Volume 320, number 3, 281–283 FEBS 12294 April 1993

Commentary

A commentary on: 'Unexpected intron location in non-vertebrate globin genes', by: Moens et al. (FEBS Letters, 312 (1992) 105–109)

Bill Pohajdak and Brian Dixon

Department of Biology, Dalhouse University, Halifax, Nova Scotia, Canada

Received 13 January 1993

We are submitting this commentary in order to prevent the confusion which Moens et al. may provoke in the minds of readers. Our comments are intended to point out the exclusion of some major issues and some errors which may mislead readers of this paper.

GENERAL COMMENTS

Many ideas presented in this article are not new. The theory of globin gene evolution presented in this paper is essentially that proposed by Blake [1] in 1981, and the figure presenting this mechanism is an update of the scheme presented by Lewin [2] in 1984. The authors cite neither of these papers.

The rationale for comparing *C. elegans*, a monomeric globin gene, and *Artemia*, a multimeric globin gene, is unclear. The authors present data for several other globin genes, including better choices to compare each gene with. If the authors' intention was to comment on globin gene evolution, why not consider all the available data, not two isolated cases?

Most importantly, the authors of this paper completely ignored the issue of the phase of the introns which they discuss. Conservation of the phase of introns is essential in order to maintain the correct reading frame. Homologous introns, those derived from a single intron in an ancestral gene, should have the same phase. Introns in similar positions but of different phase cannot be homologous. The authors claim that the central intron of the *C. elegans* globin gene is homologous to the plant central intron, but that it has moved to a more anterior position. This sliding process would have to maintain reading frame in order to produce a complete protein product, yet the central intron of plant genes is phase 0 (it cuts the gene between codons) while the *C. elegans* central intron is phase 2 (it cuts within a single codon such that the first two nucleotides are separated from the third nucleotide). How could an intron move along a gene and change phase without destroying reading frame? The only reference offered by the authors on this point does not mention phase or present a possible mechanism for intron sliding. The independent insertion of these introns into similar positions in each of these genes is a more probable explanation for the disparity in both position and phase.

Additional evidence for the independent insertion of these introns is presented in Fig. 1 of Moens et al. The *C. elegans* globin gene shares a more recent common ancestry with the *P. decipiens* globin gene than with plants. Both organisms are nematodes and the amino acid sequences of these genes are similar (35% identity), yet the the position and phase of the central intron differs between these two genes. Three of four amino acid residues in the region between the two introns are identical. It is extremely difficult to imagine a process of intron sliding, which must involve paired nucleotide loss and gain at each end of the intron to conserve reading frame, replacing a sequence of nucleotides lost at one end of the intron with one similar enough to conserve sequence identity to such a degree.

This problem is even more difficult to reconcile when one considers that the other two introns (positions B12 and G7 in Fig. 1) are identical in phase and position for seven vertebrate globin genes [3–8], five invertebrate globin genes [9–13] and five plant globin genes [14–17]. If these introns have been conserved so highly, why would the central intron be so plastic? It is interesting to note that the only exception to the above comparison is the intron at position G7 of the *Artemia* T4 globin gene, which is conserved in position but not in phase. This makes the authors' choice of the *Artemia* T4 globin gene even more questionable. The other *Artemia* globin genes contain no introns. The altered phase of the *Artemia* globin gene intron argues that it was inserted into the T4 gene following the duplication of an ancestral monomeric gene which had lost all of its introns.

Correspondence address: B. Pohajdak, Dept. of Biology Dalhousie University Halifax, Nova Scotina Canada.

SPECIFIC PROBLEMS

The concept of phase alters the conclusions made by the authors at several points in their paper. We would also like to point out some errors in the presentation of this paper which are misleading.

Our immediate problem with Fig. 1 is that there are residues missing from the *P. decipiens* sequence, notably the serine residue which should appear right before the intron separating the two repeats. The authors suggest that this intron is equivalent to the intron which separates the sequence encoding the hydrophobic leader sequence from the rest of the gene. We have never noticed such an equivalence when aligning the two repeats using the complete sequence. In addition the sequence within the brackets starting PD2 is missing a final histidine residue. This sequence is in fact the last 22 amino acids of the first repeat! Is it a linking sequence or a functional part of the first repeat? The authors cannot have it both ways. Since this region corresponds to the entire H helix of all other globins, we suggest that it is in fact a functional part of the first repeat.

