ABSTRACT

Title of Dissertation: ROLE OF TRANSCRIPTIONAL PARAMETERS ON THE FOLDING OF *TETRAHYMENA* GROUP I INTRON

Sujatha P. Koduvayur, Doctor of Philosophy, 2003

Dissertation directed by: Adjunct Professor Sarah A. Woodson

Department of Chemistry and Biochemistry

Differential elongation rates and pausing patters of polymerases are known to affect co-transcriptional folding of RNAs. The mechanism of coupling is not well studied. This study evaluates whether these factors contribute to the 20-50 fold splicing enhancement seen in vivo for the *Tetrahymena* group I intron. The splicing rates of T7 and *Escherichia coli* (*E. coli*) RNA polymerase (RNAP) transcripts were compared in vitro and in *E. coli*. T7 RNAP is a rapid and highly processive polymerase, while *E. coli* RNAP elongates transcripts more slowly. In a collaborative study with Scott Jackson,

the splicing of transcripts of eukaryotic polymerase I and II (pol I and pol II) were compared in *Saccharomyces cerevisiae*. As the stability of nascent structures can be altered by changing the sequence of transcribed RNA, I have also studied the effect of rRNA exons on the folding of this RNA by comparing the splicing rates of pre-RNAs with *E. coli* and *Tetrahymena thermophila* domain IV rRNA exons in vitro and *E. coli*. In yeast, *S. cerevisiae* and *T. thermophila* rRNA, and GFP mRNA exon sequences were compared.

My *E. coli* and in vitro comparative study reveals that co-transcriptional folding is the reason for the rate enhancement seen in vivo for this RNA. For *E. coli* RNAP transcripts this is effected by the template-dependant, site-specific pausing of the polymerase along the template. For the pre-RNA with *Tetrahymena* rRNA exons this is effected by the low stability of nascent structures that enable rapid rearrangement of non-native structures into native ones.

In yeast, we find that both the polymerase and the RNA processing events affect folding of this intron. Pol I transcripts splice 10-fold better than their pol II counterparts and mRNA processing events retard splicing of short pol II transcripts by 10-fold. Moreover, in yeast, only mutations that increase the fraction of misfolded intermediates are rescued but not those that destabilize the native structure. This folding facilitation is however not dependant on the presence of longer rRNA exons in yeast as it is in *E. coli*. This might be indicative of a different folding facilitatory mechanism in yeast from that seen in *E. coli*

EFFECT OF TRANSCRIPTIONAL PARAMETERS ON THE FOLDING OF THE *TETRAHYMENA* GROUP I INTRON

by

Sujatha P. Koduvayur

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2003

Advisory Committee:

Dr. Sarah Woodson, Chair
Dr. Douglas Julin, co-Chair
Dr. Neil Blough
Dr. John Marino
Dr. Steven Mount
Dr. David Straney

DEDICATION

To Amma, Appa, Paati, Akka, Mouli and Nannu

. TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

1. ATP	Adenosine triphosphate
2. GTP	Guanosine triphosphate
3. CTP	Cytosine triphosphate
4. TTP	Thymidine triphosphate
5. DNA	Deoxyribonucleic Acid
6. RNA	Ribonucleic Acid
7. IVS	Intervening Sequence
8. nt	Nucleotide
9. bp	Base pair
10. EDTA	ethylenediaminetetraacetic acid
11.Tris	tris(hydroxymethyl)aminomethane
12. SDS	sodium dodecylsulphate
13. TES	tris, EDTA, SDS
14. DTT	dithiothreitol
15. YPD	yeast peptone dextrose
16. SC	Synthetic Complete

Chapter 1

INTRODUCTION

RNA molecules perform many functions during the life of the cell. Apart from serving as storage material for genetic information, they also function as messengers for protein translation, regulate gene expression, form the catalytic part of large RNA-protein enzymes and also perform autocatalytic functions (reviewed by Gestland and Atkins 1993; Gestland et al 1999).

To perform these many functions RNA molecules must adopt specific tertiary structures that are conducive for each function. The mechanism of RNA folding is of great interest because, an emergent polyanionic chain attains a complex three dimensional structure in seconds despite the enormous amount of time required to search through all the available conformational states. One way to fold in a biologically relevant timescale would be to fold in a hierarchical manner, by which the linear chain first folds into secondary structures and then form tertiary interactions that lead to the formation of the active structure (Brion and Westhof, 1997).

Most RNA folding experiments have been performed on autocatalytic RNAs as its folding can be monitored by measuring the gain in catalytic activity (where folding is the rate limiting step)(Herschlag and Cech 1990). The *Tetrahymena* ribozyme is one such model system.

TETRAHYMENA GROUP I INTRON AS A MODEL SYSTEM

Location

The *Tetrahymena thermophila* group I intron is naturally found in the domain IV of the 26S rRNA gene (Wild and Gall 1979). The intron interrupts the coding sequence of the 26S rRNA at nucleotide 2215, corresponding to position 1925 in the *E. coli* 23S rRNA (Noller 1991). This region of the rRNA is highly conserved (Noller 1991) and failure to excise the intron results in loss of rRNA function (Price et al 1987). Splicing occurs during processing of the pre-rRNA (Cech and Rio 1979; Din et al 1979). Splicing of this intron is autocatalytic, with no requirement for proteins (Kruger et al 1982; Grabowski 1981) and was later shown to be a member of a group of intervening (intron) sequences that all shared the conserved core structure and splicing mechanism (Michel et al 1982).

Splicing of the Intron

Initiation of self-splicing involves base pairing of the internal guide sequence of the intron with 5' exon sequences just upstream of it, to form the P1 substrate helix (Waring et al 1985). The substrate helix is then positioned in the active site of the intron via non-base pairing interactions (Pyle et al 1992; Strobel et al 1995). Self-splicing of group I introns involves two transesterification reactions (Kruger et al 1982; Zaug and Cech 1982; Fig. 1.1). The first of these is initiated by the attack of the 3' OH group of an

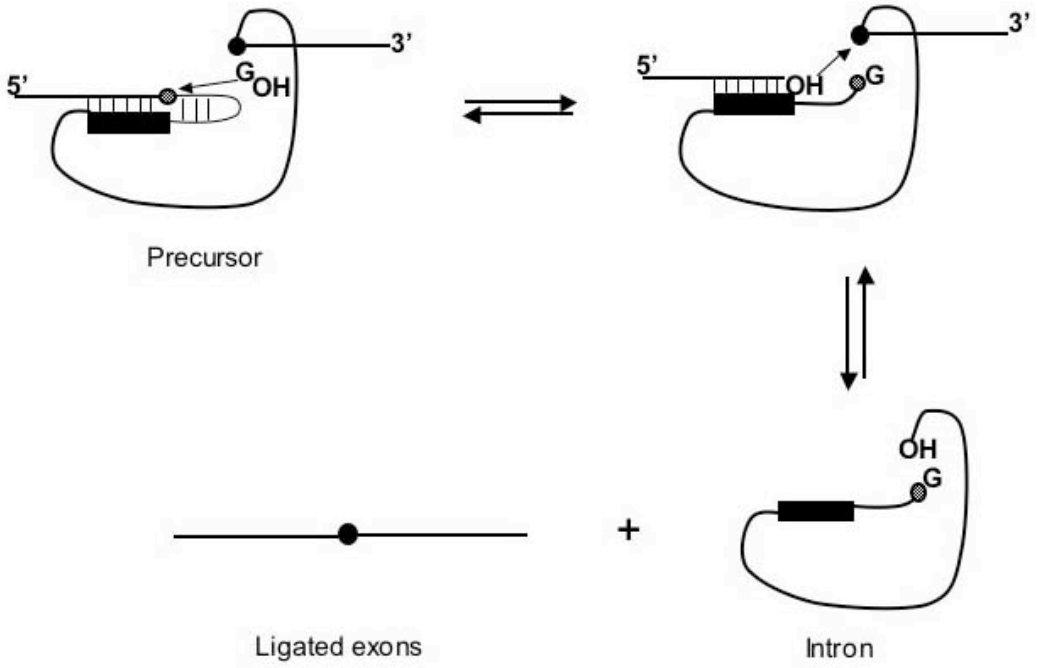


Figure 1.1: In vitro self-splicing of the *Tetrahymena* group I intron. First step of splicing occurs by the attack of the 3' OH of the external guanosine on the 5' splice site (grey circle). Second step of splicing occurs by the attack of the OH group at the 3' end of the 5' exon on the 3' splice site (black circle) leading to the ligation of the exons. Figure adapted from Woodson and Cech 1991.

exogenous guanosine residue on the 5' splice junction, resulting in the addition of the guanosine residue to the 5' end of the intron and elimination of the 5' exon (Bass and Cech 1984). This step is highly specific for the guanosine and is dependent on site-specific binding of the substrate to the intron RNA (Bass and Cech 1984; Michel et al 1989). The binding of the guanosine substrate is known to induce a conformational change in the active site (Herschlag and Khosla 1994 and McConnell and Cech 1995).

The second step of splicing involves another transesterification reaction. The free 3' OH of the 5' exon attacks the 3' splice junction resulting in ligation of the two exons and release of the free intron (Prize and Cech 1989). Recognition of the 3' splice site involves interaction of a conserved guanosine at the 3' end of the intron with the G binding site (Been and Perrotta 1991). The ribozyme must undergo a significant conformational change to enable this interaction (Bevilacqua et al 1996).

Structure of the *Tetrahymena* group I intron

The structure of the *Tetrahymena* group I intron has been extensively studied and found to be a compact, globular form with a core of solvent inaccessible residues (Latham and Cech 1989 and Heuer et al 1991). Comparative phylogenetic sequence analysis (Michel et al 1982, Davies et al 1982, Cech 1988 and Michel and Westhof 1990) and mutagenesis studies (Waring et al 1985, Burke et al 1986 and Couture et al 1990) identified conserved secondary structural elements. These are labeled from P1 to P10 (where P stands for "paired regions") as shown in Fig 1.2. The P1 region represents the substrate helix, that contains the 5' splice site and the IGS (Been and Cech 1986 and Waring et al 1986).

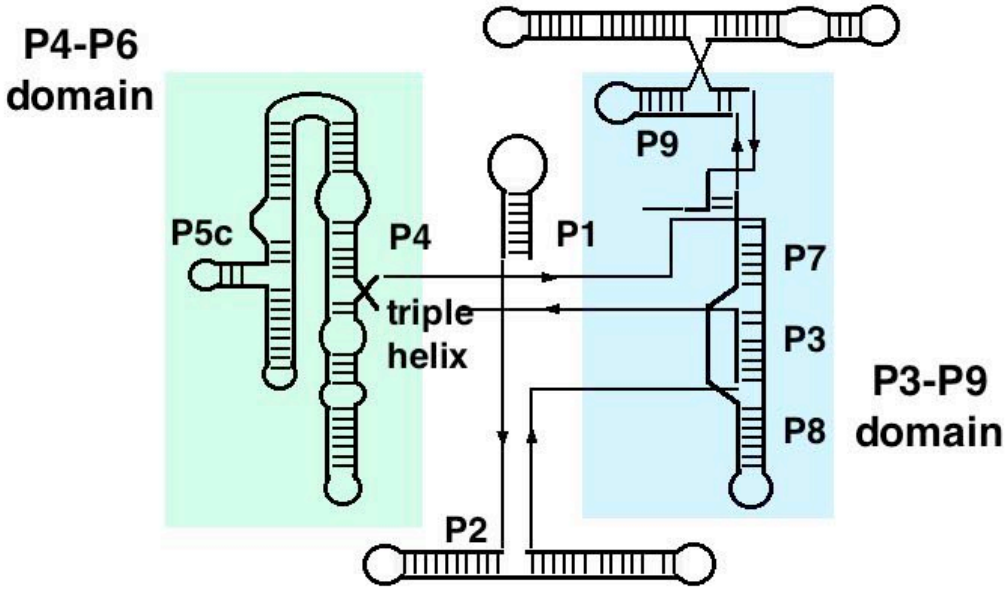


Figure 1.2: Secondary structure of the *Tetrahymena* group I intron. The paired regions are labeled P1-9. The two major domains are shown as is the substrate helix P1 that docks at the junction of the two domains on the triple helical scaffold. Adapted from Cech 1993.

