J Mol Evol (1996) 42:641–647

JOURNAL OF MOLECULAR EVOLUTION



# Evolutionary History of Introns in a Multidomain Globin Gene

**Anna M. Jellie, Warren P. Tate, Clive N.A. Trotman**

Department of Biochemistry, University of Otago, Box 56, Dunedin, New Zealand

Received: 27 September 1995 / Accepted: 15 December 1995

**Abstract.** The *Artemia* hemoglobin contains two subunits that are similar or different chains of nine globin domains. The domains are ancestrally related and are presumed to be derived from copies of an original single-domain parent gene. Since the gene copies have remained in the same environment for several hundred million years they provide an excellent model for the investigation of intron stability.

The cDNA for one of the two types of nine-domain subunit (domains T1–T9) has been sequenced. Comparison with the corresponding genomic DNA reveals a total of 17 intradomain introns. Fourteen of the introns are in locations on the protein that are conventional in globins of other species. In eight of the nine domains an intron corresponds to the B helix, amino acid B12, following the second nucleotide (phase 2), and in six domains a G-helix intron is located between G6 and G7 (phase 0). The consistency of this pattern is supportive of the introns having been inherited from a single-domain parent gene. The remaining three introns are in unconventional locations. Two occur in the F helix, either in amino acid F3 (phase 1) in domain T3, or between F2 and F3 (phase 0) in domain T6. The two F introns strengthen an interpretation of intron inheritance since globin F introns are rare, and in domains T3 and T6 they replace rather than supplement the conventional G introns, as though displacement from G to F occurred before that part of the gene became duplicated. It is inferred that one of the F introns subsequently moved by one nucleotide. Similarly, the third unconventional intron location is the G intron in domain T4 which is in G6, phase 2, one nucleotide earlier than the other G introns. Domain T4 is also unusual in lacking a B intron. The pattern of introns in the *Artemia* globin gene supports a concept of general positional stability but the exceptions, where introns have moved out of reading frame, or have moved by several codons, or have been deleted, suggest that intron displacements can occur after inheritance from an ancient source.

**Key words:** Protein evolution — *Artemia* — Intron — Exon — Hemoglobin

## Introduction

In the debate about the possible "early" or "late" origin of introns (Roger and Doolittle 1993; Dibb 1993; Hurst 1994) the hemoglobin gene is a well-studied model adopted widely in support of both sides of the argument. Hemoglobin introns, where present, are usually found in one or more of only three positions in terms of the translated protein: corresponding to amino acid residue B12 (B helix, residue 12); in a somewhat variable E helix location; and between G6 and G7.

This consistency lends some support to the introns-early idea of each intron having been inherited continuously since a single primordial origin. Not all hemoglobins have all three introns, and any missing ones, according to the introns-early theory, have been deleted at some ancestral stage. The origin of those remaining could have been coincident with an event in which primordial minigenes, coding for compact modules of protein now represented by exons, became linked together to encode more complex proteins while untranslated DNA flanking the minigenes survived as introns. Since it is

*Correspondence to:* C.N.A. Trotman

unlikely that hemoglobin appeared suddenly in a one-step process, one would envisage the original exon linkage event as creating a precursor several stages more primitive than hemoglobin as we know it, possibly containing initially duplicate modules. Alternatively, the functional protein could have existed first as a set of separately expressed modules in noncovalent association; then later a mechanism developed for excluding the intervening sequence from the mRNA. Whatever the mechanism of origin of hemoglobin, by tracing the history of introns we may gain insight into events considerably predating its first appearance.

According to an alternative or introns-late interpretation, introns were inserted into an essentially complete gene, plausibly by reverse transcription and recombination. The argument that introns divide the protein into smaller modules may be of little significance and is to some extent tautological. According to this model, insertion of an intron could occur at any time and may still be happening. These two different models have common elements in that introns inserted by recombination (i.e., later than formation of the gene) can thereafter be inherited and could be almost as old as if they were formed by exon shuffling as in the introns-early model. The distinction between the models hinges not on antiquity but on whether an intron was a consequence of the formation of the gene or was inserted later: hence the terms "formative" and "insertional" might be more apt than "early" and "late."