The valine residue shown at the end of the Art 4 sequence belongs to the beginning of of the Art 5 sequence, as it is separated from Art 4 by an intron. These errors may seem small, but missing amino acids can make a large difference in aligning protein sequences, and the authors state themselves that the 'proper alignment of the translated globin sequences is a prerequisite for the assessment of equivalent intron positions'.

As previously mentioned, the *Artemia* intron which aligns with the other introns at position G6 is a different phase from the others and therefore cannot be homologous.

In Fig. 2 the sequences underlined in the *C. elegans* sequence are supposed to be regulatory sequences, yet the upstream ones lie inside a potential 514 bp intron which separates the hydrophobic leader encoding an exon from the rest of the gene. The CAG splice consensus sequence is visible 12 bases upstream from the start of the exon box. Nematode hemoglobins are extracellular and require a hydrophobic signal sequence to exit the cell. If the exon upstream is not used as a signal sequence, it may be trans-spliced from another message, a phenomena previously reported for *C. elegans* [18]. The regulatory sequences underlined are not adequately explained. Is the ATTAT sequence underlined supposed to be a TATA box?

If this is the TATA box, there are several problems to be considered. It is not the correct sequence and is located too close to the beginning of the mRNA, and indeed the start codon. Where is the 5′ non-coding sequence of the mRNA? This region is usually 50 bases long. The protein sequence depicted in the exon box cannot start with a serine residue. The methionine following this serine does not have a good start consensus sequence. Also, methionine codons which occur within 10 bases of the beginning of the messenger are very poorly utilized as start codons [19]. (All of this is basic knowledge taught in second year biochemistry courses). Lastly, the protein is probably translated prior to the position indicated in this figure. There is an in frame codon just prior to the boxed exon which would encode a lysine residue (AAA) that is homologous to a lysine residue near the start of the mature protein of both *Ascaris* and *P. decipiens* hemoglobins.

Is the AACAAA sequence supposed to be a polyadenylation signal? The A just prior to the second boxed exon should be inside the box, in order to give the previous R codon a third nucleotide. All three nucleotides of a codon must be within the exon, as the definition of a intron is a non-coding DNA segment separating coding segments!

The authors suggest that the precoding introns of the *Artemia* globin gene are homologous yet, according to Fig. 2, the first one is phase 1 and the second is phase 0. In fact, the AG which starts the sequence within the box containing the second exon is in fact part of the preceding intron, and the AG which starts that intron is part of the previous exon. This correction reveals that these introns have the correct splice sequences as the authors claim (except that the 3′ end of the second precoding intron does not appear to have a CAG acceptor consensus sequence). Once corrected, both introns are indeed phase 0, indicating the possibility of a common origin (assuming we can trust the alignment given in Fig. 1). It is difficult to say for certain, however, since the authors do not clearly state whether or not these are the only two precoding introns in this gene. In any case what the authors refer to as the introns of *Ascaris* and *P. decipiens* are phase 1, while these, and the precoding introns of *B. reveena* and *C. elegans* are all phase 0. There-fore all precoding introns cannot have had a common evolutionary origin.

In section 3.1. the authors state that in general the central intron 'separates the haem-binding region into two structural units, F2 and F3 that make contact with the haem from opposite sides'. This passage is confusing as readers may think that the authors are referring to the F2 and F3 amino acid positions from Fig. 1, when in fact they are using Go's [20] terminology for the two exons divided by the central intron. In addition, if one examines the *C. elegans* sequence, the central intron groups almost all the haem binding residues within one exon, while the other exon has no apparent function. This is yet more evidence for the insertion of this intron as opposed to it being derived from an intron found in an ancestral gene.

The second paragraph states that intron insertion can occur up to six codons apart, but the authors argue that introns are always lost and never gained. This is confusing.