The second domain is formed by the helices P4-P6 that fold independently (Murphy and Cech 1993). The structure of the isolated domain has been studied by X-ray crystallography (Cate et al 1996) and found to consist of two coaxially stacked helices separated by a sharp bend. This domain has the same structure even in the whole ribozyme as seen in the crystal structure by Golden et al (1998).

The P3-P9 domain comprises the helices P3, P7, P8 and P9. The helix P7 contains the G binding site (Michel et al 1989). Together, the P4-P6 domain and the P3-P9 domain form a cleft which forms the catalytic core into which the P1 helix docks (Strobel and Shetty 1997 and Golden et al 1998). At the junction of the two domains (P4-P6 and P3-P9) are three sets of base triples that form a scaffold to help organize the core of the intron (Michel et al 1990, Doudna and Cech 1995 and Zarrinkar and Williamson 1996).

The *Tetrahymena* intron has peripheral regions that are characteristic of the group I C1 subgroup to which it belongs (Fig 1.2). These are the P5abc, P9.1, P9.2, P2.1 and P2.2. These peripheral regions are not essential for catalysis but enhance splicing activity (Joyce et al 1989, Beaudry and Joyce 1990 and van der Horst et al 1991). The peripheral regions interact with each other and wrap around the catalytic core (Lehnert et al 1996). Results from footprinting experiments show that deletion of one or more of these peripheral domains destabilize the core (Murphy and Cech 1993, Laggerbauer et al 1994 and Doherty and Doudna 1997). This indicates that the peripheral domains play a significant role in maintaining structural integrity of the ribozyme. This structural complexity makes this RNA an ideal candidate for studying RNA folding.

Folding of the *Tetrahymena* intron

Folding experiments conducted on the truncated intron missing the first 21 nt and the last 5 nt (L-21 ScaI), have helped tease out the details of the folding pathway of the *Tetrahymena* group I intron. These experiments showed that the P4-P6 domain folds independently (Murphy and Cech 1993). Synchrotron generated hydroxyl radical footprinting assays showed that this region folded at the rate of 1 sec^{-1} (Sclavi et al 1998). Moreover equilibrium folding studies showed that the P4-P6 domain forms tertiary contacts at lower concentrations of magnesium than the rest of the ribozyme (Celander and Cech 1991) and maintains this structure at higher temperatures (Banerjee and Turner 1993). Taken together these experiments and structural information from a high resolution crystal structure (Cate et al 1997) show that folding of the P5abc subdomain followed by the P4-P6 domain are the first steps in the folding pathway of this intron.

Oligonucleotide hybridization experiments have shown that the P3-P9 domain forms slowly at the rate of 0.7 min^{-1} (Zarrinkar and Williamson 1994). Chemical modification and footprinting assays have shown that this domain does not fold independently (Doherty and Doudna 1997). Nucleotides in the peripheral regions (P2, P2.1, P9.1), however, were protected at a rate faster than the P3-P9 domain but slower than the P4-P6 domain (Sclavi et 1998). These experiments show that the slow folding of the P3-P9 region is the rate-limiting step in the folding of the *Tetrahymena* group I intron. The folding pathway is summarized in Fig 1.3.

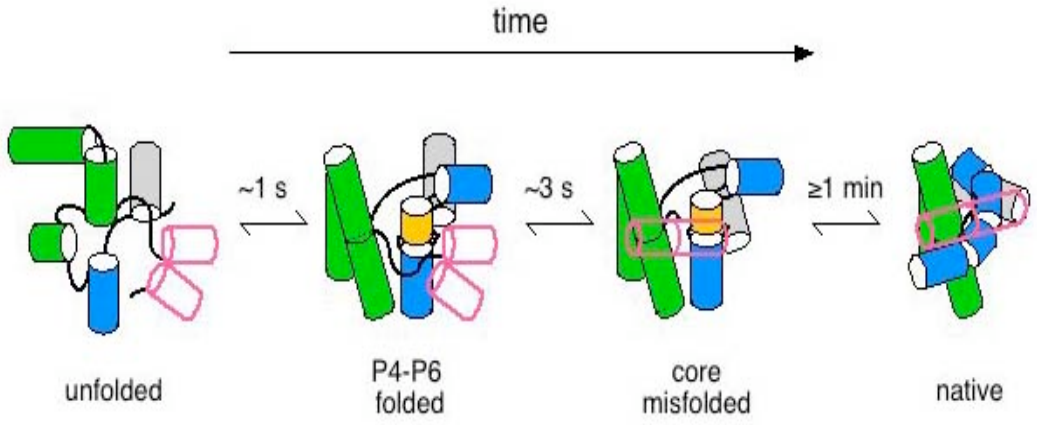


Figure 1.3: Folding of the *Tetrahymena* group I intron. Folding is hierarchical with the P4-P6 domain folded first within second. Folding involves various misfolded intermediates in some of which the core is misfolded. Resolution of these misfolds takes places slowly in the order of minutes to form the native structure (Woodson 2000)

The slow folding of P3-P9 domain provided an ideal system for the study of the nature of the intermediates formed during RNA folding. Study of these intermediates showed that they contained fully folded P4-P6 domain but a disorganized P3-P9 domain (Downs and Cech 1991; Zarrinkar and Williamson 1994; Zarrinkar and Williamson 1996). Chemical modifications and mutational studies showed that the intermediate contained alternate pairing in the P3 helix leading to the formation of the Alt P3 structure (Pan et al 2000) (Fig 1.4).

On the other hand, urea renaturation experiments and mutational studies showed that fast folding of the ribozyme is achieved by destabilizing the native interactions in the P4-P6 domain, suggesting that misfolded intermediates can be stabilized by tertiary interactions at the domain interface (Rook et al 1998; Treiber et al 1998; Zarrinkar and Williamson 1996). Taken together, these observations showed that intermediate conformations that populate the kinetic traps in the folding process can contain both native and non-native interactions (Pan and Woodson 1998; Treiber and Williamson 1998).

To see if the *Tetrahymena* intron follows a similar folding pathway when it is flanked by exon sequences, folding experiments were conducted on the pre-RNA that contains the full-length intron and minimal exon sequences. Modification interference and gel shift experiments show that formation of P3 in the pre-RNA is inhibited by the formation of Alt P3 as in the ribozyme (Pan and Woodson 1998). Formation of this alternate helix is augmented by the formation of another non-native interaction at the 5' splice site (Pan and Woodson 1998).

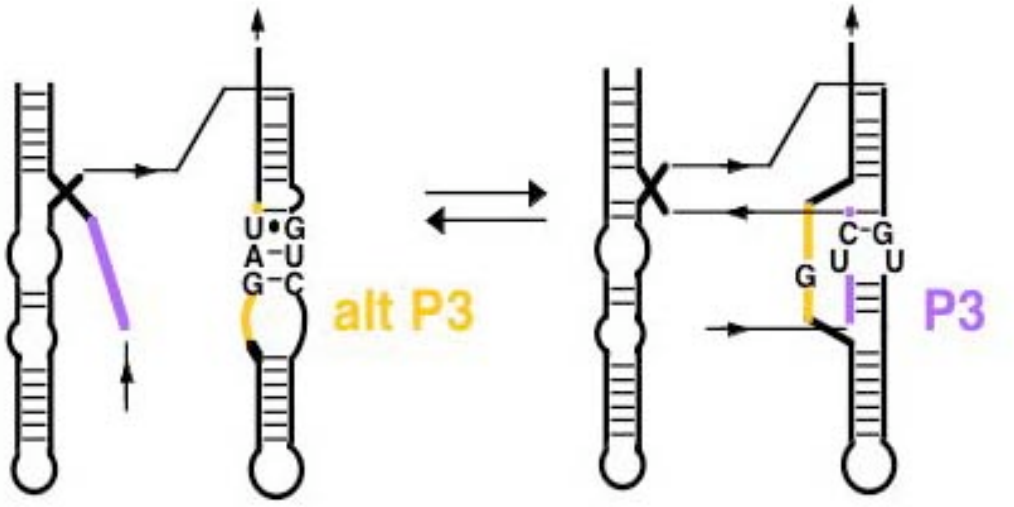


Figure 1.4: Alternate pairings in the core of the intron. The P3 helix of the P3-P8 domain has to form long-range interactions in order to attain the native structure, but slow folding of the domain was found to be due to formation of alternate pairings at the P3 helix leading to the Alt P3 structure. Adapted from Pan and Woodson 1998.

Normally, the internal guide sequence pairs with exon sequences just upstream of the 5' splice-site to form the P1 or substrate helix. However, this region of the 5' exon also forms an alternative stem-loop with sequences upstream forming the P(-1) helix. The P(-1) helix is conserved in the mature rRNA but is inactive in self-splicing (Woodson and Cech 1991). Formation of P(-1) is preferred in the native rRNA and competes with P1, thereby inhibiting splicing.

The P (-1) hairpin can be competed away by the formation of PX (Woodson and Emerick 1993). PX is a long-range interaction between sequences 82 nucleotides upstream of the splice site with sequences that form part of the stem of P (-1). Exon deletion and oligonucleotide binding assays showed that when PX is not formed, the splicing rate falls almost 5-fold (Woodson, 1992, Cao and Woodson, 1998, Woodson and Emerick, 1993) (Fig 1.5). The formation of P (-1) helix is shown to correspond to the appearance of the alt P3 structure in the core of the intron as well as a non-native interaction between the IGS and the 5' strand of P3 (Pan and Woodson 1998), which leads to a misfolded pre-RNA. Mutation experiments conducted by Pan and Woodson showed that when P(-1) is formed, the IGS can base pair with the 5' strand of P3 thus causing the core to misfold (Pan and Woodson 1998). Urea denaturation experiments show that partial unfolding of the pre-RNA accelerates folding, suggesting that formation of the native pre-RNA, like the ribozyme, requires destabilization of metastable intermediates (Pan et al 1997).

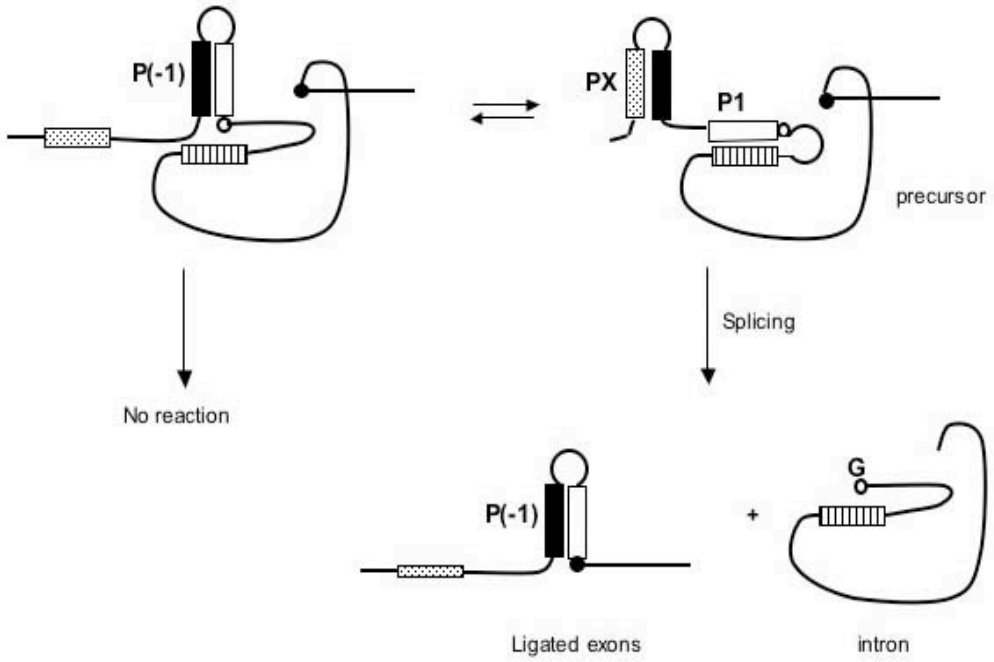


Figure 1.5: Misfolding in the 5' exon leading to loss of activity. Figure shows the competing interactions that can take place at the 5' exon-intron region leading to measurable difference in activity. The IGS is shown as a striped box. The P(-1) interaction is highly conserved and preferred in the natural rRNA. Formation of the alternate PX helix enables the formation of the P1 helix that results in splicing activity. Adapted from Woodson and Emerick 1993.