There are difficulties with either theory. If introns have been inserted repeatedly into intact genes over a long period of time (introns-late), why are they not randomly distributed? Why do they recur across widely different taxa in the equivalent structural position, such as the B, E, and G helices of globins? Conversely, it is difficult to reconcile the introns-early explanation with even small discrepancies of intron position between taxa. If the corresponding intron in a variety of genera is derived from a common ancestor, how can the intron have moved into different reading frames in different genera without disrupting the expressed protein? An intron consistently present in some genera may be missing from others: through what mechanism was it so precisely excised?

*Artemia* hemoglobin genes are a powerful model for study of the acquisition and fate of introns. All nine domains in each gene are believed to have originated from the same ancestral single-domain gene through a series of copying events within the same genome, eliminating one of the uncertainties of comparing sequences from genera that might have evolved at different rates. We have obtained the cDNA sequence for one of the two genes, confirming that it codes for a continuous polymer of nine globin domains, T1–T9 (Manning et al. 1990). This was previously thought to be the α polymer of the homodimer HbI, but partial sequence from the other gene suggests it may in fact be the β polymer of the homodimer HbIII. The designation T1–T9 will remain definitive of the present sequence. Comparison with genomic DNA shows that two of the three recognized intron sites within globin domains are represented in *Artemia* by a total of 17 introns. However, positional discrepancies reinforce an interpretation that introns were inherited with gene multiplication and incidentally suggest that the nine-domain structure arose through the fusion of intermediate trimers.

The combined evidence that the *Artemia* introns have been inherited but that certain of them have moved out of reading frame, have moved by several codons, or have been deleted is interpreted as showing that such displacements are not at variance with a possible primordial origin of introns.

## Materials and Methods

*PCR and Template Preparation.* The template for the polymerase chain reaction (PCR) with the exception of domain T4 was genomic DNA. Genomic DNA was isolated from 10 mg of *Artemia* nauplii grown for 44 h from the time of rehydration as previously described (Manning et al. 1990). Domain T4 was amplified from a genomic clone that had been isolated from a λgt10 library. The PCR was done using primers designed to flank individual domains as a whole or to flank the front or back half of domains. Sense and antisense oligonucleotides were designed using the previously published cDNA sequence (Manning et al. 1990). Thirty cycles of amplification were conducted (30 s at 94°C, 1 min at 55°C, and 5 min at 72°C to allow for a large product) using 0.3 μg of genomic DNA, *Taq* DNA polymerase, and standard buffer supplied by Gibco (BRL).

*DNA Sequencing.* PCR products were isolated by purification from agarose gels using Prep-A-Gene (Bio-Rad) and sequenced directly in an Applied Biosystems 373A automatic DNA sequencer using dye terminators.

## Results and Discussion

### *Identification of Introns*

Comparison of the genomic DNA sequence with the previously published cDNA (Manning et al. 1990) revealed a total of 17 intradomain introns. There were two in each domain except for domain T4, which contained only one intron. Table 1 summarizes the intron length and the inferred helix (A-H) of each intron. Table 2 shows the first and last nine nucleotides and the flanking DNA sequence with translation around each intron, and thereby the residue number and codon reading phase.

An interpretation of introns in terms of the conflicting theories of their origin and stability depends on their protein positions being correctly inferred. Although the crystal structure of *Artemia* hemoglobin is not known, the assignment of the translated sequence to the conven-

**Table 1.** Summary of the intradomain introns found in an *Artemia* polymeric globin; intron lengths (Kilobases) were estimated on agarose gel

| | Intron length, kb | | | |
|---|---|---|---|---|
| | Helix location | | | |
| Domain | B | E | F | G |
| T1 | 1.5 | — | — | 4.3 |
| T2 | 2.0 | — | — | 2.3 |
| T3 | 1.3 | — | 1.1 | — |
| T4 | — | — | — | 1.3 |
| T5 | 1.5 | — | — | 2.0 |
| T6 | 2.7 | — | 0.6 | — |
| T7 | 2.7 | — | — | 1.4 |
| T8 | 2.0 | — | — | 2.3 |
| T9 | 1.0 | — | — | 3.8 |

tional globin helices and turns is considerably strengthened by the unequivocal alignment of all nine domains with each other. The translated sequences also clearly align with other globins of confirmed structure (Trotman et al. 1991) and the intron pattern found in the present study further reinforces that alignment.

