A unique group of self-splicing introns in bacteriophage T4

Asad U Khan*,1, Malik Ajamaluddin1 and Masood Ahmad#

1Interdisciplinary Biotechnology Unit, A M U Aligarh 202 002, India
# Biochemistry Department Faculty of Life Sciences, A M U Aligarh 202 002, India

Received 1 February 2001; accepted 18 May 2001

We describe in this review, the salient splicing features of group I introns of bacteriophage T4 and propose, a hypothetical model to fit in the self-splicing of nrdB intron of T4 phage. Occurrence of non-coding sequences in prokaryotic cells is a rare event while it is common in eukaryotic cells, especially the higher eukaryotes. Therefore, T4 bacteriophage can serve as a good model system to study the evolutionary aspects of splicing of introns. Three genes of T4 phage were found to have stretches of non-coding sequences which belonged to the group I A type introns of self-splicing nature.

Discovery of introns in 1977 heralded a new era in the study of molecular biology of eukaryotic gene expression1. The complexity of gene organization, the combinatorial possibilities of assembling different coding exons from an RNA precursor, and the novelty of the RNA splicing process clearly indicate that eukaryotic molecular biology would be fascinatingly different from that of prokaryotic system2,3. But Morrissey and Tollervey have highlighted a striking similarity between prokaryotic and eukaryotic systems4 that ribosomal RNA processing in eukaryotes, archaeabacteria and bacteria is of common origin can be shown by the fact that the pre-rRNA processing is found in archaeon, Solfolobus acidocaldarius5,6.

In general, prokaryotic genes do not have introns because of energetic demands of rapid generation time, coupled with single origin of replication7. Therefore, it is thought that introns and other non-coding DNA sequences have been kept out of prokaryotes. But an exception to this is the occurrence of three introns in bacteriophage T4. Group I introns have been found in mitochondrial and chloroplast genomes of plants, fungi, a few nuclear rRNA, bacteriophages and also in some fungal plasmids8. Based on all known group I intron sequences, comparative folding analyses and computer search for structure with minimum free energy of folding were performed and generic secondary and tertiary structures were derived9,10.

T4 introns

The bacteriophage T4 genome contains three introns (Fig.1), all of which reside in genes involved in nucleotide biosynthesis. The first T4 intron was discovered in 1984 in the td gene coding for thymidylate synthase11, whereas the other two introns are in the nrdB gene coding for small subunit of ribonucleotide reductase12 and the nrdD gene coding for small subunit of anaerobic ribonucleoside triphosphate reductase13. The reason of the occurrence of introns only in the genes involved in nucleotide biosynthesis is still unanswered. These introns differ in sequence and size; the nrdD and td introns comprise 1033 and 1016 nucleotides respectively, while nrdB intron has only 598 nucleotides. All these introns comprise Open Reading Frames (ORFs). The ORFs are not homologous and occur at three different positions, at least two of which looped out from a secondary structure. Autocatalysis of these introns is enabled by the folding of introns into characteristic secondary structures required for intron excision and exon ligation10,14. The critical ribozyme secondary structure comprises 10 RNA pairing regions (P1-P10) which alongwith conserved primary sequence elements namely P, Q, R and S identify the T4 intervening sequences as group I intron15,16. The P4-P6 domain directs the folding of Tetrahymena ribozyme core17. It is not established that whether the tertiary structure of group I intron is essential for catalytic activity. Basic features of predicted three dimensional (3 D) tertiary structure have been studied by different techniques like photochemical cross-linking reactions, affinity cleavage reagents, NMR spectroscopy and computer algorithm18,22.

---

*Author for correspondence at his present address:
Biochemistry Department, RWJMS, University of Medicine and Dentistry New Jersey 675 Hoes Lane, Piscataway, New Jersey 08854-5635, USA; E mail: asadok@usa.net
Fig. 1—Secondary structure for intron sequences from the nrdB, td and sunY (nrdD) genes
Self splicing of group I intron

Cech and coworkers\textsuperscript{23} found that a ribosomal precursor RNA of Tetrahymena could remove its own 413 nucleotides long intron in the absence of proteins \textit{in vitro}. This reaction proceeds by two consecutive trans-esterification reactions and requires a divalent cation (Mg\textsuperscript{2+} or Mn\textsuperscript{2+}) as well as guanosine (or a guanosine phosphate)\textsuperscript{24}. Divalent cation is needed for the formation of stable secondary structure of group I intron\textsuperscript{25,26}. Recently it has been found that Cobalt ion [Co(III)(NH\textsubscript{3})\textsubscript{6}]\textsuperscript{3+} (ref. 27-29) and other monovalent cation (Li\textsuperscript{+} or even NH\textsubscript{3}\textsuperscript{+}) at higher concentration can substitute the Mg\textsuperscript{2+} (ref. 30). These findings reveal that the fundamental requirement for RNA catalysis is simply one of a relatively positive charge. The 3' end of the 5' exon forms a hybrid with an internal guide sequence (IGS) in the 5' end of the intron. The trans-esterification reaction is initiated by the external guanosine which is held in a guanosine binding site of the intron hydrogen bridges\textsuperscript{31}. The 3'-OH of this guanosine attacks the phosphorus atom at the 5' splice site and forms 3',5'-phosphodiester bond with the first nucleotide of intron. In the second trans-esterification step, the free 3' hydroxyl group at the 3' splice site brings about the ligation of exons and removal of the linear intron (linear intervening sequences), while the extra guanosine remains at its 5' end. During this reaction, the conserved 3' guanosine from the intron remains bound at the G-binding site. Subsequently, the intron itself undergoes a self-catalyzed trans-esterification reaction. The conserved terminal 3' guanosine residue still bound in the G-binding site attacks a phosphorus atom near the end of the molecule resulting in intron cyclization (cyclic intervening sequences) and removal of short 5' oligonucleotide\textsuperscript{32}.