FOLDING OF LARGE RNAS VIA KINETIC PARTITIONING MECHANISM

The slow folding of the *Tetrahymena* ribozyme and pre-RNA show that kinetic folding of large multidomain RNAs is plagued by kinetically trapped misfolded intermediates that indicate a rugged free energy landscape for folding (Trieber and Williamson 1999). Moreover, the presence of these varied intermediates indicate that the RNA can fold via multiple pathways. The energy landscape models proposed to explain folding of proteins (Dill and Chan 1997) is used to explain folding of this RNA via a Kinetic Partitioning Mechanism (KPM) (Thirumalai and Woodson 1996).

According to this mechanism, the RNA folds through multiple pathways. Starting from a pool of unfolded RNA, a fraction (Ø) of the RNA collapse into the native structure while the rest of the RNAs follow a pathway that is filled with folding intermediates (Fig 1.6). This partitioning of the RNA population was substantiated by experimental data by Emerick and Woodson (1993). Using native gel electrophoresis, they have shown that for the *Tetrahymena* pre-RNA a mixed population of native and non-native intermediate conformers are in slow exchange at 30˚C (Emerick and Woodson 1993). Approximately 70% of the pre-RNA is misfolded after transcription at 30˚C, but renaturation of the RNA decreases this to 20%. This suggests that the misfolded population can refold into the native conformation (overcome energy barriers) only upon unfolding via denaturation.

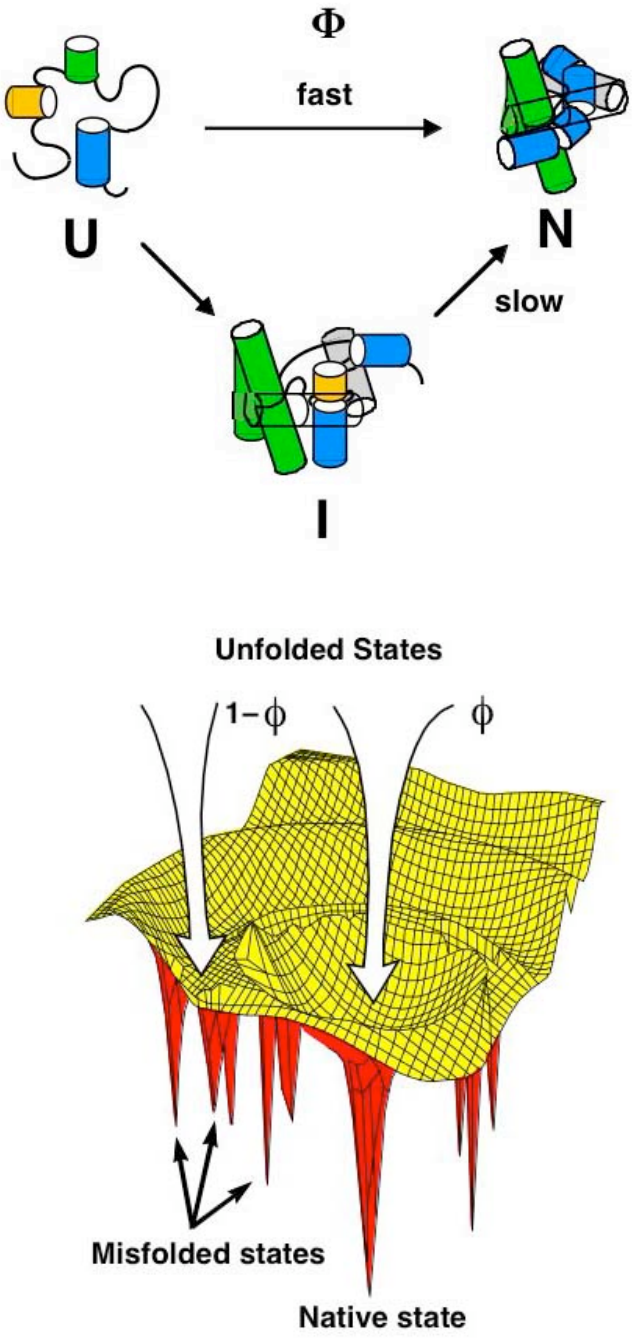


Figure 1.6: Folding of large RNAs. The top panel shows the multiple folding pathway seen for the *Tetrahymena* group I intron where a fraction (Ø) of the unfolded RNA collapses quickly into the native structure while the remaining (1- Ø) RNAs are trapped in misfolded states that refold at a slow rate to the native state. The bottom panel shows the rugged free energy landscape that results in the multiplicity of in vitro folding pathway of large RNAs. Figure from Pan et al 1999 and Thirumalai and Woodson 1996.

Results from experiments on the *Tetrahymena* group I intron and circular permutation studies conducted on another large RNA, the RNase P RNA (Pan and Sosnick 1997; Pan et al 1999), show that the folding landscape for these large RNAs is rugged (Fig 1.6). The intermediates formed during folding populate the local energy minima, and the longer folding times reflect the need to overcome these energy barriers to reach the native state. This leads to topological frustration, which increases with the length of the RNA sequence.

The need to overcome energy barriers to achieve native conformation suggests that in vivo, folding of these RNAs must be assisted, perhaps via chaperones.

FOLDING OF LARGE RNAS IN VIVO

Our knowledge of RNA folding in vivo comes from studies conducted on catalytic RNA, because folding can be measured from their catalytic activity. The hairpin ribozyme is a small autocatalytic RNA whose in vitro and in vivo folding has been extensively studied (Fedor, 2001). Mutation studies and cleavage assays show that the folding of this RNA is the same in yeast and in vitro (Donahue and Fedor, 1997, Donahue et al, 2000 and Yadava et al, 2001). This suggests that folding of these small RNAs occurs by a concerted collapse into the native structure, forming no metastable intermediates, as seen in vitro.

The group I intron in the thymidylate synthase mRNA of phage T4 is a large RNA that folds inefficiently in vitro in the absence of StpA protein (Zhang et al 1995b; 1996). The in vivo folding pathway of this intron has been shown to follow the same

principle of in vitro folding. Chemical probing of the wild type and mutant intron structure in vivo via DMS has shown that this intron folds hierarchically with the P4-P6 domain folding first and the P3 helix folding last (Waldisch et al 2002). Moreover, exons were also found to play a significant role in the folding of the *td* pre-RNA in vivo. Mutational studies showed that splicing in vivo is inhibited by interactions of the 5' exon with the 3' end of the intron, which was disrupted by translating ribosomes (Semrad and Schroeder 1998).

The group I intron of the *Tetrahymena* 26S rRNA, however splices 20-50 fold faster in vivo than in vitro (Brehm and Cech 1983), suggesting a different folding pathway (Zaug and Cech 1980). Splicing of this intron from an mRNA in *E. coli* showed that this intron can splice in prokaryotes at a rate sufficient to support gene expression (Prize and Cech 1985).To see if this in vivo facilitation requires any species-specific protein factor splicing rates of this RNA from the *E. coli* 23S rRNA was studied in *E. coli* (Zhang et al 1995a). The intron was inserted into 23S rRNA of *E. coli* after nucleotide 1926U, which is nearly homologous to the natural splice junction in the *Tetrahymena* rRNA. This preserves the conserved U•G wobble at the base of the P1 stem (Barford and Cech 1989). In order to maintain complementarity with the *E. coli* exon sequence, the internal guide sequence of the intron was changed and this altered intron was called EC-IVS RNA (Fig 1.7). The Ec-IVS RNA showed wild type splicing activity in vitro. More importantly this pre-RNA showed similar splicing enhancement in *E. coli* (Zhang et al 1995a) as seen in *Tetrahymena* (Brehm and Cech 1983). This showed that the splicing enhancement seen in vivo is not due to any species-specific

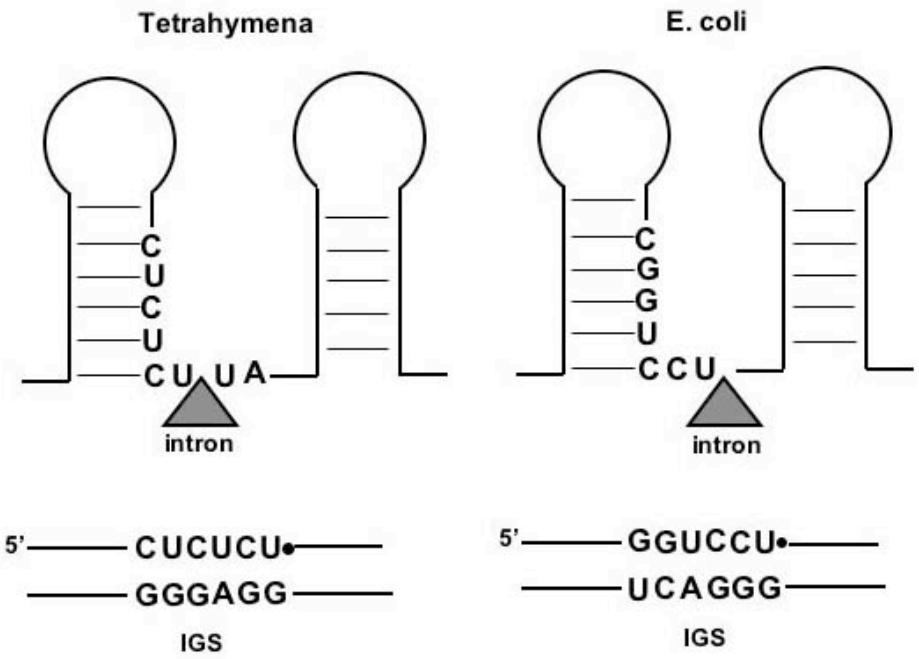


Figure 1.7: Location of the intron in the 23S rRNA of *E. coli*. In *Tetrahymena* the group I intron is present after the nucleotide 1925U in the 26S rRNA as shown on the left. For studies in *E. coli*, the intron was inserted after the nucleotide 1926U in the 23S rRNA as shown on the left in order to maintain the conserved U•G wobble. To ensure complementarity with the 5' exon, the IGS was changed from the sequence on the left to that on the right. Adapted from Zhang et al 1995a.

protein factor and suggested that splicing enhancement in vivo is due to some intrinsic facilitation of folding of the RNA.

To understand this folding facilitation, Nikolcheva and Woodson (1999) measured the effect of mutations on splicing rates of the pre-RNA in *E. coli*. It was shown earlier that splicing of the *Tetrahymena* pre-RNA is enhanced in vitro in the presence of longer exon sequences (Woodson 1992). To see if the length of the exons affect in vivo splicing (and hence folding) in a manner similar to that seen in vitro, Nikolcheva and Woodson (1999) studied the effect of long and short exon sequences on splicing in *E. coli*. Their findings showed that point mutations in the exons that reduce in vitro folding of the wild type pre-RNA about 100-fold were rescued in *E. coli*. However, mutations in the intron that reduce the rate of folding in vitro caused splicing to be cold-sensitive in *E. coli*. The cold-sensitivity of mutations in the intron core was partially relieved when the exons included longer rRNA. As folding facilitation in *E. coli* seems to involve secondary structure of the exons, Nikolcheva and Woodson (1999) propose that this could involve competition between native and non-native secondary structure interactions like P1 and P(-1) as seen in vitro.

FACTORS THAT AFFECT FOLDING OF RNA

Given the complexity of the RNA folding problem, it is important to identify and study the various factors that could affect this process both in vivo and in vitro, in order for us to be able to better understand it.

Proteins in RNA folding

Studies conducted on trans splicing of the *td* intron and accelerated folding of the hammerhead ribozyme have brought to light several proteins with varied cellular function that show RNA chaperone activity in vitro (Zhang et al 1995b). Some of these are the *E. coli* ribosomal protein S12, the transcriptional regulator StpA, eukaryotic hnRNP protein A1, transcription factor p53, prion protein PrP and HIV-1 nucleocapsid protein NCp7 (Weeks 1997, Coetzee et al 1994, Herschlag et al 1994, Nedbal et al 1997, Gabus et al 2001 and Tsuchihashi et al 1993). In vivo chaperone activity has been attributed to StpA, S12, Hfq and NCp7 (Clodi et al 1999, Zhang et al 2002 and Schroeder et al 2002). These RNA chaperone proteins are proposed to interact nonspecifically with single stranded regions in the intron and promote folding by either resolving non-native conformations or impeding their formation (Herschlag 1995). However, there are instances of proteins that facilitate RNA folding by binding to specific sites on the RNA. The group I intron maturases fall under this category.