### *Conventionally Located Introns*

All introns except two were positioned in the conventional globin first (B helix) or last (G helix) intron locations (Table 1). There were no middle (E helix) introns. All domains except T4 were found to have a B intron, which is located following the second nucleotide (phase 2) of the codon for amino acid residue B12 (Table 2). This is exactly the same position and phase as in all species in which a B intron is present and in which an overall sequence alignment is reasonably certain, including invertebrates (Sherman et al. 1992; Jhiang and Riggs 1989; Titchen et al. 1991; Dixon et al. 1992), plants (Hyldig-Nielsen et al. 1982; Welters et al. 1989; Landsmann et al. 1986; Bogusz et al. 1988; Jensen et al. 1981), and vertebrates (Blanchetot et al. 1983) but excluding *Chlamydomonas* (Couture et al. 1994), where the alignment indicates substantial deletions.

The usual position of the G helix intron in globins is G6/G7 (phase 0) (references as for B introns). In six *Artemia* domains this intron was located conventionally at G6/G7 but unusually in domain T4 it was one nucleotide earlier at G6, phase 2 (Table 2). In domains T3 and T6 this G intron was missing but introns were anomalously present at F3 and F2/F3, respectively, instead.

The positions of the *Artemia* B and G introns can be argued in support of either a great antiquity for introns if their consistency is emphasized, or conversely in support of independent, more recent insertions if it is argued that inheritance could never be accompanied by a change of reading phase. Thus, the B introns are particularly indicative of an origin predating the growth of the *Artemia* globin polymer from one to nine domains, since all B introns are coincident and in phase with other species as if inherited from a common ancestor. Antiquity alone, however, does not distinguish between an "early" origin of introns in untranslated sequence flanking minigenes (Gilbert 1987, 1978; Darnell and Doolittle 1986; Blake 1978; Gilbert and Glynias 1993) and a "late" insertion into an existing gene at an early evolutionary date.

If intron survival reflects solely a pattern of inheritance it must be inferred that where introns are missing—for example, the B introns of *Artemia* domain T4, of *Chironomus* (Antoine and Niessing 1984; Kao et al. 1994), and of a mouse-globin-like gene (Nishioka et al. 1980)—they have been deleted subsequently. The phase discrepancy of the G intron in domain T4 would be explained as intron slippage. Notions of introns slipping, moving, or being deleted, however, are viewed more skeptically in the insertional or introns-late theory, as any mutation that caused a change of reading phase could be fatally disruptive. An introns-late interpretation (Cavalier-Smith 1991; Stoltzfus et al. 1994) of the *Artemia* data would allow the B intron to have been inserted into the equivalent position on more than one and possibly eight separate occasions; similarly, it would allow the G intron to have been acquired up to seven times and the F intron twice. Numerous hybrid mechanisms are of course possible, such as the B intron being acquired insertionally and replicated through the series of domain duplications, or different introns being acquired at different times.

### *Unconventionally Located Introns*

The F introns may provide the key to choosing between the formative and insertional models in *Artemia*. Globin F introns are rare, the only other reported possibilities being in the protists *Paramecium* (Yamauchi et al. 1992) and *Chlamydomonas* (Couture et al. 1994) in which significant deletions of exon sequence render the intron assignments debatable. Despite the rarity of the F intron in globin genes it appears in two *Artemia* domains and its position is supported by the convincing internal alignment of nine domains underpinned by 15 other introns. Superficially this position appears to be at variance with a primordial origin, since there is no evidence for an original F intron, and at variance with the minigene model, since it conflicts with the exon junctions proposed by Go (1981). The F introns are midway between the conventional globin second (E helix) and third (G helix) intron positions. However, partly because E introns have not been found in *Artemia,* and every domain contains either an F or a G intron but not both or neither, we consider the F introns to be displaced G rather than displaced E introns.

**Table 2.** *Artemia* hemoglobin intradomain introns showing the first and last nine nucleotides of each intron (lowercase)[a]