The second paragraph of section 3.2. is devoted to correcting an alignment of the AB region which places an intron illogically in the A helix. The only reference

given for this alignment is one of our papers which includes no alignment of that region or discussion of the alignment in that region. Where did this erroneous alignment come from? In addition, the authors present a five residue revision of this alignment, but the intron in question is located within the codon for the residue at position B12. How does a five base error place this intron within the A helix?

The authors offer Table I as a correction of the above problem, but in their revised alignment the two introns they dub as the precoding introns of *P. decipiens* are no longer aligned as they state in Fig. 1 and the text.

The *P. decipiens* globin gene in Fig. 3 are incorrectly drawn as the exon encoding the hydrophobic leader sequence is omitted. The separation of the molluscs into two groups, one of which evolved in conjunction with the vertebrates and one which evolved along a separate path, is certainly at odds with the generally accepted theories of evolution. Again no mention is made of the differences in phase of the central introns and therefore no attempt is made to reconcile these differences.

Perhaps the greatest criticism of this evolutionary scheme is that it is in fact not a scheme at all. The radiation of each group independently from a common ancestral gene provides no explanation about how differences within each taxa arose and is not very likely.

REFERENCES

[1] Blake, C.C.F. (1981) Nature 291, 616.
[2] Lewin, R. (1984) Science 226, 328.
[3] Leder, A., Miller, H.J., Hamer, D.H., Seidman, J.G., Norman, B., Sullivan, M. and Leder, P. (1978) Genetics 75, 6187–6191.
[4] Nishioka, Y. and Leder, P. (1979) Proc. Nat. Acad. Sci. USA 77, 2806–2809.
[5] Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Cell 21, 627-638.
[6] Spritz, R.A., Jagadeeswaran, P., Choudary, P.V., Biro, P.A., Elder, J T., DeRiel, J.K., Manley, J.L., Gefter, M.L., Forget, B.G. and Weissman, S.M. (1981) Proc. Nat. Acad. Sci. USA 78, 2455–2459.
[7] Jeffreys, A.J., Barrie, P.A., Harris, S., Fawcett, D.H. and Nugent, Z.H. (1982) J. Mol. Biol. 156, 487–503.
[8] Blanchetot, A., Wilson, V., Wood, D. and Jeffreys, A.J. (1983) Nature 301, 732–734.
[9] Jhiang, S.M. and Riggs, A.F. (1989) J. Biol. Chem. 264, 19003-19008.
[10] Titchen, D.A., Glenn, W.K., Nassif, N., Thompson, A.R. and Thompson, E.O.P. (1991) *Biochim. Biophys. Acta* 1089, 61–67.
[11] Naito, Y., Riggs, C.K., Vandergon, T.L. and Riggs, A.F. (1991) Proc. Nat. Acad. Sci. USA 88, 6672–6676.
[12] Dixon, B., Walker, B., Kimmins, W. and Pohajdak, B (1992) J. Mol. Evol. 35, 131–136.
[13] Sherman, D.R., Kloek, A.P., Krishnan, B.R., Guinn, B. and Goldberg, D.E. (1992) Proc. Nat. Acad. Sci. USA 89, 11696–11700.
[14] Ostergaard, E., Paludan, K., Hyldig-Nielsen, J.J., Jorgensen, P. and Marcker, K.A. (1981) Nature 291, 677–679.
[15] Landsmann, J., Dennis, E.S., Higgins, T.J.V., Appleby, C.A., Krott, A.A. and Peacock, W.J. (1986) Nature 324, 166–168.
[16] Bogusz, D., Appleby, C.A., Landsmann, J., Dennis, E.S., Trinick, M.J., Peacock, W.J. (1988) Nature 331, 178–180.
[17] Welters, P., Metz, B.A., Schell, J. and de Bruijn, J. (1989) Nucl. Acids Res. 17, 1253.
[18] Krause, M. and Hirsh, D. (1987) Cell 49, 753–761.
[19] Kozak, M. (1987) Nucl. Acids Res. 15, 8125–8147.
[20] Go, M. (1981) Nature 291, 90–92.