Some group I introns require proteins to splice under physiological conditions (Herschlag 1995). In these, stabilization of RNA's native structure occurs with the help of proteins (Weeks 1997). These proteins are either encoded by the intron itself (maturases) or are host proteins present in the cell (Weeks and Cech 1995). In vitro formation of native structure is facilitated by high levels of magnesium ions or by addition of proteins. Thus the role of proteins in folding of the RNA is similar to the role of ions, i.e. stabilization of native structure.

Two nuclear encoded proteins have been studied extensively and found to promote splicing in vitro and in vivo: CBP2 and CYT18. CBP2 is a yeast protein that is

found to specifically promote splicing of the fifth intron (bI5) in the mitochondrial cytochrome b (COB) gene (McGraw and Tzagoloff 1983). CTY18 is the mitochondrial tyrosyl tRNA synthetase of *Neurospora crassa*, which promotes the splicing of various group I introns in the mitochondrion (Mannella et al 1979 and Wallweber et al 1997). The CBP2 protein promotes splicing of the bI5 intron by a tertiary structure capture mechanism. The secondary structure of this intron is largely formed in the absence of the protein and folds slowly to form the P3-P9 domain and a peripheral structure called the P7.1 (Shaw and Lewin 1995 and Weeks and Cech 1995b). This tertiary structure is the partial binding site for the protein and is not a stable structure in low magnesium concentrations. The protein however binds strongly to this structure and stabilizes the active intron (Weeks and Cech 1995a, 1995b and 1996). CBP2 also stabilizes the tertiary interaction between the intron and the 5' domain (P1-P2). The association rate for protein binding is limited by the equilibrium between the unfolded and folded states of the bI5 intron (Weeks and Cech 1996).

The CYT-18 protein, binds to the folded P4-P6 domain (Weeks and Cech 1995) and stabilizes the triple helical geometry at the junction of P4 and P6 helices. This protein-RNA complex is thought to provide a scaffold for the assembly of the rest of the intron, which provides more protein binding sites (Saldanha et al 1995, 1996 and Caprara et al 1996a, 1996b and 2001).

Group I intron maturases facilitate folding by a mechanism that is different from the other two protein assisted RNA folding mechanisms described above. Most group I intron maturases are also DNA endonucleases that promote lateral transfer of their encoding intron via a mechanism called homing (Jurica and Stoddard 1999 and

Chevalier and Stoddard 2001). The well-studied maturase protein is the one encoded by the *Aspergillus nidulans* (A.n.) mt COB group I intron (I-AniI), which has shown site-specific DNA endonuclease activity (Ho et al 1997). RNase T1 footprinting assays show that I-AniI binds to the A.n. COB intron that has little or no tertiary structure in the absence of the protein (Ho and Waring 1999). Deletion experiments showed however, that the peripheral domains that stabilize the intron tertiary structure are necessary for protein-assisted splicing activity, suggesting that extensive RNA folding is necessary for protein recognition (Geese and Waring 2001).

Recent kinetic and splicing assays along with RNA deletion analysis have helped answer the question of how I-AniI recognizes the unfolded or misfolded RNA (Solem et al 2002). These experiments have helped propose a model for this protein assisted folding mechanism. According to this model, the protein first preassociates with an unfolded intron producing a labile interaction that facilitates correct folding of the intron catalytic core, either by resolving misfolds or narrowing the number of conformations sampled in the search for native RNA. The active intron structure is then stabilized by specific interactions of the protein to its intron binding site (Solem et al 2002).

The group I intron of the *Tetrahymena* 26S rRNA gene does not need any protein chaperone activity to fold into the active structure in vitro. Despite having a higher splicing rate in vivo, there is no species-specific protein factor involved in the folding enhancement in vivo for this intron (Zhang et al 1995). This indicates that the in vivo folding facilitation for this intron could be due to generic cellular mechanisms. One such mechanism could be folding during transcription. Folding during transcription has been known to enhance folding of RNA. This is discussed in the following section.

Co-transcriptional folding of RNA

As mentioned earlier, hierarchical folding of RNA can best explain how RNA folds much faster than predicted if it were to sample all available conformations (Brion and Westhof 1997). This hierarchical folding has been shown to take place during synthesis of the RNA strand. Using MDV-1 RNA transcribed by Qβ replicase, Kramer and Mills (1981) showed that the nascent RNA is capable of forming secondary structures. Moreover, using ribonuclease cleavage assays they have shown that these nascent structures are capable of reorganizing into other structures as the chain grows. Since then, a number of studies have shown that formation of metastable folding intermediates during transcription is a central theme in various regulatory mechanisms.

The maintenance of plasmid ColE1 copy number involves interactions of the primer precursor RNA and its antisense RNA, RNA I (Tomizawa and Itoh 1981 and Lacatena and Cesareni 1981). Biochemical evidence for a particular structural reorganization during primer transcription involving the first 200 nucleotides has been obtained (Wong and Polisky 1985). This correlates with the observation that the inhibition can take place only in a particular window during transcription (Masukata and Tomizawa 1986). Using genetic and biochemical assays, Polisky et al (1990), showed that the conformational rearrangement of the primer RNA involves a dynamic helix exchange in a transcription window between positions 140 and 144 which is necessary for loop-loop interactions with the RNA I. They also provided evidence that this conformational rearrangement is due to the pausing of the polymerase at specific sites along the template.

Co-transcriptional folding is also found to play a significant role in alternative splice site selection and exon skipping during processing of eukaryotic mRNA. Roberts et al (1998) used spacer elements to delay the synthesis of a downstream inhibitory element, showing that the decision to splice or repress exon 3 of the □-tropomyosin gene occurs during a limited window following transcription in smooth muscle cells. Introducing pause sites upstream of this inhibitory element enhances inclusion of this exon suggesting that pausing of the polymerase during transcription plays a significant role in mRNA splicing. Recent studies on mRNAs in human cells, *Drosophila* and *S. cerevisae* (de la Mata et al and Howe et al 2003) have shown that splicing efficiency and fidelity is influenced by elongation rate of the polymerase, thus strengthening the observations of Roberts et al (1998).

Apart from plasmid replication and post-transcriptional processing reactions, metastable RNA structures formed during synthesis is also known to play a significant role in translational control. Evidence for this comes from studies conducted on the *hok*/*sok* locus of the plasmid R1 that exhibits the post-segregational killing mechanism (PSK) (Gerdes et al 1997). The PSK mechanism restricts Hok toxin production to plasmid-free cells and this occurs post-transcriptionally. This mechanism involves two small metastable hairpins found at the 5'end of the hok mRNA transcript that prevents formation of the longer tac stem-loop that is essential for translation (Gultyaev et al 1997). In vitro experiments show that these structures are indeed formed during transcription thus preventing initiation of translation before the whole RNA is transcribed (Nagel et al 1999). Genetic and biochemical structure probing assays have

shown that these metastable intermediates are formed during in vivo transcription too (Moller-Jensen et al 2001).

These studies not only provide evidence for the occurrence of RNA folding during transcription but also argue for the biological relevance of this phenomenon. Studies conducted on catalytic RNA have shown that co-transcriptional folding affects formation of native structures (Pan et al 1999). Depending on the folding pathway, cotranscriptional folding can either enhance or slow down the formation of the native structures. If sequestering upstream sequences so that they are not available to form non-native interactions is important, then co-transcriptional folding should enhance the formation of native structures. If the nascent structures formed represent stable non-native interactions that must be disrupted to form the native conformation, then cotranscriptional folding would retard the folding process.

Transcription by T7 RNAP was found to produce defective *E. coli* rRNAs (Lewicki et al 1993) suggesting that elongation rate of the polymerase affects co-transcriptional folding. Studies on folding of the RNase P RNA showed that transcription of the wild type or a circularly permuted variant (CP) of this RNA by *E. coli* or T7 RNAP changed the folding pathway from those observed during Mg^{2+}-induced refolding (Pan et al 1999). Moreover, pausing of *E. coli* RNAP was found to play a significant role in this folding process as folding of the wild type *E. coli* RNAP CP RNA transcript was 10-fold faster in the presence of NusA than that of a mutant *E. coli* RNAP transcript that was deficient in pausing and NusA response.

To summarize, co-transcriptional folding of RNA could significantly alter formation of metastable intermediates. Polymerase elongation rates, pause sites and the

stability of nascent RNA structure all affect this process. The folding of the *Tetrahymena* group I intron is plagued by the presence of various metastable intermediates in vitro, but its splicing and hence folding in vivo is enhanced by an as yet unknown mechanism.

I wanted to see if this enhancement could be due to a difference in the synthesis of this RNA. In previous in vitro studies, the RNA was transcribed by T7 RNAP that has an elongation rate of about 230 nt per sec (Golomb and Chamberlin 1974). Conversely, in vivo studies have relied on transcripts by the host polymerase that is either *E. coli* RNAP (elongation rate of about 10-35 nt per sec; Chamberlin et al 1979 and Wang et al 1998) or eukaryotic RNA pol I (elongation rate of about 60 nt/sec in yeast; French et al 2003). Both of these polymerases have a significantly lower elongation rate than phage RNAP and in the case of *E. coli* RNAP there is significant pausing during transcription. This could more significantly affect the folding pathway of the transcripts as seen for T7 synthesized rRNA transcripts (Lewicki et al 1993).

Because the choice of polymerase can either enhance or retard folding, I decided to study the folding of the *Tetrahymena* intron in different exon contexts. The two different exon contexts we chose were the domain IV of *E. coli* 23S rRNA and the native *Tetrahymena* 26S rRNA. These exons have a highly conserved secondary structure, but different primary sequence. We measured folding by assaying splicing activity as the observed splicing of this pre-RNA is determined by the fraction of correctly folded RNA. The specific goals for this project are outlined below.

SPECIFIC GOALS

As discussed in this chapter, in vitro folding of the *Tetrahymena* group I intron is slow due to the presence of misfolded intermediates that result from the ruggedness of the folding landscape. We infer that folding is more efficient in vivo as splicing is 20-50 faster in vivo (Brehm and Cech 1983 and Zhang et al 1995a). In my thesis, I address this differential folding by comparing splicing rates of pre-RNA in *E. coli*, in vitro and in yeast.

1. **Effect of Polymerases and Exons on the Splicing of *Tetrahymena* group I intron in *E. coli***. Polymerases with differing elongation rates and pause sites are found to affect the folding pathway of RNAs. This polymerase effect is dependent on the stability of the nascent structures that are formed during chain elongation. In this study I measure the effect of T7 and *E. coli* RNA polymerases on the splicing of the *Tetrahymena* intron in *E. coli*. The templates that these two polymerases transcribe contain either the *E. coli* 23S rRNA domain IV sequence or the *Tetrahymena* 26S rRNA sequence. This comparative study will give us some insight into the folding facilitation found in *E. coli*.

2. **Effect of Polymerases and Exons in the Splicing of the *Tetrahymena* Group I intron in vitro**. Co-transcriptional folding assays on this *Tetrahymena* group I intron have shown that folding during transcription is two-fold faster than Mg^{2+}-induced folding in vitro (Heilman-Miller and Woodson 2003). This rate increase does not depend on the nature of the polymerase that is transcribing it. This is still a small increase compared to the 20-50 fold splicing enhancement seen in vivo. However, this study was

conducted on just the ribozyme and not on the pre-RNA. In order to make a better comparison between in vitro and in vivo splicing conditions, I compare the splicing of intron from similar pre-RNAs that I used for studying splicing in *E. coli*. I also study the effect of S30 extracts on the self-splicing of this intron from the Ec-T7 pre-RNA to see if any ribosomal proteins could be the reason for splicing enhancement seen in vivo.