Self-splicing of group I intron is very sensitive to mutation in core of this structure but many of the peripheral stem-loop structures can be deleted without loss of splicing function \textit{in vivo}. These structures may, however, play a role in stabilizing the intron or providing binding sites for proteins which facilitate or regulate self splicing \textit{in vivo}. Introns themselves may in fact encode a protein, involved in their own structural stabilization\textsuperscript{33}.

Splicing of T4 introns

The T4 introns have the ability to self-splice \textit{in vitro}\textsuperscript{13,34}. Autocatalysis is enabled by folding of the intron into characteristic secondary structure that facilitates a series of trans-esterification reactions necessary for intron excision and exon ligation\textsuperscript{8}. The T4 \textit{td} intron contains four long range base paired regions in P3, P6, P7 and P10. The remaining base paired regions are stem-loop structure including P7.1 and P7.2 which coupled with specific variants in the P, Q, R and S sequence elements further sub classify the T4 introns as group IA\textsuperscript{8,15}.

The secondary structure model for group I intron was initially postulated on the basis of phylogenetic comparisons of intron sequences from various organisms which predicted a common secondary structure\textsuperscript{35,36}. Since then, mutational analysis of autocatalytic group I intron in \textit{Tetrahymena}, T4 and yeast mitochondria has identified many splicing defective \textit{cis}-acting intron mutations that would destabilize predicted RNA pairing regions, thus generally supporting the model. In T4, for example, the sequence changes for 31 splicing defective \textit{td} intron mutants (19 phage mutants, 12 mutations from a cloned \textit{td} gene) have been determined and the vast majority disrupt proposed intron RNA helices\textsuperscript{37,39}.

Single base changes in predicted RNA pairing regions which disrupt splicing, provides proof that a nucleotide is essential for the splicing reaction\textsuperscript{40,41}. Genetic verification that a base pair (and thus a pairing region) exists, can be demonstrated by the isolation of a pseudorevertant that harbours a compensatory mutation at the nucleotide predicted to base pair with the original mutant nucleotide. The double mutant would be predicted to regenerate a Watson-Crick base pair in an essential RNA helix, thus permitting intron RNA to fold, resulting in the restoration of some degree of autocatalytic splicing\textsuperscript{42}. Such experiments have been carried out using site-specific mutagenesis to verify the existence of group I intron RNA pairing region in \textit{Tetrahymena} and yeast mitochondria\textsuperscript{43-45}. For the T4 introns, however, only the P6 pairing region of the \textit{td} intron has been proven by isolation of such second site suppressor mutations\textsuperscript{38}.

Conclusion

Our study demonstrated that 3' region of ORF in the \textit{nrdB} intron could also play significant role in splicing\textsuperscript{40}. In contrast, genetic studies on \textit{td} and other group I introns demonstrated that intron ORF is not required for active splicing conformation\textsuperscript{39}. This could probably be due to the small size of ORF of \textit{nrdB} intron (294 bases) compared to that of \textit{td} intron (778 bases) which looped out of the secondary structure models for RNA folding\textsuperscript{36,46}. Present study of the \textit{nrdB} mutants of phage T4 and their revertants
Further leads us to conclude that the conserved sequences of \textit{nrdB} intron that are important in pre-mRNA splicing, especially the extreme ends of \textit{nrdB} intron, were more critical for auto-catalysis. Moreover, the exon sequences immediately adjacent to the splice site were also important for splice site selection\(^{40,42}\). Finally, the data also strongly suggest that the cleavage of intron might require a well defined secondary structure which on folding would probably acquire a typical tertiary structure involving the P9.0 and P10 loops. These secondary and/or tertiary structures are crucial for the splicing of \textit{nrdB} intron like other group I type of introns\(^{47}\).

In view of the above discussion, we would like to propose a plausible model to show the involvement of various domains of \textit{nrdB} intron resulting in the formation of typical tertiary structure (Fig. 2). This hypothetical model for the auto-catalysis of \textit{nrdB} intron is very close to that proposed earlier in case of \textit{Tetrahymena} intron\(^{48}\).

\textbf{References}

2. Green M R (1986) \textit{Annu Rev Genet} 20, 671-708