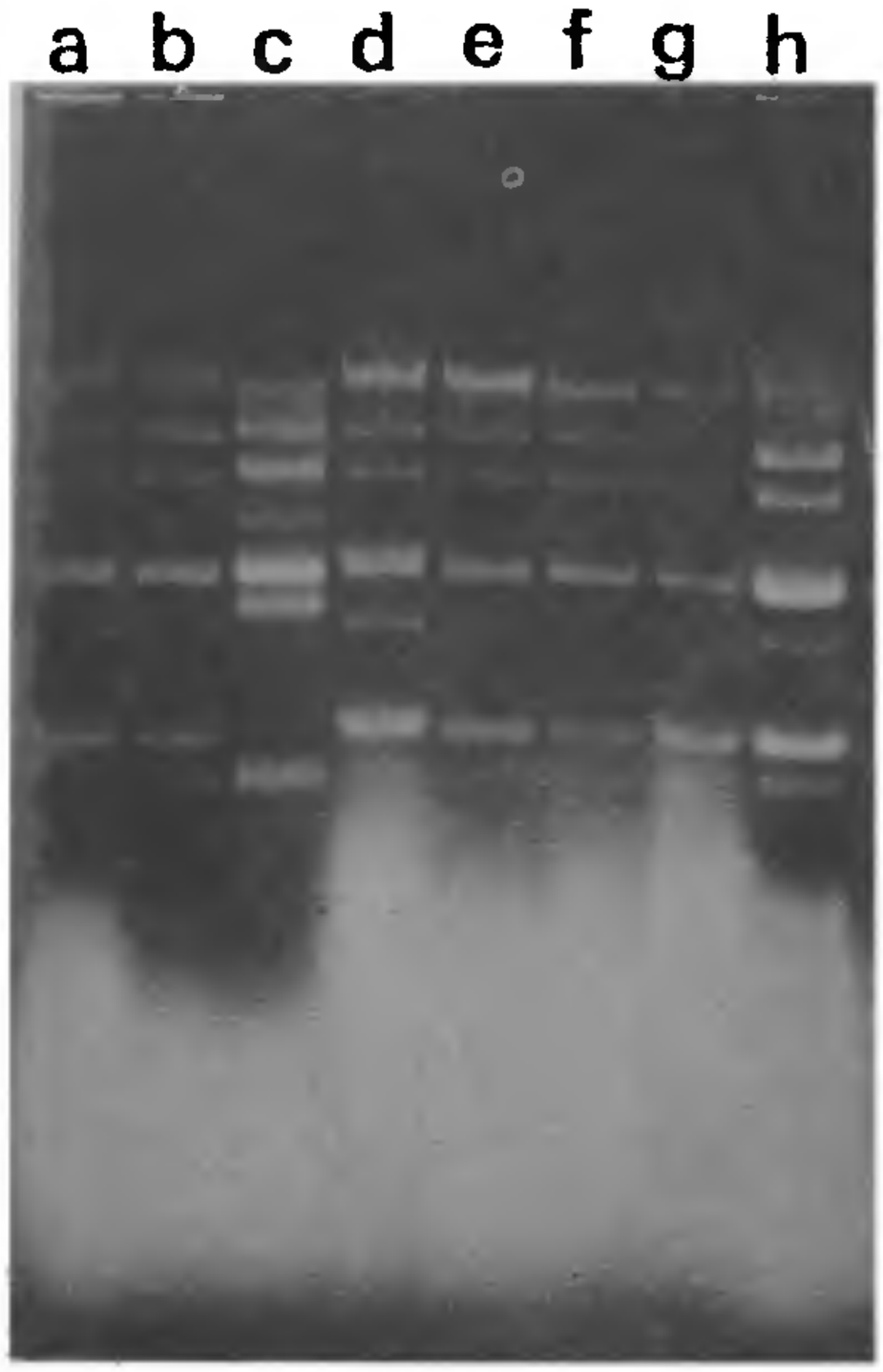

Figure 1. Effect of different concentrations of sodium chloride and sugar solutions on the isolation of pUC18. (a) 150 mM NaCl; (b) 300 mM Nacl; (C) 500 mM NaCl; (d) 1% glucose; (e) 3% glucose; (f) 5% glucose; (g) 8% sucrose; (h) Control.

with restriction enzymes (data not shown).