3. **Effect of Polymerases and Exons on the Splicing of *Tetrahymena* Group I intron in Yeast**. As *Tetrahymena* is a eukaryote we decided to study the effect of eukaryotic RNA polymerases and exons on the folding of this intron in yeast. We compared splicing rates of the intron when transcribed by either pol I or pol II polymerase. To study the effect of exons on the splicing of this intron in yeast, we compared the splicing of this intron from pre-RNA containing minimal native *Tetrahymena* 26S rRNA exons, minimal Saccharomyces cereviseae 26S rRNA exons, or mRNA (GFP) exons. To study effects of polI transcription on in vivo folding we used the clone that contains the *Tetrahymena* intron in every rDNA chromosomal repeat. The yeast experiments were conducted as part of a collaborative project with Scott A. Jackson, who helped me with yeast growth, RNA isolations, northern blots and data analysis.

Chapter 2

EFFECT OF TRANSCRIPTIONAL PARAMETERS ON FOLDING OF THE *TETRAHYMENA* GROUP I INTRON IN *E. COLI*

INTRODUCTION

The splicing rate of the *Tetrahymena* group I intron in *Tetrahymena* is 20-50 fold faster than the self-splicing rate in vitro (Brehm and Cech, 1983 and Bass and Cech, 1984). In vitro it was found that the rate of the chemical step is very high and the slower splicing rate seen is thought to be due to slower conformational steps (Herschlag and Cech 1990). This indicated that the enhanced splicing seen in vivo is due to facilitated folding of the intron. To see if this facilitation is due to a species-specific protein factor, the intron was inserted near the homologous position in the domain IV of the 23S rRNA of *E. coli* (Zhang et al, 1995a). The splicing of this heterologous pre-RNA in *E. coli* was comparable to that seen in *Tetrahymena* (Brehm and Cech 1983 and Zhang et al 1995a). This suggested that the folding facilitation seen in vivo was not due to any species-specific protein factor such as a maturase.

Role of exons in folding facilitation

As local and long-range RNA interactions can alter the balance between active and inactive structures during in vitro folding of the pre-RNA (Woodson and Cech 1991, Woodson 1992 and Pan and Woodson 1998), studying the role of exon sequences on in vivo splicing could give us insights into the in vivo folding of this intron. Nikolcheva

and Woodson (1999) showed that point mutations in the 3' exon that reduce in vitro folding to about a hundred-fold (Woodson, 1992) had no effect on in vivo splicing of this intron. However, intron mutations that cause misfolding in vitro reduced splicing in *E. coli* (Nikolcheva and Woodson 1999). They also showed that splicing of mutant precursors with natural rRNA exons were less cold sensitive than those with non-rRNA exons. This suggests that the ribosomal RNA exon sequences facilitate folding of this intron in vivo.

If a species-specific protein does not facilitate in vivo splicing of the *Tetrahymena* group I intron, perhaps conserved features of the rRNA exon sequences play a role in this facilitation process. In addition, sequential folding of the ribozyme was shown to be two fold faster than refolding in vitro (Heilman-Miller and Woodson, 2003). In this study I test the effect of sequential folding during transcription on the in vivo splicing of this pre-RNA by changing elongation rate of the polymerase. I also compare rRNA exons from *Tetrahymena* and *E. coli*.

MATERIALS AND METHODS

Plasmids

Plasmid pTN112 contains the EC variant of the *Tetrahymena* IVS inserted after U1926 in *E. coli* 23S rRNA (Zhang et al 1995a) and is described in Nikolcheva and Woodson, (1997). Primers TN112BAM 5'E and TN112BAM 3'E (primer sequences in Appendix A) that introduce the BamH I restriction enzyme recognition site at both ends of the insert were used to PCR amplify the pre-RNA region of wild type and C260G mutant

(Nikolcheva and Woodson, 1999) of pTN112. The purified PCR product containing the pre-RNA sequences was ligated into vector pTZ19R digested with Sma I. Positive clones were selected by picking blue colonies from a β-galactosidase screen. The vector containing the desired fragment was digested with BamH I. The 750 bp fragment was purified and then ligated into BamH I digested pET21 (Novagen) and pSE380 vector (Invitrogen). The former has an IPTG-inducible T7 promoter upstream and a T7 terminator downstream of the insert. The latter has an IPTG-inducible *E. coli trc* promoter upstream and an rrnB termination sequence downstream of the insert. The final clones were called pTN21 and pTN380, respectively (plasmid maps in Appendix B). Wild type and C260G mutant plasmids pSW21 and pSW380 were constructed by a similar method using primers SW012BAM 5'E and SW012BAM 3'E (primer sequences in Appendix A) and the parental plasmid pSW012 as a PCR template (Woodson 1992 and Pan and Woodson, 1998). Plasmid maps are shown in Appendix B.

The resulting pre-RNAs transcribed from these clones have the *Tetrahymena* group I intron flanked by either *E. coli* (Ec) rRNA or *Tetrahymena thermophila* (Tt) rRNA exons and transcribed by either T7 or *E. coli* (Eco) RNA polymerase. A diagram of these is shown in Fig 2.1. As can be seen from the plasmid maps in Appendix B, the 380 clones have extra vector sequences (~30 nt at 5' end and ~100 nt at 3' end) at either ends of the insert. In vitro splicing experiments showed that these vector sequences however, did not retard splicing of the pre-RNA (data not shown).

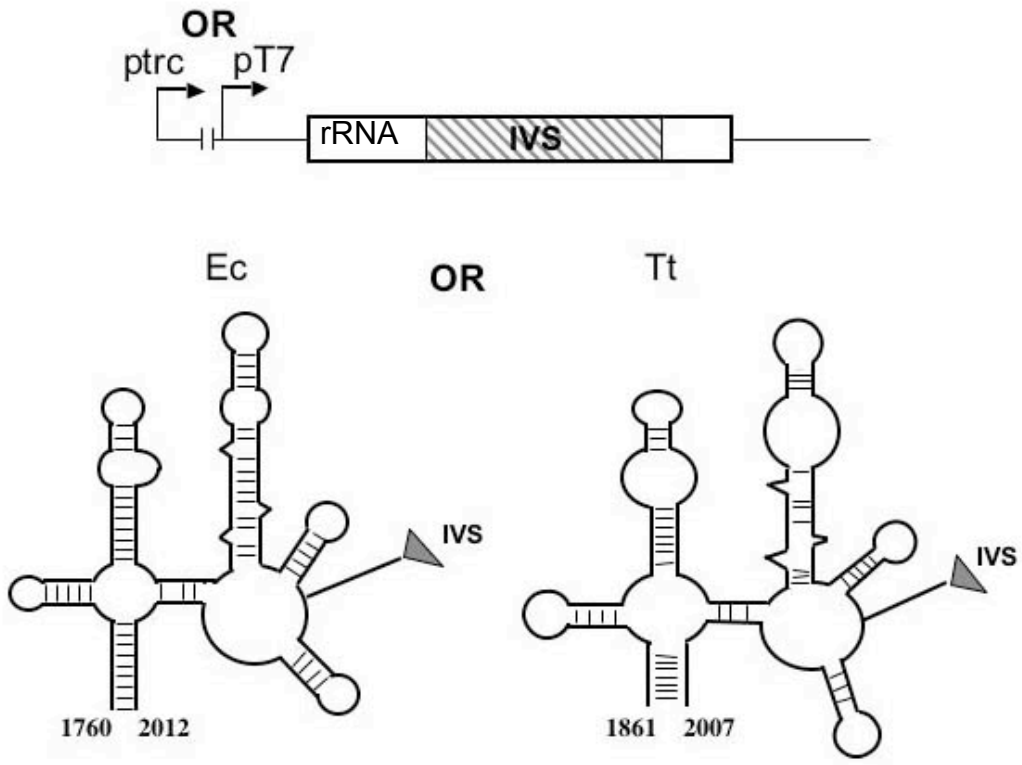


Figure 2.1: Constructs used in this study. Top panel shows the two promoters used in this study. Ptrc is the *E. coli* RNAP promoter and pT7 is the phage T7 RNAP promoter. The bottom panel shows the secondary structures of the exon sequences and the location of the intron (IVS) insertion site. The numbers indicate the starting and ending nucleotide for each rRNA exon. Numbering is according to *E. coli* rRNA residues (Noller 1981).

Primer Extension Assay

Primer extension assays were conducted as described in Zhang et al (1995a). 2 μM 5' ^{32}P-labeled primer IP006 (sequence in Appendix A), that is complementary to the 5' end of the intron, was annealed to intron present in 250 ng of total RNA. The primer was extended by AMV reverse transcriptase (USB) and an equimolar mixture of, dATP, dGTP and dTTP(Zhang et al 1995a). ddCTP was used to stop primer extension. 0.5 μg of total RNA was used in primer extension assays conducted on mutant pre-RNA. The resulting extension products were run on a denaturing 20% polyacrylamide gel. The amounts of radioactivity in the intron and precursor bands were quantitated by storage phosphorescence (Molecular Dynamics).

Steady state analysis

E. coli strain BL21 was grown in 100 ml of LB containing 100 mg/ml of ampicillin at 37˚C until cell density as measured by absorbance at A_{600} was between 0.2 and 0.4. 0.05 mM IPTG (final concentration) was added to induce transcription. For experiments at other temperatures, the cells were shifted to 25˚C, 30˚C or 42˚C at the time of induction. 20 ml aliquots of liquid culture were taken at various times after induction and 66.6 μl of chloramphenicol (30 mg/ml) was added to stabilize the ribosomes. Growth was stopped by freezing cells in ice-ethanol slurry for 1 min. Total RNA was extracted from the cells using the RNA extraction kit from Eppendorf-3 Prime Inc. containing guanidinium isothiocyanate. The ratio of intron and precursor RNA was calculated by primer extension assays as described above.

Intron and precursor decay rates

Decay rates were measured by incubating cells at 25˚C, 30˚C, 37˚C or 42˚C for 90 to 120 minutes after induction of transcription with IPTG. Steady-state expression levels were approached in this time. Transcription was stopped by the addition of 0.8 mg/ml rifampicin. Total RNA was extracted from 20 ml aliquots of liquid culture at various times after stopping transcription as described above. The fraction of free IVS RNA remaining was determined by primer extension and plotted versus time. This was fit to a first order rate equation from which the decay rate was calculated. Precursor decay rates were calculated in a similar manner, except cells were harvested at shorter intervals (1-10 and 30 min).

In vivo splicing rates

Splicing in *E. coli* was measured as described in Zhang et al (1995a). The splicing rate in vivo was taken to be equal to the product of the decay rate of the intron and the ratio of free intron to precursor at steady-state (Brehm and Cech 1983), as per the following scheme:

$$\text{pre} \xrightarrow{k_{splice}} \text{ivs} \xrightarrow{k_{decay}}$$

At steady state

$$k_{splice} = k_{decay} \bullet \frac{[ivs]}{[pre]}$$

RESULTS

In vivo splicing rates of pre-RNA at 37˚C

In order to compare the effects of polymerase and exons on the folding of the intron in *E. coli*, in vivo splicing rates were calculated by measuring steady state ratios of precursor to free intron and decay rates of free intron in cells transformed with the pre-RNA expression vectors shown in Fig 2.1 (Zhang et al 1995a). Induction with IPTG resulted in good expression of plasmid-encoded pre-RNA. As shown in Fig. 2.2, a large amount of spliced intron is visible after 30 min. The levels of precursor and intron RNA reached a plateau after 90 min, which was taken to be steady-state. To measure decay rate of the intron, 0.8 mg/ml of rifampicin was added after 90 min of induction with IPTG (after steady-state) to stop transcription. The amount of free intron at various times was visualized by autoradiographs of primer extension assays. The decay rate was calculated from the plot of fraction of intron against time as shown in Fig 2.3.

Splicing rates are summarized in Table 2.1. At 37˚C, the splicing rate of *E. coli* RNAP transcripts was two to three-fold faster than the T7 transcripts. The pre-RNA with *E. coli* rRNA exons spliced 3-4 fold faster than the pre-RNA with *Tetrahymena* rRNA exons (Table 2.1).

