] most of which are related to the regulation of gene expression, and which can perform their regulatory roles in either the sense or the antisense direction. Some have even suggested that noncoding RNA is the 'dark matter' of genomes. It is easy to see how RNA genes could be embedded within pseudogenes. Compared with protein-coding genes, RNA genes are usually much smaller, have a wider range of potential promoters, have only a relatively weak nucleotide sequence composition bias*, and are much more diverse.[16] This field of research is wide open.

Glossary

Antisense RNA—An RNA molecule that, for whatever reason, is transcribed in a backwards (that is, tail to head, or 3' to 5') direction.

Complementarity—The pairing off of nucleotides, between a strand of RNA and a strand of DNA, in the following manner: Adenines (A) with uracils (U), and the cytosines (C) with the guanines (G).

Composition bias—the difference between the sequence of coding DNA from noncoding DNA.

Murine—referring to the rodent family Muridae, which includes rats and mice.

Ortholog, **Orthologous**—a member of the same family of genes and/or pseudogenes occurring within different organisms, and usually believed by evolutionists to have arisen from a common ancestral gene or genes.

Paralog, **Paralogous**—a member of the same family of genes and/or pseudogenes occurring within an organism, and usually believed by evolutionists to have arisen from a common ancestral gene or genes

Peptide—Synonymous with protein, a chain of amino acids linked together.

Purifying Selection—The preferential die-off of organisms that contain a harmful mutation that, in the context of this paper, prevents a protein from functioning at an optimum level. Both creationists and evolutionists recognize the existence of purifying selection.

Transcription—The intracellular construction of a messenger RNA (mRNA) molecule that is complementary to the DNA molecule of the gene. The DNA molecule serves as a template for transcription.

Translation—The conversion of the sequence of a messenger RNA molecule into a series of amino acids that are subsequently joined together to form a specific peptide.

References

1. Riley, D.E. and Krieger, J.N., Diverse eukaryotic transcripts suggest short tandem repeats have cellular functions, *Biochemical and Biophysical Research Communications* **298**:581–586, 2002.

2. Zuckerkandl, E., Why so many noncoding nucleotides? The eukaryote genome as an epigenetic machine, *Genetica* **115**:105–129, 2002.

3. Standish, T.G., Rushing to judgment: functionality in noncoding or 'junk' DNA, *Origins (Loma Linda)* **53**:7–30, 2002.

4. Max, E.E., Plagiarized errors and molecular genetics, <www.talkorigins.org/faqs/molgen/>, 19 March 2002.

5. Shah, A.A. *et al.*, Computational identification of putative programmed translational frameshifting sites, *Bioinformatics* **18**(8)1046–1053, 2002.

6. Woodmorappe, J., Unconventional gene regulation and its relationship to classical pseudogenes, *Proceedings of the Fifth International Conference on Creationism*, 2003, (in review).

7. Woodmorappe, J., Are pseudogenes 'shared mistakes' between primate genomes? *CEN Tech. J.* **14**:55–71, 2000.

8. McCarrey *et al.*, Determinator-inhibitor pairs as a mechanism for threshold setting in development: a possible function for pseudogenes, *Proc. Nat. Acad. Sci. USA* **83**:679–683, 1986.

9. Inouye, M., Antisense RNA: its functions and applications in gene regulation—a review, *Gene* **72**:28–29, 1988.

10. Weil *et al.*, Antisense transcription of a murine *FGFR-3* pseudogene during fetal development, *Gene* **187**:115–122, 1997.

11. Zhou, B.-S., Beidler, D.R. and Cheng, Y.-C., Identification of antisense RNA transcripts from a human DNA topoisomerase I pseudogene, *Cancer Res*earch **52**:4280, 4284, 1992.

12. Andrea, J.E. and Walsh, M.P., Identification of a brain-specific protein kinase Cζ pseudogene (ψPKCζ) transcript, *Biochemical J.* **310**:843, 1995.

13. Weil *et al.*, Ref. 10, p. 121.

14. Korneev, S.A., Park, J.-H. and O'Shea, M., Neuronal expression of neural nitric oxide synthase (nNOS) protein is suppressed by an antisense RNA transcribed from a NOS pseudogene, *J. Neuroscience* **19**:7711–7720, 1999.

15. Harrison, P.M., Echols, N. and Gerstein, M.B., Digging for dead genes, *Nucleic Acids Research* **29**:818, 2001.

16. Korneev, S. and O'Shea, M., Evolution of nitric oxide synthase regulatory genes by DNA inversion, *Molecular Biology and Evolution* **19**:1228–1233, 2002.

17. Korneev *et al.*, Ref. 14, p. 7715.

18. Korneev *et al.*, Ref. 14, pp. 7717–7718.

19. Korneev *et al.*, Ref. 14, p. 7718.

20. Korneev and O'Shea, Ref. 16, p. 1229.

21. Korneev and O'Shea, Ref. 16, p. 1230.

22. Korneev and O'Shea, Ref. 16, p. 1232.

23. McCarrey and Riggs, Ref. 8, pp. 681–682.

24. Zhou *et al.*, Ref. 11, p. 4281.

25. Korneev and O'Shea, Ref. 16, p. 1228.

26. Korneev *et al.*, Ref. 14, p. 7719.

27. Eddy, S.R., Computational genomics of noncoding RNA genes, *Cell* **109**:137–140, 2002.

28. Korneev *et al.*, Ref. 14, p. 7713.

John Woodmorappe has an MA in geology and a BA in biology, from a midwestern US state university. He is a science educator by profession. He is the author of *Studies in Flood Geology*, *Noah's Ark: A Feasibility Study*, and *The Mythology of Modern Dating Methods*.

Implausible mutations

'Upon close examination it appears difficult at best to make a plausible case for mutations as an important evolutionary factor. Gene duplications may well occur, but contrary to the idea that such an event would free the organism to "invent" new functions with one gene while the other is keeping house, both copies would suffer mutations at the same rate. The presence of duplicate genes does not protect either one from truly random events and while the organism is perhaps safer with two genes of one kind, certainly it is no more free to experiment because there is conceptually no way to restrict mutational activity to one of the copies.'

Schwabe, C.
Mini review: theoretical limitations of molecular phylogenetics and the evolution of relaxins,
Comp. Biochem. Physiol. **107B**(2):170, 1994.

Erratum

In the article 'Speed of light slowing down after all? Famous physicist makes headlines', (*TJ* **16**(3):7-10, 2002) it was stated that physicist Keith Wanser was '…a member of the RATE group sponsored by ICR/CRS…". However, although Keith Wanser has been an invited guest to a RATE meeting and has helped review a few RATE reports, he is not a member of the group.