*Artemia* globin B helix introns

| Domain | Exon | | | Intron | Exon | | |
|---|---|---|---|---|---|---|---|
| | B 1 0 | B 1 1 | B 1 2 | | B 1 2 | B 1 3 | B 1 4 |
| T1 | T T C | C T T | A G | g t a a c a t t t ... t c c t c t t a g | T | G T C | T T T |
| | **F** | **L** | | | **S** | **V** | **F** |
| T2 | T T C | C A A | A G | g t a c t t t t c ... t t t t t t c a g | A | A T G | A T C |
| | **F** | **Q** | | | **R** | **M** | **I** |
| T3 | T T T | G G A | A A | g t a a g t a a c ... t t t t t t t a g | A | C T C | T T C |
| | **F** | **G** | | | **K** | **L** | **F** |
| T4 | T T C | A T G | A G | [no intron] | G | A T G | T T C |
| | **F** | **M** | | | **R** | **M** | **F** |
| T5 | T T C | G C C | A A | g t a a g a g t t ... t t t t c t t a g | A | T T G | T T G |
| | **F** | **A** | | | **K** | **L** | **F** |
| T6 | T T C | A C A | A G | g t a t t a t t t ... t g c t t t a a g | A | C T T | T T C |
| | **F** | **T** | | | **R** | **L** | **F** |
| T7 | T T T | A A A | A G | g t a a g a a a g .. a t t t t t c a g | C | C T T | T T C |
| | **F** | **K** | | | **S** | **L** | **F** |
| T8 | T T C | A C A | A A | g t a a g t t t t ... t t a t t a c a g | G | G T A | T T A |
| | **F** | **T** | | | **K** | **V** | **F** |
| T9 | T T C | A A A | C A | g t a a g t g c a .. t t c t t t t a g | G | C T G | T T T |
| | **F** | **K** | | | **Q** | **L** | **F** |

*Artemia* globin G and F helix introns

| Domain | Exon | | | Intron | Exon | | |
|---|---|---|---|---|---|---|---|
| | G 4 | G 5 | G 6 | | G 6 | G 7 | G 8 |
| T1 | C A C | T T T | G A G | g t a a g a a t c ... c t t t t t t a g | | G C G | T T T |
| | **H** | **F** | **E** | | | **A** | **F** |
| T2 | C A T | T T C | C A G | g t a a t a g a t ... t t c c a t t a g | | A A T | T T T |
| | **H** | **F** | **Q** | | | **N** | **F** |
| T4 | C A T | T T T | A G | g t a a g a a a a ... t t t t t g c a g | A | A G C | T T C |
| | **H** | **F** | | | **R** | **S** | **F** |
| T5 | C A G | T T C | G A T | g t a a g t t g t ... a c t t t a c a g | | C A A | T T T |
| | **Q** | **F** | **D** | | | **Q** | **F** |
| T7 | A T G | T T C | A A G | g t a a a t t t t ... t a t t t t t a g | | A G C | T T C |
| | **M** | **F** | **K** | | | **S** | **F** |
| T8 | C A T | T T C | C A G | g t a g g a a t g ... t a a t t c t a g | | G C C | T T C |
| | **H** | **F** | **Q** | | | **A** | **F** |
| T9 | C A C | T T C | A A G | g t a t g t t t t ... t c t t a a t a g | | G A T | T T T |
| | **H** | **F** | **K** | | | **D** | **F** |
| | F 1 | F 2 | F 3 | | F 3 | F 4 | F 5 |
| T3 | A T C | A A A | G | g t a a a c a g t ... t g t t t a a a g | A A | T T G | G G T |
| | **I** | **K** | | | **E** | **L** | **G** |
| T6 | C T A | A A G | | g t a t c t a t a ... g t t g t c t a g | G A C | C T T | G G T |
| | **L** | **K** | | | **D** | **L** | **G** |

[a] Flanking exon sequences are shown in uppercase under their inferred amino acid positions in helix notation; the translations are given below in bold type

The rarity of F introns suggests that a displacement to this position is more likely to have happened once before a duplication event that gave rise to domains T3 and T6 than twice after this duplication. The single-nucleotide phase discrepancy between the F introns of domains T3 and T6 would then have occurred after domain duplication. By whatever mechanism, one of the F introns (more probably in T3 since T6 is in phase with the six G introns) has been moved out of phase, like the G intron in T4.

The mechanism for a one-nucleotide displacement need not be complex, nor fatally disruptive, and of course an apparent movement of an intron could reflect a historical deletion or insertion in an exon upstream. A previous deletion of the first nucleotide of the codon for amino acid G6 in domain T4, or insertion of a nucleotide in the codon for F2 in T3, would have maintained the exon consensus -AG- splice sites while introducing phase discrepancies now observed. The pivotal requirement for the phase change to become established was survival of the gene until a compensating event occurred downstream (or the whole process in reverse). A gene product having repetitive functional domains may have a survival advantage if one domain is inactivated. Signifi-

cantly, the alignment of the *Artemia* domains (and globins generally) provides abundant evidence that they have survived numerous variations in exon length including, for example, in the critical C, CD, and D heme environment and in the GH turn (Manning et al. 1990; Trotman et al. 1991).