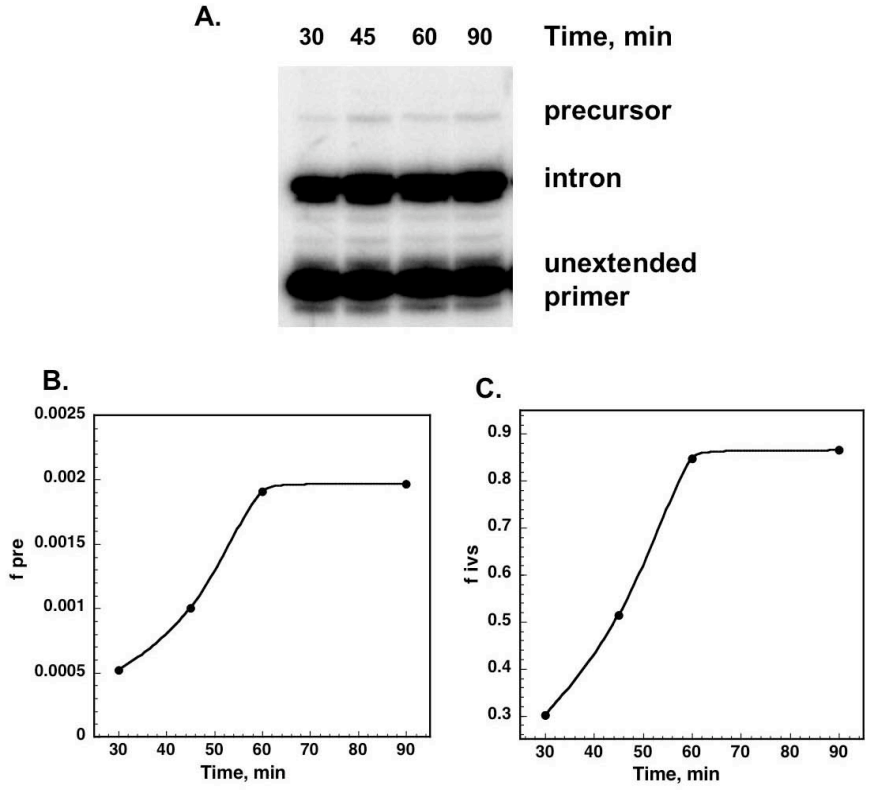


Figure 2.2: Steady-state levels of intron and precursor. A. Primer extension of total RNA from *E. coli* strain BL21 transformed with pTN380 and grown at 42˚C. Times after induction with 0.05 mM IPTG are indicated above each lane. B. Ratio of cpm of precursor band over cpm of unextended primer band (fpre) per lane was plotted versus time to get the steady-state expression levels of precursor. C. Ratio of cpm of intron band over cpm of unextended primer band (fivs) per lane versus time was plotted to get steady-state expression levels of intron. These were used to calculate steady-state ratio of intron to precursor as described in Zhang et al 1995a.

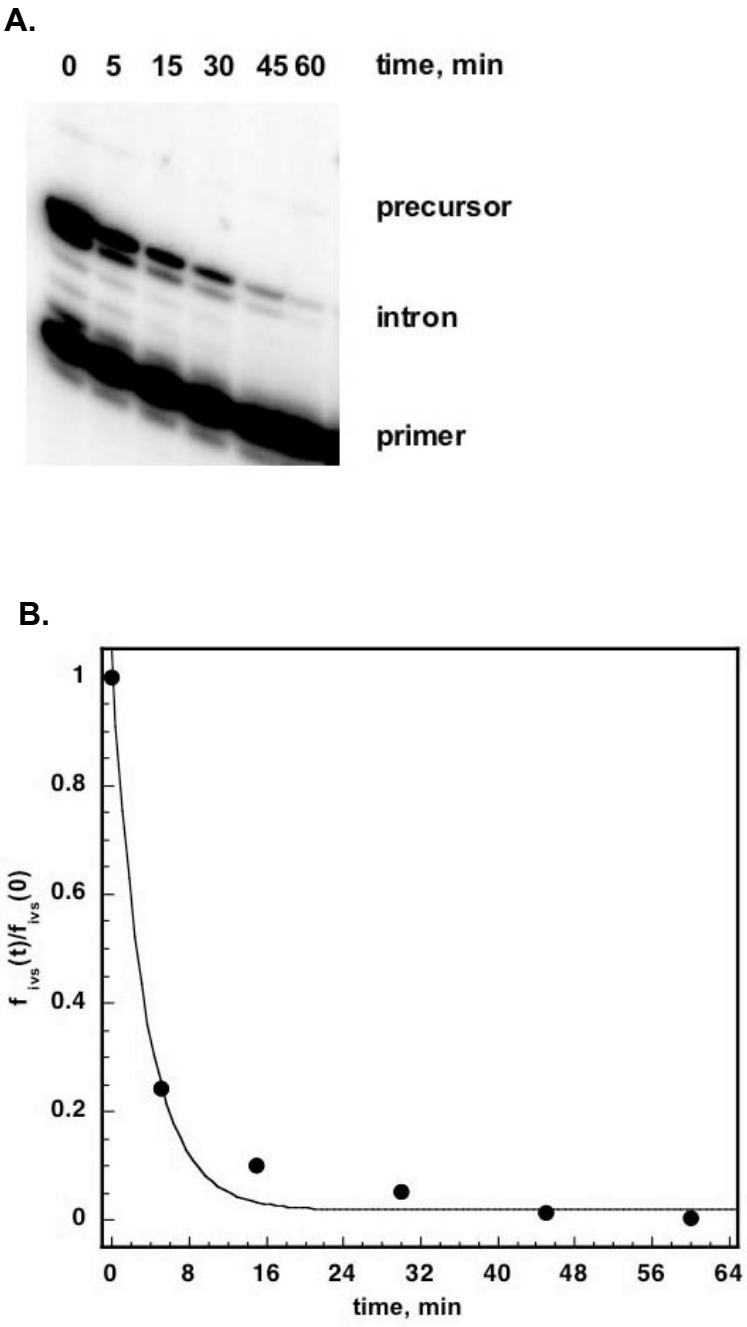


Figure 2.3: Intron decay rate of Ec-Eco transcript at 42˚C. A. Primer extension gel showing the disappearance of the spliced intron band over time after addition of rifampicin (0.8 mg/ml). B. The fraction of free intron present at a given time was plotted versus time and fit to:
$Y = A*(\exp^{-kt})+C$; where k = rate of decay, A = amplitude and C = constant.
The rates from similar graphs at different temperatures are shown in Table 2.2. The product of these values and the steady-state ratios of intron to precursor were used to calculate splicing rate in *E. coli*.

Table 2.1. Splicing rates (min^{-1}) of the *Tetrahymena* pre-RNA in *E. coli*.

Pre-RNA	Splicing rate, min^{-1}				E_a kcal•mol^{-1}•K^{-1}
	42˚C	37˚C	30˚C	25˚C	
Wild Type core					
Ec-T7	32.8 ± 9	12.0 ± 0.6	10.0 ± 0.1	5.0 ± 0.7	18
Tt-T7	5.0 ± 0.5	3.0 ± 0.5	0.50 ± 0.06	0.3 ± 0.0	31
Ec-Eco	38.5 ± 12	23.0 ± 0.8	19.0 ± 1.5	7.0 ± 0.7	16
Tt-Eco	5.0 ± 0.4	9.0 ± 1.3	1.0 ± 0.5	0.30 ± 0.04	ND[a]
C260G core					
Ec-T7	24 ± 1	8.0 ± 0.4	2.0 ± 0.1	0.50 ± 0.04	34
Tt-T7	1.0 ± 0.0	0.40 ± 0.03	0.05 ± 0.01	0.01 ± 0.00	40
Ec-Eco	19 ± 1	9.0 ± 0.6	3.0 ± 0.3	0.50 ± 0.04	20
Tt-Eco	0.10 ± 0.01	0.05 ± 0.01	0.02 ± 0.00	0.02 ± 0.00	14

In vivo splicing rates were calculated from primer extension assays using intron decay rates shown in Table 2.2. Variations are standard errors of the mean of 3 trials for the pre-RNA with wild type intron core and 4 trials for pre-RNA with mutant intron core. These values agree well (within error) with previously reported data for splicing in *E. coli* (Nikolcheva and Woodson 1999).

[a]ND, Not Done, Activation energy for Tt-Eco transcript was not calculated as splicing rates at 42˚C is skewed due to the higher decay rate of the pre-RNA.

Table 2.2: Decay rates (min^{-1}) in *E. coli*.

Pre-RNA	[a]Precursor decay rate	[b]intron decay rate			
		25˚C	30˚C	37˚C	42˚C
Wild Type core					
Ec-Eco	0.24	0.03	0.05	0.08	0.1[c]
Tt-Eco	0.64	0.02	0.05	0.18	0.3
C260G core					
Ec-Eco	0.83	0.07	0.1	0.24	0.4
Tt-Eco	1.30	0.04	0.1	0.3	0.5

Decay rates were calculated from primer extension analysis conducted on 0.25-0.50 µg of total RNA.
[a] precursor decay rates were measured at 37˚C.
[b] intron decay rates were measured at different temperatures. For wild type intron decay values variations between 2-3 trials was 10-25%. Rates at lower temperatures for both wild type and mutant intron agree well with those published earlier (Nikolcheva and Woodson 1999).
[c] decay rate taken from Zhang et al 1995a

Although these values were reproducible, the effect of polymerase or exons on in vivo splicing at 37˚C is subtle. To see if these differences vary under different cell growth conditions, I measured the splicing rate of this intron at different temperatures.

Temperature-dependence of splicing of wild type and Ec pre-RNA

In vitro folding experiments conducted on the Tt pre-RNA showed that the RNA becomes trapped in misfolded intermediates (Pan and Woodson 1998 and Rook et al 1998). These can be rescued by heat renaturation of the RNA or by the addition of chemical denaturants (Pan and Woodson 1998 and Pan and Sosnick 1997). More significantly, in vivo splicing was found to be cold sensitive. In vivo splicing experiments conducted on wild type and mutant RNA showed that splicing of a mutant intron was more cold sensitive than the wild type intron (Nikolcheva and Woodson 1999). This suggested that reducing the temperature drove more of the mutant RNA into misfolded intermediates, which were rescued at higher temperatures. As in vitro study of temperature dependence of splicing directly measures the change in partitioning between native and misfolded fractions. To determine if T7 transcripts misfold more than *E. coli* transcripts, I studied the effects of these parameters on splicing at lower temperatures.

Splicing rates were measured at 25˚C, 30˚C and 42˚C. The data are summarized in Table 2.1 and a representative primer extension assay is shown in Fig 2.4. The natural logarithm of the k_{obs} values were plotted versus the inverse of temperature to get an Arrhenius plot (Fig 2.5). The activation energies were calculated from this plot and found to be in the same range for both transcripts of the Ec pre-RNA (Table 2.1). Thus

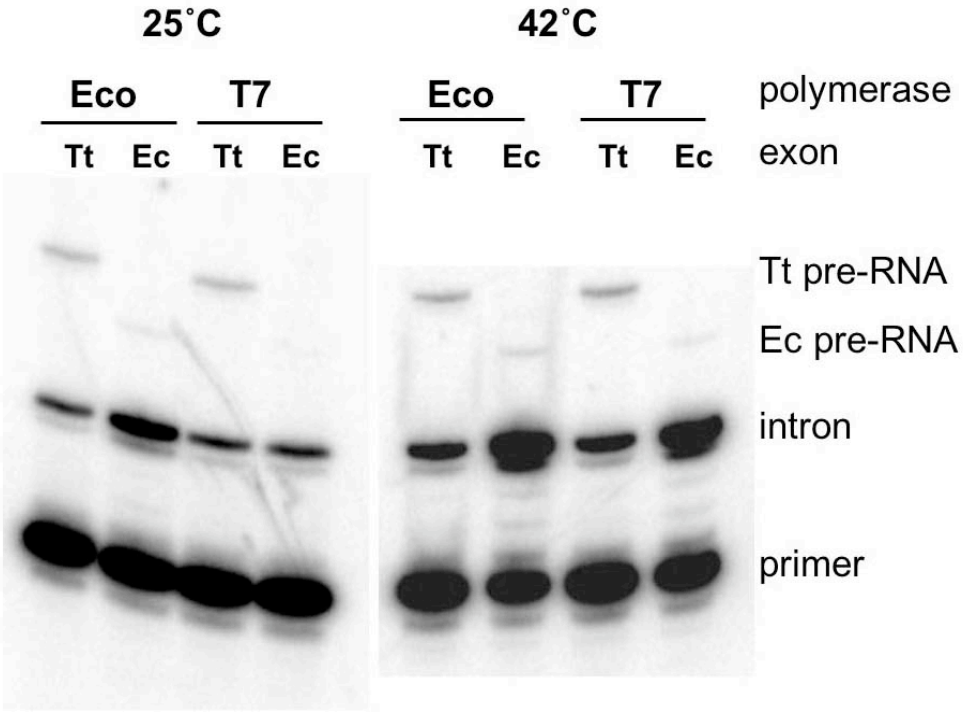


Figure 2.4: Splicing of *Tetrahymena* intron in *E. coli*. 0.25-0.5 g of total RNA from cells grown with IPTG at 25 or 42˚C was used in primer extension assays using primer complementary to the 5' end of intron. Spliced intron, unspliced precursor and unextended primer bands were separated on a 20% denaturing gel. Expressions of precursor and intron RNA is at steady state as explained in Materials and Methods (Fig 2.2). Tt: pre-RNA with *Tetrahymena* exons and wild type intron core. Ec: pre-RNA with *E. coli* exons and wild type intron core. T7, phage T7 RNAP; Eco, *E. coli* RNAP.

