Volume 320, number 3, 284–287 FEBS 12295 April 1993

Commentary Response

L. Moens[a], J. Vanfleteren[b], I. De Baere[a], A.M. Jellie[c], W. Tate[c] and C.N.A. Trotman[c]

[a]*Department of Biochemistry, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium,* [b]*Laboratory of Animal Morphology and Systematics, University of Ghent, K.L. Ledeganckstraat 35, B-9000 Ghent, Belgium* and [c]*Department of Biochemistry, University of Otago, Dunedin, New-Zealand*

Received 25 February 1993

Our paper in FEBS Letters [1] provoked a reaction by Pohajdak and Dixon, mainly claiming that the evolution of the intron/exon pattern in globin genes can only be explained in terms of intron loss and independent intron gain. A crucial point in their commentary is the operational definition of intron homology. By definition the adjective 'homologous' refers to common evolutionary origin [2,3]. Thus homologous introns are assumed to have a common origin. Introns located both in conserved position and phase, when the proteins are properly aligned, are thought to be homologous. The introns located in the B and G helices of vertebrates and higher plant globin genes meet these criteria and it is generally accepted that they were present in the very same positions in their common ancestor. However, the definition of homologous introns would not exclude those introns that have moved from an ancestral position to a secondarily acquired one. Pohajdak and Dixon argue that such a displacement is too unlikely, and they state that "The splice junctions must conserve phase and sequence on each side, before they can be said to have a common origin" [4]. The immediate consequence is that they have to classify introns that do not meet these criteria as non-homologous introns resulting from independent insertion events. This leads to bizarre evolutionary schemes involving frequent intron loss and intron gain as depicted in Fig. 2 of the cited work [4].

Since our paper and the one by Dixon and Pohajdak [4] were published, three more globin gene structures have been reported. Much like *Pseudoterranova decipiens*, the *Ascaris suum* globin gene consists of two repeats, or domains, presumably as a result of a gene duplication event, and an intron separates the exon containing the leader sequence from the remainder of the gene. Three more introns are localized in sequence regions of the first repeat encoding the B (B12/phase 2), E (E8/phase 1) and G (G6/G7; phase 0) helices, respectively. However, the second repeat of the *Ascaris* gene has also three introns in identical positions and phases, whereas the second repeat of the *Pseudoterranova* gene has retained only the intron localized in the region encoding the B helix [5]. A clear example of intron loss in closely related species. The *Paramecium caudatum* haemoglobin gene is a single gene, which is interrupted by only one intron in the E helix [6]. An alternative alignment places the intron at position (F2/F3; phase 0). Either way, its position does not coincide with the plant central intron. Is this one more example of independent intron insertion?

The globin gene of the green alga *Chlamydomonas eugametos* has three introns [7]. The first intron divides the B helix (B5/B6; phase 0), the second intron is located near the end of the E helix (E19/E20; phase 0), the third one near the end of the F helix (F9/F10; phase 0) (alignment assuming the D helix and part of the CD interhelix region to be absent). None aligns with the position of any other globin intron reported thus far. Yet, there are three introns as in plants; are they all three resulting from independent insertion events? Instead, we believe that these introns are, in fact homologous.

We have interpreted intron location at non-conservative positions or phase as resulting from 'sliding'. The term sliding was first used by Craik et al. [8,9] to describe alteration of intron position in the genes for serine protease and dihydrofolate reductase. These alterations coincided with internal length differences in the polypeptide. We have extended the meaning of 'sliding' to include every displacement of an intron along a sequence irrespective of the precise mechanism involved. Obviously, the mechanism of the displacement will be different; to avoid confusion, we will therefore use the neutral term 'displacement' in our further discussion.

The fact that the mechanism of intron displacement is unknown does not preclude it happening. There are many gene structures, the evolution of which can be reasonably understood by assuming intron displacement events. For example, *Caenorhabditis elegans* pos-

Correspondence address L. Moens, Department of Biochemistry, University of Antwerp, Universiteitsplein 1, B-2610 Antwerp, Belgium. Fax: (32) (3) 820 2248

Published by Elsevier Science Publishers B.V.