The fact that the positions of the 17 introns are not randomized supports the conclusion that single-nucleotide slippage is rare and that substantial displacements are even rarer. The more substantial movement of an intron from G6/G7 to F2/F3 is not interpreted therefore as a repetition of single-nucleotide steps. Single mutational events could nevertheless cause major shifts. For example, deletion of a section of exon would appear to advance the start of the next intron; creation of a new splice site could convert part of an exon into intron or vice versa. The critical requirement again is that the gene survives until acceptable compensations occur. Machinery exists that hypothetically could excise an intron and reinsert it elsewhere (Derr and Strathern 1993; Martinez et al. 1989; Sharp 1985), albeit with little obvious selectional motivation, but if a recombinant mechanism has operated in the history of *Artemia* hemoglobin, the deletion of the B intron from domain T4 would be the most conspicuous possibility.

There is no reason to doubt that the nine-domain gene is derived from a single domain precursor, which is as likely as a modern gene to have contained introns. If *Artemia* replicated this gene complete with introns then we see evidence of introns later moving out of reading phase (last intron, domains T4 and T3), moving a distance of 16 codons (from G to F helix, domains T3 and T6) and being deleted (first intron, domain T4). While each of these three phenomena may superficially seem contradictory to a model of intron inheritance, the evidence that they have occurred in *Artemia* outweighs theoretical objections. If mechanisms of intron inheritance and movement have operated since the divergence of the *Artemia* line, they are likely to have operated in other organisms over a similar or longer time scale.

### *Implications for the Origin of Introns*

By showing within one species that intron displacements, while exceptional, can occur and survive, our findings suggest that no limit can be placed on the potential age of these introns. They could be as ancient as hemoglobin or one of its molecular antecedents. If introns can move, then discrepancies of intron position between genera do not negate models for the origin of larger protein genes from smaller exons. On the other hand, however convincingly the stability of the globin B intron appears to support an inheritance dating back toward a primordial origin, possibly even predating hemoglobin as the molecule is now recognized, a recombinant origin for that intron cannot be precluded.

Intron positions, and by implication splice-site consensus sequences, are evidently much more highly conserved than intron length, which can differ considerably between homologous examples. The inference is that most of the sequence of an intron does not have much biological significance, whereas the loss of a crucial splice site is potentially catastrophic to the expressed protein, resulting in the translation of an intron or the nontranslation of an exon. The pattern we observe of predominantly stable position but unstable content is more supportive of inheritance, coupled with relatively infrequent shifts and even rarer complete deletions, than of numerous independent insertions.

### *Origin of the* Artemia *Polymeric Globin*

Phylogeny analysis by the program Protpars (Felsenstein 1988) suggests that the ninefold multiplication of the single-globin gene in the ancestors of *Artemia* occurred at least 700 MY to 500 MY ago (Trotman et al. 1994). The estimate is approximate and minimal because a correction for multiple mutations at single sites by the Poisson distribution, which probably gives an underestimate, indicates a saturation of amino acid changes between domains (between 95% and 143%). The sequence history has been further obscured by deletions and insertions in coding regions and by the intronic events being discussed. An evolutionary tree derived from domain differences is therefore of limited resolution and cannot be accepted uncritically as specifying the precise pathway of duplications. Additional data inputs, however, including certain threefold features of the alignment and now the intron locations, suggest that triplets of domains were a possible evolutionary intermediate (Fig. 1).

One supporting piece of information is that domains T7, T8, and T9 diverge from the other six domains at the base of the evolutionary tree (Trotman et al. 1994). Some features of the alignment between domain sequences (Manning et al. 1990; Trotman et al. 1991) may be vestiges of a trimeric stage. For example, the putative trimer of domains T7-T8-T9 is characterized by an unusual hydrophobic A14 Leu, an FG4 Val, and a G19 Asp. Similarly, domains T1-T2-T3 include -Arg-Gly-Thr- for the EF turn and an uncharacteristically hydrophobic residue, Val or Tyr, at G10. Possible signs of a duplication of trimers include an unusual hydrophobic Val at CD2 in domains T1 and T4, the hydrophobic Leu at F6 in domains T3 and T6, and the shortness (two or three residues) of the GH turn in domains T1-T2 repeated in T4-T5.