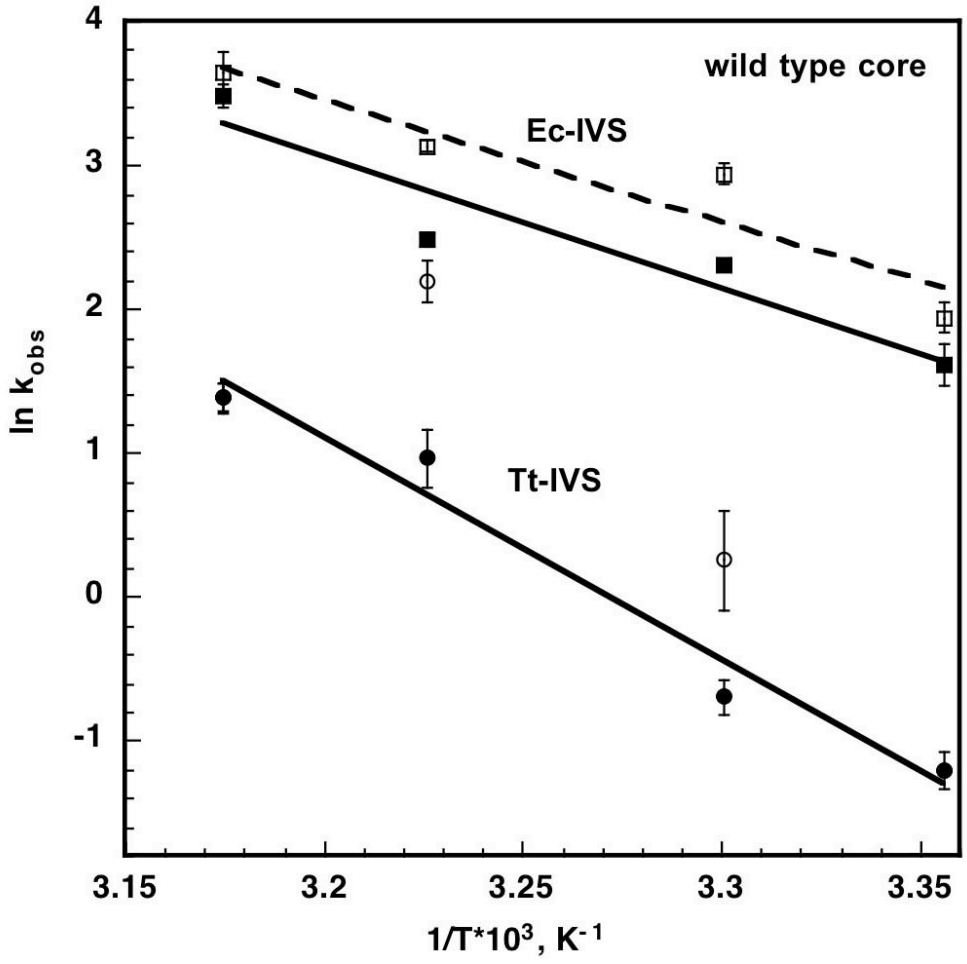


Figure 2.5: Temperature dependence of in vivo splicing of pre-RNA with wild type intron core. Observed rate constants (min^{-1}) as measured in Fig 2.4 are plotted versus inverse of temperature (K^{-1}). Open squares and dashed lines are *E. coli* RNAP transcripts of Ec-IVS. Open circles represent *E. coli* RNAP transcripts of Tt-IVS pre-RNA. Filled symbols and solid lines represent T7 RNAP transcripts. Ec-IVS: pre-RNA with *E. coli* exons and Tt-IVS: pre-RNA with *Tetrahymena* exons. Error bars indicate variations from the average of 3 trials. The data from Tt-Eco pre-RNA are not fit to the Arrhenius equation as splicing rates at high temperatures are hard to measure due to fast decay of the precursor.

while the *E. coli* transcripts are slightly more active at 30˚ and 37˚C, there appears to be little difference in energy barrier for splicing.

The activation energy for the T7 transcript of the Tt pre-RNA was however found to be about 2-fold higher than the T7 transcript of the Ec pre-RNA. Activation energy for the Tt-Eco pre-RNA could not be calculated with accuracy as the splicing rates at high temperatures were skewed due to an increase in the degradation of the precursor RNA (Table 2.2). However, splicing of this transcript would also have a high activation energy, if one extrapolates from splicing rate at 37˚C. Apart from the difference in the activation energies for the Ec and Tt pre-RNA, there is a considerable difference in the Y- intercepts. This is a reflection of the difference in splicing rates between these two pre-RNAs. As described in the previous section, the difference in splicing rate of these two pre-RNAs is about 3-4 fold at 37˚C, with the Ec pre-RNA splicing better than the Tt pre-RNA. This difference is, however, increased about 20-fold at lower temperatures (Table 2.1) suggesting that splicing is affected by the stability of the conformation of exon sequences. If this were true, then these effects would be enhanced in the splicing of an intron mutant that destabilizes the folded structure. To test this I studied the effect of exons and polymerase on the splicing of a folding mutant at various temperatures.

Temperature-dependence of splicing of mutant pre-RNA

The point mutation C260G disrupts the triple helix at the junction of the P4 and P6 domain at the core of the intron causing it to misfold (Zarrinkar and Williamson 1996 and Pan and Woodson 1998). In *E. coli*, splicing of this mutant intron was found to be

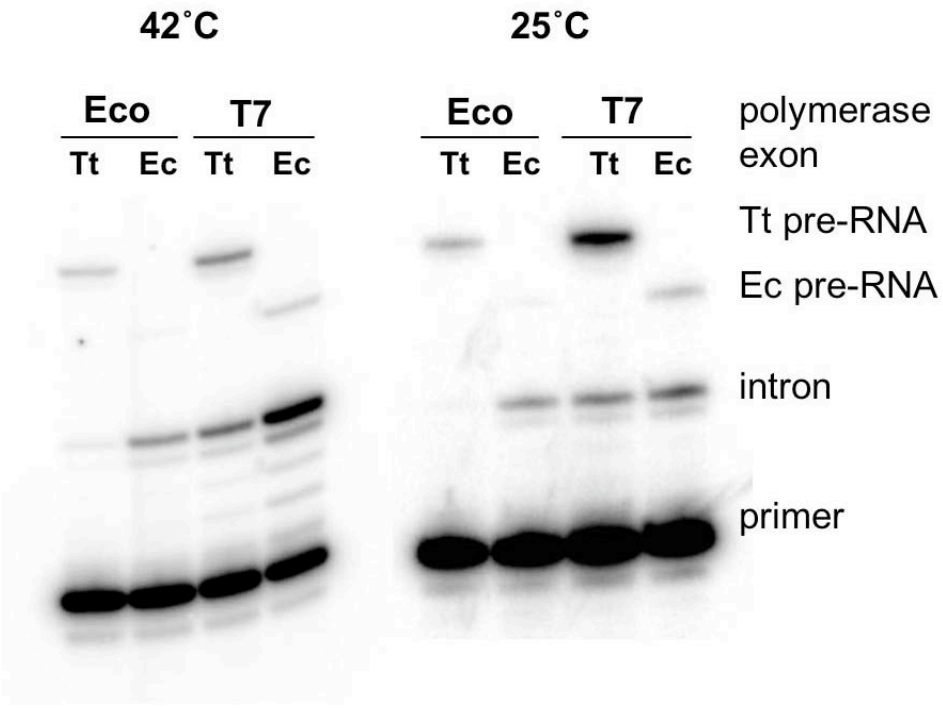


Figure 2.6: Splicing of pre-RNAs with mutant intron core in *E. coli*: 0.5□g of total RNA from cells grown at 25 or 42˚C is used in primer extension assays using primer complementary to the 5' end of intron. Spliced intron, unspliced precursor and unextended primer bands are separated on a 20% denaturing gel. The figure depicts steady-state expression levels of precursor and intron RNA as explained in Materials and Methods. Labels as in Fig 2.4.

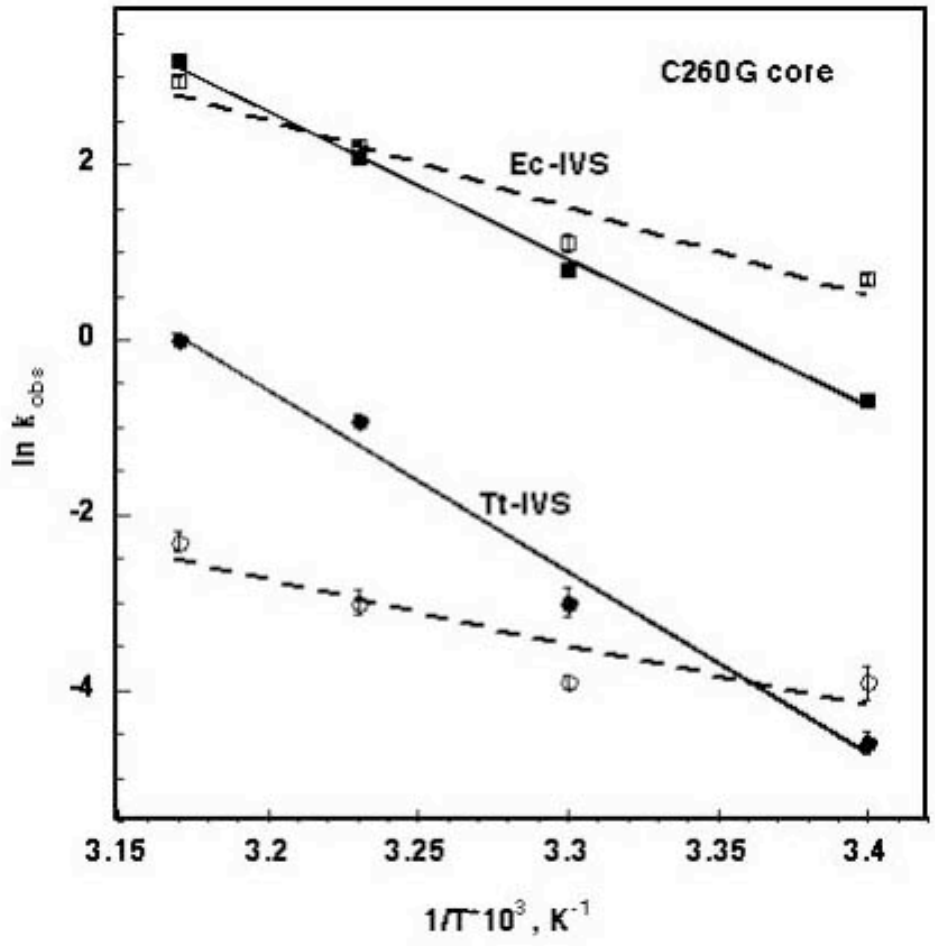


Figure 2.7: Temperature-dependence of in vivo splicing of pre-RNA with mutant intron core. Observed rate constants (min^{-1}) as measured in Fig 2.6 are plotted versus inverse of temperature (K^{-1}). Open symbols and dashed lines are *E. coli* RNAP transcripts. Filled symbols and solid lines represent T7 RNAP transcripts. Ec-IVS: pre-RNA with *E. coli* exons and Tt-IVS: pre-RNA with *Tetrahymena* exons. Error bars indicate variations from the average of 4 trials.

more cold-sensitive in the absence of rRNA exon sequences than in the presence of minimal rRNA exons (Nikolcheva and Woodson 1999). I used this well characterized folding mutant to study the effect of polymerase and exons on the folding of the *Tetrahymena* intron. Splicing rates at various temperatures were measured as described above and activation energies were calculated for each transcript (Table 2.1, Fig 2.6 and Fig 2.7). From Fig 2.7, it can be seen that T7 transcripts of the mutant pre-RNA have a steeper slope (higher activation energy) than the analogous *E. coli* transcripts. This suggests that T7 transcripts are trapped in more misfolded structures than *E. coli* RNAP transcripts. It is also clear from Fig 2.7, that as in the case for the pre-RNA with a wild type intron core, both Tt-IVS pre-RNAs have lower splicing rate overall (Y intercept lower) than their Ec-IVS counterparts. This is more obvious for the *E. coli* transcript of the Tt-IVS pre-RNA. This lower apparent splicing activity could be due to unstable exon structure that begins to unfold at high temperatures. If this were so, the unstructured pre-RNA would be degraded and we should see very little pre-RNA at steady-state times. This is in fact the case as one can see from Fig 2.6, where the band corresponding to the Tt-T7 precursor is more intense than the one for Tt-Eco pre-RNA at 42˚C. This suggests that the Ec pre-RNA forms more stable structures than the Tt pre-RNA.