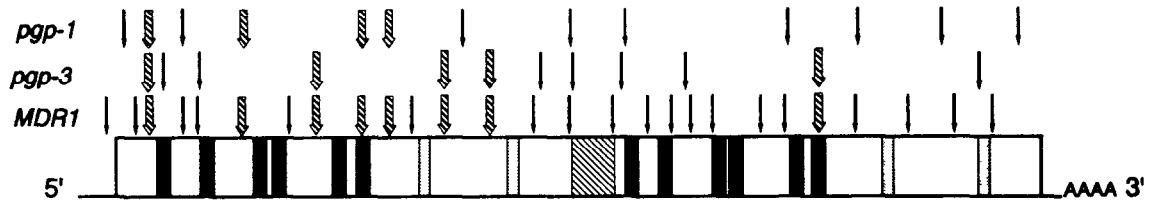

Fig. 1. Evolutionary conservation of intron positions in *C. elegans pgp* and human *MDR* genes. The rectangle represents the coding region of a P-glycoprotein cDNA. Transmembrane domains are symbolized by filled bars, the nucleotide binding consensus motifs by stippled bars. The hatched box represents the so-called linker region. Arrows indicate intron positions relative to the cDNAs. Thick, hatched arrows denote introns that are conserved between at least two sequences with respect to intron phase and coding context. (Figure taken from [12] with permission of J. Mol Biol.).

sesses four GAPDHase genes (*gpd-1* through *gpd-4*). *Gpd-1* and *gpd-4* are nearly identical (99% identity) with respect to coding sequence information and to the position and phase of two introns. The genes *gpd-2* and *gpd-3* are also nearly identical to each other and they are very similar to the *gpd-1*, *gpd-4* couple, but their two introns are inserted at different positions with respect to the former [10]. Is it realistic to preclude these introns from being homologous? Internal similarities in the rabbit muscle phospho-fructokinase gene clearly suggest that this gene evolved from an ancestral gene by duplication and divergence. The amino-terminal half of this gene contains 12 introns, the second half has 9 introns. None of these matches any intron in the first half of the gene [11]. Must we accept that all nine result from independent insertion events? The P-glycoprotein gene family of *C. elegans* consists of four homologues termed *pgp-1* (mapped to chromosome IV), *pgp-2* (chromosome I), *pgp-3* and *pgp-4* (chromosome X). The detailed structure of *pgp-1* and *pgp-3* was recently studied by Lincke et al. [12]. Both genes are analogous to mammalian P-glycoprotein genes in structure and deduced protein sequence. However, the mammalian gene has 26 introns in its coding sequence, *pgp-1* has 13 introns of which four are inserted at a conserved position and phase, and *pgp-3* has 12 introns of which 5 are homologous to their mammalian counterpart by the criteria of Dixon and Pohajdak [4], yet *pgp-1* and *pgp-3* share only one intron by these criteria (Fig. 1). We do believe that these examples point to the combined action of intron loss and intron displacement rather than de novo insertion.

Pohajdak and Dixon raise the unknown mechanism of displacement of introns as an insurmountable obstacle, but by which mechanism would non-homologous introns insert repeatedly? The problem of intron displacement has been addressed by Martinez et al. [13] at length. Their model comprises four consecutive steps: excision of the intron, re-insertion into a nearby site of the pre-mRNA via reversible transesterification, reverse transcription of the modified pre-mRNA, and homologous recombination via gene conversion. Crucial steps such as RNA-mediated recombination and gene conversion were recently shown to occur in the yeast *Saccharomyces cerevisiae* [14]. This model avoids problems of phase alteration, and implies that genuine homologous introns are re-inserted (or eventually removed: intron diminution) in the original gene or a copy of that gene. This model also predicts that intron loss would tend to be more readily achieved than re-insertion, and this is precisely what we attempted to visualize in our Fig. 3 [1]: loss of introns from a common ancestral globin gene has occurred independently in the course of evolution. The legend clearly states: "note that the diagram depicted is not a phylogenetic representation". On the bottom line is the 4 exon, 3 intron ancestral globin gene, and on the top line are the intronless genes; the general evolutionary tendency is to remove introns irrespective of the divergence path. Eventually it should be possible to reconstruct the real genealogy of life and to fit the evolution of the globin gene to each branch of the tree precisely. But this is certainly premature at present; phylogeny is still poorly understood [15] and data on globin gene structure are too fragmentary.