The method described here uses reagents that are stable at room temperature. In addition, this method uses lower concentration of NaOH and SDS for lysis than that recommended by Birnboim and Doly[1]. Milder conditions of lysis as the one employed here enable the release of small molecules including compact supercoiled plasmids while trapping larger molecules such as denatured chromosomal DNA inside the cell. Moreover, the precipitation time with 3 M potassium acetate is also reduced to 15 min as longer incubations lead to slow leaching of the chromosomal DNA into the solution. The method is consistent, reproducible and gives better yields.

1. Birnboim, H C. and Doly, J., *Nucl. Acids Res.*, 1979, **7**, 1513–1523.
2. Holmes, D. S. and Quigley, M., *Anal. Biochem*, 1981, **114**, 193–197.
3. Kado, C. I. and Liu, S. T., *J. Bacteriol.*, 1981, **145**, 1365–1373.
4. Kamal Chowdhury, *Nucl. Acids Res.*, 1991, **19**, 2792.
5. Kieser, T., *Plasmid*, 1984, **12**, 19–36.
6. Serghini, M. A., Ritzenthaler, C. and Pinck, L., *Nucl. Acids Res.*, 1989, **17**, 3604.
7. Maniatis, T., Fritsch, E. F. and Sambrook, J., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Lab., Cold Spring Harbor, New York, 1982.
8. Cohen, S. N., Chang, A. C. Y. and Hsu, L., *Proc. Natl. Acad. Sci. USA.*, 1972, **69**, 2110–2115.

ACKNOWLEDGEMENTS. We thank Mr V. Ramanathan for testing the procedure on large Agrobacterium plasmid, and DST, New Delhi, for financial support.

Received 18 October 1993; revised accepted 31 January 1994

SHAIK YAZDANI BASHA
P. PALANIVELU

Department of Biotechnology
Madurai Kamaraj University
Madurai 625 021, India

COMMENTARY

Paradoxes in the evolution of introns and genes

G. Naresh Kumar

The recent discovery of class II type of self-splicing introns in bacteria[1] has provided additional information to probe into the perplexing role of introns in the evolution of genes[2–4]. The problem begins with the origin of introns. Introns can be classified into five different categories—class I, class II, class III, spliceosomal mRNA introns, and the unique small tRNA introns. Classes I and II introns both undergo self-splicing *in vitro* but require proteins for efficient splicing *in vivo*[5,6]. They differ from each other in their secondary structures and in their mechanisms of splicing. Class I introns are found in the genes of organelles, bacteria[6–9] and in eukaryotic nuclear genes whereas class II introns, until their recent discovery in

bacteria, had been detected only in organelles. Class III introns are found in the chloroplasts of *Euglena* and *Astasia* and are relatively small in size[10]. Class III introns contain only a subset of class II intron domains and thus require additional domains in *trans* for splicing. The fourth category, mRNA introns found in the protein-coding regions of eukaryotes, vary considerably in size and sequence even between closely related genes, and lack self-splicing ability. A ribonucleoprotein complex, called spliceosome, catalyses their removal from mRNA. mRNA introns have not so far been detected in prokaryotes and in anaerobic eukaryotes. The small tRNA introns are characterized by yet another mechanism of splicing involving removal of the intron by an endonuclease and joining of the exons by an ATP-dependent ligase[11,12]. Two variations in introns are described as outrons and twintrons. Outrons have only one exon adjacent to them and they participate in *trans*-splicing[13] and twintrons are introns within introns which are composite group II and group III introns[10]. A sequential splicing of introns is essential for generating a mature transcript.

Since the discovery of catalytic RNA (ribozymes), a new hypothetical world, the RNA world, has emerged in which organisms contained only RNA[14,15] functioning both as the genetic material and as a catalytic molecule. Organisms developed DNA and catalytic proteins only later. Accordingly, both class I and class II introns, possessing catalytic activity, may have been derived from ribozymes and are probably the most primitive class of introns. Similarities between class II introns, class III and mRNA introns, therefore, have been interpreted as suggesting that evolution of introns occurred in that order[10,16–19].

Self-splicing class I and class II introns encode site-specific recombinase and reverse transcriptase-like protein-coding regions[16,20–22]. Both classes of introns are also mobile. It is therefore important to understand the evolutionary relationship between transposons and introns[23,24]. Since the RNA world is believed to have preceded the present-day DNA world, transposons have possibly been derived from self-splicing introns. The ancestral self-splicing introns probably had no open reading frames (ORFs). Interestingly cyanobacteria, which are considered to be ancient[25], are devoid of ORFs in their group I introns[7,8]. In the early stages, self-splicing introns containing ORFs facilitating splicing would have been selected and then at a later stage, after the transition of genetic material into DNA, ORFs that enabled propagation of introns might have been selected. This view is supported by the role played by the ORFs of the self-splicing introns both in splicing and in their mobility[5,6]. There is relatively greater consensus on these issues; the present controversy is about the origin of mRNA introns. Two different hypotheses view the origin of mRNA introns in living systems as Early and Late.