The intron pattern now adds a new data set to this information and provides further support for a duplication of triplet domains. The model (Fig. 1) proposes that the parent gene of *Artemia* globin contained the first (B

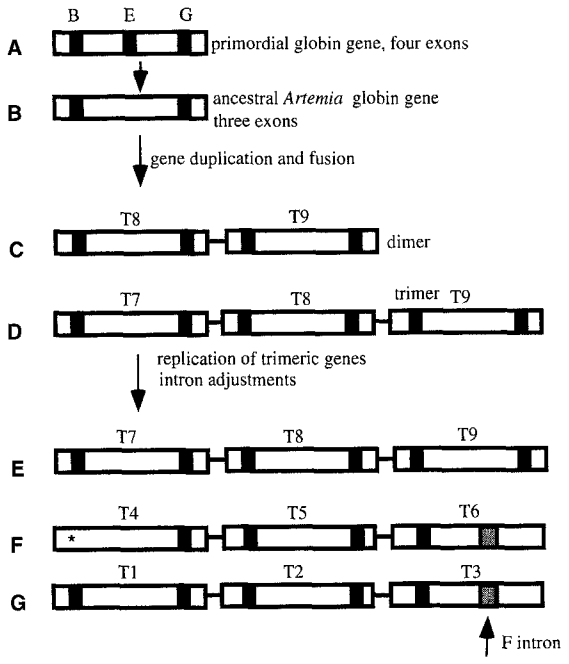


**Fig. 1.** Main stages in a possible origin of the gene for an *Artemia* hemoglobin polymer of nine domains. **A** The primordial globin gene contained four exons linked by three introns in positions corresponding to the B, E, and G helices of the protein. **B** On the branch of the arthropod line leading to *Artemia,* the E intron was deleted. **C** A succession of gene fusion events led to a dimer and then to a trimer **(D)** which duplicated **(E)**. **F** A duplicate trimer underwent transposition of one G intron to an F helix position. **G** The trimer containing the new F intron was duplicated. Subsequently but in unknown order, single-nucleotide adjustments of intron position occurred to the G intron in domain T4 and the F intron in domain T3 (see Table 2), the B intron was deleted (*) from domain T4, and the three trimers became fused into the nine-domain composite gene (not shown).

helix) and third (G helix) introns, but not the second (E helix) intron represented at least sporadically on the insect branch of the arthropod line (Kao et al. 1994). The single domain gene, by duplication and fusion or unequal crossing over, multiplied to become a dimer and then a trimer, represented by domains T7-T8-T9. This trimer gene duplicated to generate domains T4-T5-T6. Domain T6 later acquired the unique marker of an F helix intron and eventually the T4-T5-T6 gene duplicated to produce T1-T2-T3, the F intron being copied from T6 to T3. Subsequently domain T4 lost its first intron; also, the G6/G7 intron of T4 and the F2/F3 intron of T3 were repositioned by one nucleotide in opposite directions. This minimum model allows the fusion of the trimer genes, yielding the present nine-domain molecule, to have occurred either before or after the last intron adjustments and leaves open the possibility of intermediate stages such as an unequal fusion of hexamers.

*Acknowledgments.* We are thankful for financial support from the New Zealand Lottery Grants Board; the Marsden Fund; the Todd Foundation; the EG Johnstone Trust; the JDS Roberts Trust; the Nellie Milnes Trust; and the Otago Medical Research Foundation; and for a grant to A.M.J. from the Health Research Council of New Zealand.

## References

Antoine M, Niessing J (1984) Intron-less globin genes in the insect *Chironomus thummi thummi.* Nature 310:795–798

Blake CCF (1978) Do genes-in-pieces imply proteins-in-pieces? Nature 273:267

Blanchetot A, Wilson V, Wood D, Jeffreys AJ (1983) The seal myoglobin gene: an unusually long globin gene. Nature 301:732–734

Bogusz D, Appleby CA, Landsmann J, Dennis ES, Trinick MJ, Peacock WJ (1988) Functioning haemoglobin genes in non-nodulating plants. Nature 331:178–180

Cavalier-Smith T (1991) Intron phylogeny: a new hypothesis. Trends Genet 7:145–148