We would like to reply to some specific objections raised by Pohajdak and Dixon. Obviously we wanted to confirm Lewin's hypothesis, quoted as Reference 3 in [1], and extend it to the globin gene structures available today. The putative *C. elegans* globin and the *Artemia* T4 repeat were not at all presented as examples but because they were never published as such.

The drafting errors in Fig. 2 are regretted, and the corrected figure is enclosed (Fig. 2). We have probed a cDNA library from *C. elegans* and confirmed expression of the globin and removal of the central intron [16]. Some potential signal sequence was found to precede the first methionine but since the gene from which it originates has not yet been identified it is premature to conclude that it is a signal sequence. We are surprised to hear from Pohajdak and Dixon that "nematode globins are extracellular", no such rule of nature having been substantiated, and even more so that "the other *Artemia* globin genes contain no introns" since this is not known.

In our Fig. 1, the H helix, clearly bracketed and headed as such, was sited at both ends of the alignment as a service to readers. In commenting on Table I, Pohajdak and Dixon have confused the intron under dis-

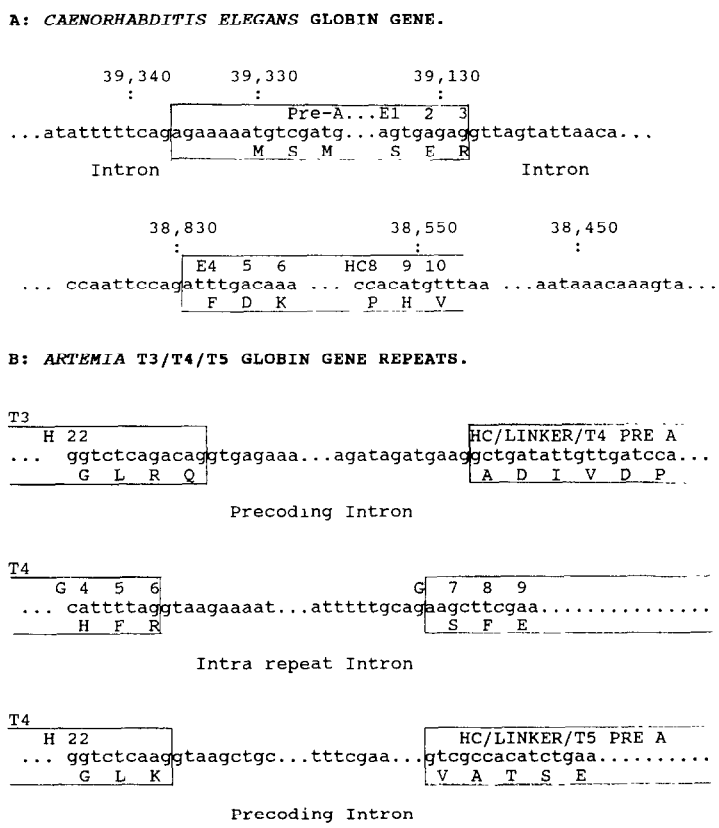

Fig. 2. Gene organization of the *C elegans* globin and of repeat T4 of *Artemia*. (A) Partial nucleotide sequence of the *C. elegans* globin gene showing splice junctions Exons are boxed. Numbers refer to the sequence of the template strand insert in cosmid zk 637 (from Ref. 6 in [1]). Deduced coding strand sequence is shown for convenience. Helix notation as in Fig 1. Amino acids are given by the single letter code underneath the second base of each codon. (B) Partial nucleotide sequence of *Artemia* T3/T4/T5 globin repeats showing splice junctions. Exons are boxed. Helix notation as in Fig 1.

cussion (not B12, but the intron preceding it). We agree with Dixon et al. [17] that the alignment of nematode globin sequences with other species is convincing between residues B5 and F9. Proximal alignment is difficult but should nevertheless be addressed. Options are limited if B5-F9 are accepted and further limited if structure within the globin family is as conserved as is suggested by presently known globin structures. Although no alignment is unequivocal in that region, the alternatives offered in Fig. 1 and Table I score highest against Bashford's template and are worthy of discussion. The first implies a 5-residue AB region and an intron near A3, both of which are unprecedented. The second requires deletion of B1–B4 and of AB, and places an intron between domains, all three of which points are precedented.