The early-intron hypothesis has been proposed to account for the origin of diverse and complex proteins[26]. The role of the introns was to generate diversity and complexity without destroying the functional integrity of protein (sub) domains encoded by exons. Interestingly, experimental evidence supporting exon-shuffling was found in terms of recombination events between self-splicing introns[27,28]. Recently, excision of an exon circle from RNA containing group I[29] and group II introns[30] has been demonstrated. This process is known as inverse splicing. This finding supports early hypothesis since a process exactly reverse of inverse splicing could mediate exon shuffling. The occurrence of *trans* splicing of pre-mRNA amongst plants, animals and protists[31] may be a relic of earlier events envisaged in the early intron hypotheses. Analysis of glyceraldehyde phosphate dehydrogenase (GAPDH) gene homologies between bacterial contiguous genes and discontiguous genes of chloroplasts and eukaryotic nuclear genomes, has indicated that the eukaryotic intron-exon organization predates the divergence of eukaryotes and prokaryotes[32].

The early-intron hypothesis predicts that introns would predominantly be found at the boundaries of (sub) domain-encoding regions, rather than within regions encoding (sub) domains. Although there are a few cases that do not conform to this expectation, intron distribution in a variety of genes has been found to be in agreement with such a view[33,34]. Based on the protein domain structure, a new intron position was predicted for the triosephosphate isomerase gene[33], and such an intron was detected in the mosquito *Culex tarsalis*[35]. The early hypothesis also explains the fact that many substrate/cofactor binding (sub) domains have common structural features even amongst proteins with diverse functions[36,37], as well as the widespread occurrence of conserved sequence motifs among otherwise unrelated proteins. Additional surprising evidence came from the nature of exon-intron junction sequences, AGgt, which were suggested to have been derived from terminator codons[38]. In a number of instances, the last codon of exons was derivable from the amber stop codon by a single base change. In some cases the 3′ end codon of exons was found to be the stop codon UAG and translation of these codons in mRNA occurs through suppressor tRNAs, thus supporting the early hypothesis of exons as independent genetic units[39] or microgenes[40].

Since self-splicing requires conservation of certain sequences and tertiary structures, the extensive recombination events in introns envisaged by the early hypothesis could be limited by the loss of self-splicing and catalytic ability[41]. All the undisputed examples of intron-mediated shuffling of exons involve 'modern' eukaryotic proteins without prokaryotic counterparts[41]. A further limitation to the early intron hypothesis discussed by Patthy[41] is that the generation of 'new' genes encoding functional proteins by exon shuffling, is possible only for a minority of intron recombination events. This is based on the observation that the size of all exons is not $3n$, hence the sequence of the protein encoded by an exon is dependent on the reading frame of the exon at its 5′ end. This criticism may be rendered less serious by realizing that present-day exons need not represent their ancestral counterparts because of changes caused not only by mutations but also by the mobility of introns themselves. The early intron hypothesis envisages ancestral exons as microgenes[40] which corresponded to independent protein coding domains and would therefore have been of phase 0 class only.

A weakness of the early-intron hypothesis is the difficulty in accounting for the absence of introns in most of the protein-coding genes of phylogenetically distinct organisms, viz. bacteria, archaebacteria and lower eukaryotes. Some of the reasons which could have been responsible for the loss of introns from these organisms are: (i) the need to decrease the load of maintaining 'unnecessary' genomic DNA in order to achieve faster growth rates to compete better with other

organisms; (ii) elimination of errors in splicing which result in the deletion of some regions of exons leading to mutant proteins; and (iii) the problem of accidental translation of intron regions. (This is a serious problem in the case of prokaryotes since transcription and translation can take place simultaneously and non-functional proteins could be easily generated by the translation of mRNAs before complete splicing). The introns could have got eliminated by reverse transcription of mature mRNA followed by homologous recombination with the chromosomal counterparts[42]. Such an intron loss mediated by reverse transcripts has been demonstrated in *Saccharomyces cerevisiae*[43].