Couture M, Chamberland H, St-Pierre B, Lafontaine J, Guertin M (1994) Nuclear genes encoding chloroplast haemoglobin in the unicellular green alga *Chlamydomonas eugametos.* Mol Gen Genet 243:185–187

Darnell JE, Doolittle WF (1986) Speculations on the early course of evolution. Proc Natl Acad Sci USA 83:1271–1275

Derr LK, Strathern JN (1993) A role for reverse transcripts in gene conversion. Nature 361:170–173

Dibb NJ (1993) Why do genes have introns? FEBS Lett 325:135–139

Dixon B, Walker B, Kimmings W, Pohajdak B (1992) A nematode hemoglobin gene contains an intron previously thought to be unique to plants. J Mol Evol 35:131–136

Felsenstein J (1988) Phylogenies from molecular sequences: inference and reliability. Annu Rev Genet 22:521–565

Gilbert W, Glynias M (1993) On the ancient nature of introns. Gene 135:137–144

Gilbert W (1987) The exon theory of genes. Cold Spring Harb Symp Quant Biol 52:901–905

Gilbert W (1978) Why genes in pieces? Nature 271:501

Go M (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291:90–92

Hurst LD (1994) The uncertain origin of introns. Nature 371:381–382

Hyldig-Nielsen JJ, Jensen EO, Paludan K, Wiborg O, Garrett R, Jorgensen P, Marcker K (1982) The primary structures of two leghemoglobin genes from soybean. Nucleic Acids Res 10:689–701

Jensen EO, Paludan K, Hyldig-Nielsen JJ, Jorgensen P, Marcker KA (1981) The structure of a chromosomal leghaemoglobin gene from soybean. Nature 291:677–679

Jhiang SM, Riggs AF (1989) The structure of the gene encoding chain c of the hemoglobin of the earthworm, *Lumbricus terrestris.* J Biol Chem 264:19003–19008

Kao W-Y, Trewitt PM, Bergtrom G (1994) Intron-containing globin genes in the insect *Chironomus thummi.* J Mol Evol 38:241–249

Landsmann J, Dennis ES, Higgins TJV, Appleby CA, Kortt AA, Peacock WJ (1986) Common evolutionary origin of legume and non-legume plant haemoglobins. Nature 324:166–168

Manning AM, Trotman CNA, Tate WP (1990) Evolution of a polymeric globin in the brine shrimp *Artemia.* Nature 348:653–656

Martinez P, Martin W, Cerff R (1989) Structure, evolution & anaerobic regulation of a nuclear gene encoding cytosolic glyceraldehyde-3-phosphate dehydrogenase from maize. J Mol Biol 208:551–565

Nishioka Y, Leder A, Leder P (1980) Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. Proc Natl Acad Sci USA 77:2806–2809

Roger AJ, Doolittle WF (1993) Why introns in pieces? Nature 364: 289–290

Sharp PA (1985) On the origin of RNA splicing and introns. Cell 42:397–400

Sherman DR, Kloek AP, Krishnan BR, Guinn B, Goldberg DE (1992) *Ascaris* hemoglobin gene: plant-like structure reflects the ancestral globin gene. Proc Natl Acad Sci USA 89:11696–11700

Stoltzfus A, Spencer DF, Zuker M, Logsdon JM Jr, Doolittle WF (1994) Testing the exon theory of genes: the evidence from protein structure. Science 265:202–207

Titchen DA, Glenn WK, Nassif N, Thompson AR, Thompson EOP (1991) A minor globin gene of the bivalve mollusc *Anadara trapezia*. Biochim Biophys Acta 1089:61–67

Trotman CNA, Manning AM, Moens L, Tate WP (1991) The polymeric haemoglobin molecule of *Artemia*. Interpretation of translated cDNA sequence of nine domains. J Biol Chem 266:13789–13795

Trotman CNA, Manning AM, Bray JA, Jellie AM, Moens L, Tate WP (1994) Inter-domain linkage in the polymeric hemoglobin molecule of *Artemia*. J Mol Evol 38:628–636

Welters P, Metz BA, Schell J, de Bruijn J (1989) Nucleotide sequence of *Sesbania rostrata* leghemoglobin (siglb3) gene. Nucleic Acids Res 17:1253

Yamauchi K, Ochiai T, Usuki I (1992) The unique structure of the *Paramecium caudatum* hemoglobin gene: the presence of one intron in the middle of the coding region. Biochim Biophys Acta 1171:81–87