We stress that the evolution of nuclear, intron-containing genes, here exemplified by the globin genes, by a combination of loss, displacement and strict conservation, is fully compatible with the assumption that introns were involved in the initial assembly of primordial protein-coding genes (introns early) [18–20]. Introns are, indeed, frequently inserted at the borders of functional protein modules. Exceptions to this rule can be easily explained as resulting from secondary displacement. As the ancestral globins were likely intracellular and monomeric, precoding and inter-repeat introns are more recently acquired. In at least one case there is clear evidence that the insertion of an inter-repeat intron has accompanied globin gene duplication event [21]. The origin of these introns is thus different from the three intra-repeat introns which are relics of the ancestral globin gene assembly.

We cannot definitely disprove the hypothesis of generalized independent intron insertion that was advanced to explain the origin and the further evolution of introns (introns late) [22,23]. According to this hypothesis all introns would be derived from self-inserting elements. The regularities in the distribution of the introns are more difficult to reconcile with the hypothesis, however, and the molecular evolution of nuclear genes is more readily explained by a model which combines exon shuffling to create the primordial genes with a molecular drive of selective conservation, displacement and loss of introns to create the variability observed in extant globin genes.

REFERENCES

[1] Moens, L., Vanfleteren, J., De Baere, I., Jellie, A.M., Tate, W. and Trotman, C. (1992) FEBS Letters 312, 105–109.
[2] Lewin, R. (1987) Science 237, 1570.
[3] Reeck, G.R., de Haën, C., Teller, D., Doolittle, R.F., Fitch, W.M., Dickerson, R.E., Chambon, P., McLachlan, A.D., Margoliash, E., Jukes, T.H. and Zuckerkandl, E. (1987) Cell 50, 667.
[4] Dixon, B. and Pohajdak, B. (1992) Trends Biochem. Sci. 17, 486–488.
[5 Sherman, D.R., Kloek, A.P., Krishnan, B.R., Guin, B. and Goldberg, D.E. (1992) Proc. Natl. Acad. Sci. USA 89, 11696–11700.
[6] Yamauchi, K., Ochiai, T. and Usuki, I. (1992) Biochem. Biophys. Acta. 1171, 81–87.
[7] Couture, M., Guertin, M. and Savard, F. (1993) Unpublished; EMBL entry CELI637.
[8] Craik, C.S., Sprang, S., Fletterick, R. and Rutter, W.J. (1982) Nature 299, 180–182.
[9] Craik, C.S., Rutter, W.J. and Fletterick, R. (1983) Science 220, 1125–1129.
[10] Huang, X.Y., Barrios, L.A.M., Vonkhornporn, P., Honda, S., Albertson, D.G. and Hecht, R.M. (1989) J. Mol. Biol. 206, 411–424.
[11] Lee, C.P., Kao, M.C., French, B.A., Putney, S.D. and Chang, S.H. (1987) J. Biol. Chem. 262, 4195–4199.
[12] Lincke, C.R., The, I., Van Groeningen, M. and Borst, P. (1992) J. Mol. Biol. 228, 701–711.
[13] Martinez, P., Martin, W. and Cerff, R. (1989) J. Mol. Biol. 208, 551–565.
[14] Derr, L.K. and Strathern, J.N. (1993) Nature 361, 170–173.
[15] Brusca, R.C. and Brusca, G.J. (1990) Invertebrates, Sinauer, Underland, Massachusetts.
[16] Mansell, J.B., Timms, K., Tate, W., Moens, L. and Trotman, C. (1993) Unpublished data.
[17] Dixon, B. and Pohajdak, B. (1991) Proc. Natl. Acad. Sci. USA 88, 5655–5659.
[18] Doolittle, F. (1978) Nature 272, 581–582.
[19] Darnell, J.E. (1978) Science 202, 1257–1260.
[20] Doolittle, F.W. and Stoltzfus, A. (1993) Nature 361, 403.
[21] Naito, Y., Riggs, C., Vandergom, T.L. and Riggs, A. (1991) Proc Natl. Acad. Sci. USA 88, 6672–6676.
[22] Cavalier-Smith, T. (1991) Trends Genet. 7, 145–148.
[23] Rogers, J.H (1989) Trends Genet. 5, 213–215.