The alternate late-intron hypothesis postulates that intron evolution was independent of the evolution of protein-coding genes and that the capacity of introns for self-propagation has been responsible for their predominance in higher eukaryotes[44]. According to this hypothesis, self-splicing introns invaded the genome of the eukaryotic ancestor by way of endosymbionts which subsequently evolved into mitochondria and chloroplasts containing class II introns. The mechanism of intron invasion into eukaryote genomes is postulated to have involved their insertion into mRNA followed by reverse transcription and recombination with genomic DNA[45]. These later evolved into mRNA introns requiring complex spliceosomes for their removal. Similarities between class II introns, class III and spliceosomal RNAs suggest that fragments from class II introns may have contributed towards the emergence of spliceosomes. The late-intron hypothesis explains the absence of mRNA introns in bacteria and in primitive eukaryotes. The recent discovery of class II introns in bacteria[1] and especially the finding that they do contain reverse transcriptase-like ORFs strongly supports this hypothesis. Propagation of these introns could be similar to the *in vivo* transposition of group II introns which seemed to be mediated by reverse transcriptase[22, 46, 47]. The mechanism by which self-splicing introns got converted into spliceosomally excised mRNA introns seems to have involved class III introns as intermediates. Splicing catalysed by spliceosomes would then render the maintenance of self-splicing ability unnecessary and would also minimize the formation of different mRNAs as a consequence of inverse splicing or *trans* splicing events. A limitation of the late hypothesis is that it requires an additional hypothesis to explain the generation of complex proteins and to account for similarities amongst proteins of diverse functions (i.e. the presence of conserved motifs).

An interesting hypothesis has been suggested as an intermediate between these two contrasting views[4], which is a synthesis of the strong features of both. It shares with the Early hypothesis the proposal that early gene and protein evolution occurred by fusion and recombination of protein domain-encoding regions corresponding to exons. The later stages proposed in this hypothesis correspond to those of the late-intron hypothesis. Accordingly, primitive organisms in the RNA world contained small RNAs corresponding to exons coding for functional peptides. These organisms contained ribozymes with ligation and transesterification properties, which could combine the exonic RNAs to generate many novel combinations. This resulted in the formation of complex proteins. However, these RNAs did not contain introns. Many changes, including the transition to DNA as the genetic material, are expected to have taken place in those organisms, which resulted ultimately in the formation of genes and genomes similar to prokaryotes and lower eukaryotes. Later events of this hypothesis are similar to those in the late-intron hypothesis. Some of the ribozymes would have transformed into introns by gaining the ability to propagate. After the establishment of endosymbiosis between ancestors of eukaryotic cells and organelles, introns could have inserted into mRNA molecules at the junction sites of original exonic RNAs and invaded the genomic DNA by reverse transcription followed by recombination. Insertion of introns into mRNAs is envisaged to be non-random and biased at the relatively unstructured connecting loops between RNA domains which would also correspond to protein (sub)domains.

Although the intermediate hypothesis accounts for many known facts, a weakness could be that it does not explain why ribozyme/introns did not get incorporated into mRNAs and then into the genes of primitive organisms, especially since the hypothesis invokes the occurrence of such an event after the emergence of organelles in eukaryotes. However, this limitation may not be very serious, since it could be argued that the problems encountered in generating functional proteins from intron-containing genes, which are described in the early-intron hypothesis, may have been responsible for the selection of organisms without introns. Another problem with the intermediate hypothesis is that the explanation offered for non-random insertion is not in agreement with the observation that most of the present day prokaryotic and eukaryotic mRNAs do not possess significant secondary structure.

Clearly, the finding of class II type introns in bacteria is a step ahead in our understanding of the puzzling role of introns during early evolution. This finding obviously supports the late hypothesis and has in fact provoked the postulation of the intermediate hypothesis. However, evidence firmly supporting any of the above hypotheses should come from experiments designed to mimic the events postulated to have occurred on earth a few billion years ago. It should also be realized that the reverse transcriptase encoded by class II introns could have played a pivotal role in the transition of the genetic material from RNA to DNA. None of the above hypotheses seems to consider the mechanism of the RNA to DNA transition and that this transition might have been closely associated with the evolution of introns. Another important aspect of gene evolution, not considered by any of the hypotheses presented above, is the *de novo* generation of novel genes—with and without introns—at later stages of evolution, which was essential for the emergence of diversity in prokaryotes and the complexity in eukaryotes.

1. Ferat, J.-L. and Michel, F., *Nature*, 1993, **364**, 358–361.
2. Doolittle, W. F. and Stoltzfus, A., *Nature*, 1993, **361**, 403.
3. Roger, A. J. and Doolittle, W. F., *Nature*, 1993, **364**, 289–290
4. Wistow, G., *Nature*, 1993, **364**, 108–109.
5. Saldana, R., Mohr, G., Belfort, M. and Lambovitz, A. M., *FASEB J.* 1993, **7**, 15–24.
6. Lambovitz, A. M and Belfort, M , *Annu. Rev. Biochem.*, 1993, **62**, 587–622.
7. Kuhsel, M G., Strickland, R. and Palmer, J. D., *Science*, 1990, **250**, 1570–1573.
8. Xu, M -Q , Kathe, S. D., Goodrich-Blair, H , Nierzwicki-Bauer, S. A. and Shub, D. A., *Science*, 1990, **250**, 1566–1570.
9. Reinhold-Hurek, B. and Shub, D. A., *Na-*

ture, 1992, **357**, 173–176.
10. Copertino, D. W. and Hallick, R. B., *Trends Biochem.*, 1993, **18**, 467–471.
11. Abelson, J., *Science*, 1992, **255**, 1390.
12. Phizicky, E. M. and Greer, C. L., *Trends Biochem.*, 1993, **18**, 31–34.
13. Blumenthal, T., *BioEssays*, 1993, **15**, 347–348.
14. Joyce, G. F., *Nature*, 1989, **338**, 217–224.
15. Lamond, A. I. and Gibson, T. J., *Trends Genet.*, 1990, **6**, 145–149.
16. Long, D. M. and Uhlenbeck, O. C., *FASEB J.*, 1993, **7**, 25–30.
17. Jacquier, A., *Trends Biochem.*, 1990, **15**, 351–354.
18. Newman, A. J. and Norman, C., *Cell*, 1992, **68**, 743–754.
19. Madhani, H. D. and Guthrie, C., *Cell*, 1992, **71**, 803–817.
20. Dujon, B., *Gene*, 1989, **82**, 91–114.
21. Perlman, P. S. and Butow, R. A., *Science*, 1989, **246**, 1106–1109.
22. Belfort, M., *Science*, 1993, **262**, 1009–1010.
23. Hickson, R. E., *J. Theor. Biol.*, 1989, **141**, 1–10.
24. Lambovitz, A. M., *Cell*, 1989, **56**, 323–326.
25. Schopf, J. W., *Science*, 1993, **260**, 640–646.
26. Gilbert, W., *Nature*, 1978, **271**, 501.
27. Hall, D. W., Liu, Y. and Shub, D. A., *Nature*, 1989, **340**, 574–576.
28. Bryk, M. and Belfort, M., *Nature*, 1990, **346**, 394–396.
29. Puttaraju, M and Been, M. D., *Nucleic Acids Res.*, 1992, **20**, 5357–5364.
30. Jarrell, K. A., *Proc. Natl. Acad. Sci. USA*, 1993, **90**, 8624–8627.
31. Bonen, L., *FASEB J.*, 1993, **7**, 40–46.
32. Shih, M.-C., Heinrich, P. and Goodman, H. M., *Science*, 1988, **242**, 1164–1166.
33. Gilbert, W., Marchionni, M. and McKnight, G., *Cell*, 1986, **46**, 151–153.
34. Traut, T. W., *Proc. Natl Acad. Sci. USA*, 1988, **85**, 2944–2948.
35. Tittiger, C., Whyard, S. and Walker, V. K., *Nature*, 1993, **361**, 470–472.
36. Dorit, R. L., Schoenbach, L. and Gilbert, W., *Science*, 1990, **250**, 1377–1382.
37. Campbell, W. H. and Kinghorn, J. R., *Trends Biochem.*, 1990, **15**, 315–319.
38. Senapathy, P., *Proc. Natl. Acad. Sci. USA*, 1988, **85**, 1129–1133.
39. Gilbert, W., *Cold Spring Harbor Symp. Quant. Biol.*, 1987, **52**, 901–905.
40. Seidel, H. M , Pompliano, D. L. and Knowels, J. R., *Science*, 1992, **257**, 1489–1490.
41. Patthy, L., *BioEssays*, 1991, **14**, 187–192.
42. Fink, G. R., *Cell*, 1987, **49**, 5–6.
43. Derr, L. K. and Strathern, J. N., *Nature*, 1993, **361**, 170–173.
44. Cavalier-Smith, T., *Trends Genet.*, 1991, **7**, 145–148.
45. Rogers, J. H., *Trends Genet.*, 1989, **5**, 213–216.
46. Mueller, M. W., Allmaier, M., Eskes, R. and Schweyen, R. J., *Nature*, 1993, **366**, 174–176.
47. Sellem, C. H., Lecellier, G. and Belcour, L., *Nature*, 1993, **366**, 176–178.

G. Naresh Kumar is in the Department of Biochemistry, M. S. University of Baroda, Vadodara 390 002, India.