Precursor decay in *E. coli* at 37˚C

Results described in the previous section suggest that the *E. coli* transcripts containing Ec exons are more stable than those containing Tt exons. For the Tt pre-RNA, the rapid synthesis of the entire pre-RNA (transcription by T7 takes approximately 2 sec) helps in the formation of stable active structures. In order to see if this stability translates into a

lower precursor decay rate for the Ec-Eco pre-RNA, I measured the half-life of *E. coli* RNAP transcripts at 37˚C. The reason for this is as follows: In the cell, the pre-RNA either splices or is degraded. Thus, the disappearance of the pre-RNA is a sum of its splicing and degradation rates. Since I know the splicing rate of the pre-RNA, the intrinsic decay rate can be determined from the difference between the overall rate of disappearance and the rate of splicing (Donahue and Fedor 1997 and Donahue et al 2000).

To measure the disappearance rate of the precursor, I grew liquid cultures at 37˚C in the presence of IPTG until steady state levels of expression is reached (60 min) before adding rifampicin to inhibit further transcription initiation. Cells were harvested 1-30 min after addition of rifampicin, and the amount of pre-RNA was measured by primer extension as described above.

From the data (Table 2.2; page 38), it is clear that the disappearance of all the pre-RNAs is slower than their splicing rates. Normally one would expect the apparent decay rate to be greater than the rate of splicing. This discrepancy can be reconciled by assuming that the measured decay rates represent the decay of only the misfolded population of pre-RNA. This is not an implausible assumption as, the Tt pre-RNA has been shown to have a biphasic splicing rate in vitro (Emerick et al 1996) and this biphasic nature is directly related to the partitioning of the RNA into fast folding and slow folding populations (Emerick et al 1996, Thirumalai and Woodson 1994).

My preliminary results show that for pre-RNAs with a wild type intron core, the Ec pre-RNA decays about two-fold slower than the Tt pre-RNA (Table 2.2; page 38). For the pre-RNAs with a mutant intron core however, the decay rate of both the Ec and

Tt pre-RNAs are similar, but faster than the decay of the wild type transcripts. This is consistent with my assumption that the half-life of the pre-RNA primarily represents the decay of misfolded and inactive transcripts, as RNAs carrying C260G mutation are expected to misfold more than transcripts with a wild type core. These results suggest that at least for the *E. coli* transcripts, *E. coli* exons form more stable folded structures than those formed by the *Tetrahymena* exons

DISCUSSION

As explained in the introduction, the cause of in vivo splicing enhancement of the *Tetrahymena* group I intron is not known. Presumably most of the RNA folds co-transcriptionally in vivo. Moreover, in vivo splicing of this intron was facilitated by the presence of rRNA exons flanking the ribozyme (Nikolcheva and Woodson 1999). In vitro, these rRNA exons have been shown to form either active or inactive structures due to alternative base-pairing interactions 5' of the splice site (Emerick and Woodson 1996, Pan and Woodson 1998 and Cao and Woodson 1998). In this study I wanted to see if perturbing transcriptional folding by changing elongation rate and stability of rRNA exons affects the balance between active and inactive structures, which can be studied for this pre-RNA by measuring splicing rates in vivo. From my results it is clear that these factors play a significant role in splicing. The effects of these two factors are discussed separately below.

Effect of polymerase on in vivo splicing

Sequential folding of RNA during transcription was found to enhance folding of the *Tetrahymena* ribozyme (Heilman-Miller and Woodson 2003). Although this enhancement was not affected by the elongation rate of the polymerase, changing the order in which sequences were transcribed was important. For certain permuted RNAs sequential folding was enhanced by more than 10-fold compared with refolding under the same conditions (Heilman-Miller and Woodson 2003). This suggests that perturbing sequential folding during transcription will change the folding pathway of the intron.

My results show that in vivo splicing of pre-RNA with a wild type intron core is affected by changing the kind of polymerase used for transcription. *E. coli* transcripts have two to three-fold faster splicing rates than their T7 counterparts (Table 2.1) at 37˚C. It is interesting to note that this is the same extent of folding enhancement seen for the ribozyme in vitro, when co-transcriptional folding and refolding were compared. However, in vitro there was no difference between *E. coli* and T7 polymerase (Heilman-Miller and Woodson, 2003).

Although the temperature-dependence of splicing of the Ec pre-RNAs with a wild type intron core did not show any polymerase specific effects, transcription by *E. coli* RNAP improves splicing of C260G pre-RNAs at lower temperatures. Studying splicing at different temperatures helps us tease out interesting details of folding of the RNA. If two RNAs have a similar temperature dependence of splicing but one RNA is more active than the other at all temperatures, it suggests that the rate-limiting step involves a smaller activation entropy. This change in entropy could be due to the folding parameters or to different partitioning between native and misfolded structures. By

contrast, the simplest explanation for a higher activation energy, (steeper slope) is greater enthalpic barriers to refolding of the RNA. Thus, the shallow slopes in the Arrhenius plots for the *E. coli* RNAP transcripts of mutant RNA suggests that these pre-RNAs are not trapped in as many misfolds as their T7 counterparts for which the slope is steep (Fig 2.7). This folding facilitation by *E. coli* RNA polymerase, however, is template dependent, as it is more enhanced in the Ec-IVS template than in the Tt-IVS template. This is discussed further in the following section.

Effect of exons on in vivo splicing of the *Tetrahymena* intron

Comparing splicing of wild type pre-RNA at 37˚C, we find that the pre-RNA with *E. coli* exons splice approximately two-fold faster than pre-RNA with *Tetrahymena* exons. This splicing enhancement increases to about twenty-fold at lower temperatures (25˚C and 30˚C) (Table 2.1). This shows that *E. coli* exons facilitate folding of the intron at lower temperatures by relieving misfolds. This is probably by tilting the balance towards the formation of native interactions over non-native ones. This is because at lower temperatures the energy necessary for rearrangement of non-native interactions increases. This difference between the two exons is enhanced in the C260G mutant intron background. Extrapolating the splicing rate from 37˚C (Fig 2.5), we expect the *E. coli* transcript of Tt-IVS to have the same activity as Ec-Eco pre-RNA. However, activity of both Tt-IVS transcripts is much lower than Ec-IVS at 42˚C. This difference is more extreme in the C260G mutant (Fig 2.7). Fig 2.7 shows reduced activity at all temperatures for this RNA (Table 2.1). One possible reason for this apparently low activity is that the misfolded Tt pre-RNA is unstable and is rapidly degraded. This

conclusion is corroborated by Fig 2.6, which shows a lower amount of unspliced Tt-Eco pre-RNA with no corresponding increase in free intron. This indicates that the mutant Tt-Eco RNA is degraded more rapidly than other pre-RNAs.

Studies on mRNA decay in *E. coli* have shown that stability of the RNA is greatly increased by the presence of stable secondary structures at either end of the transcript while unstructured RNA is rapidly degraded (Higgins et al 1993, Bechhofer 1993 and Regnier and Arraiano 2000). Thus the rapid decay of the Tt-Eco transcript suggests lower stability of this pre-RNA. This is further strengthened by the precursor decay rate measurements shown in Table 2.2 where this pre-RNA has the highest decay rate.

In conclusion, *E. coli* RNAP facilitates folding of RNA in vivo, but this is more pronounced on the Ec pre-RNA template. The Tt pre-RNA is less stable, perhaps because it is more structurally dynamic, and this effect is more pronounced when it is transcribed by *E. coli* RNAP. Thus, the facilitation of folding shown by *E. coli* RNAP is template-dependent, in so far as the polymerase choice has an opposite effect on the two pre-RNAs. Template-dependence must be due to a mechanism that involves interaction of the polymerase and template DNA. One such mechanism is site-specific pausing of polymerases. It is well established that *E. coli* RNAP pauses at specific places along the template and this pausing involves interactions with the nascent RNA strand apart from the DNA template (von Hippel et al 1995, Artismovitch and Landick 2000 and Pan et al 1998b). Thus it is conceivable that the polymerase could pause at places that not only prevent the formation of non-native interactions but also promote the formation of local native interactions.

To see if site-specific pausing enhances splicing of the intron and to compare my in vivo data with the effect of polymerases and exons on in vitro folding, I have mapped the pause sites of *E. coli* RNAP and measured self-splicing rates in vitro. These results are discussed in the next chapter

..

Chapter 3

EFFECT OF TRANSCRIPTIONAL PARAMETERS ON IN VITRO FOLDING OF THE *TETRAHYMENA* GROUP I INTRON.

INTRODUCTION

As explained in Chapter I, results from co-transcriptional folding on circularly permuted variants of the *Tetrahymena* ribozyme indicated that the sequential folding of RNA during transcription is sensitive to the order of RNA sequences which is altered by circular permutation (Heilman-Miller and Woodson 2003a). For the *Tetrahymena* pre-RNA this could affect the balance between inactive and active structures, by changing base-pairing interactions in the 5' exon and between the 5' exon and the intron.

These alternative helices, described in detail in chapter I, include the active PX and inactive P(-1) in the 5' exon and the inactive P(-1) and the active P1 in the 5'exon and intron. Formation of the inactive P(-1) helix is shown to correlate with the formation of an alternate helix in the intron core, the alt P3 (Pan and Woodson 1998). Together these alternative base pairs lead to a misfolded RNA intermediate. Thus by manipulating sequential folding conditions, by using polymerases of different elongation rates or using different exons, we can perturb folding of the pre-RNA in vitro which can be measured as a change in self-splicing activity.

In this chapter, I study the effect of polymerases and exons on in vitro self-splicing to see if *E. coli* RNAP shows the same kind of template-dependent folding

enhancement as seen in *E. coli*. I also test to see if site-specific pausing of *E. coli* RNAP could be the cause for this folding facilitation.

MATERIALS AND METHODS

Plasmids

The plasmids pTN21 and pSW21 described in chapter 2 were used to prepare T7 transcripts in this study. Templates for T7 RNAP were prepared by digesting pTN21 and pSW21 with *EcoR* I. This gives transcripts of 710 nts and 695 nts respectively. These transcripts have 3 extra vector nucleotides at the 3' end of the corresponding *E. coli* transcript. Templates for *E. coli* RNAP were prepared by PCR amplification of the T7 templates using an upstream primer containing the trc promoter (primer name E-21, sequence in Appendix A) and the downstream primers TN112BAM 3'E and SW012BAM 3'E (Appendix A). The resulting templates were called TN E-21 and SW E-21. All DNA templates were purified by phenol chloroform extraction and ethanol precipitation.

In vitro transcription

To prepare T7 transcripts, 40 μl transcription reactions containing 1μg linear template DNA were incubated at 30˚C for 30 minutes. The transcripts were uniformly labeled using 10 μCi of α [^{32}P] ATP. The pre-RNA was purified by size exclusion spin column chromatography (TE-100, Clontech) (Emerick and Woodson 1993).

To prepare *E. coli* transcripts, *E. coli* RNAP holoenzyme (Epicenter Technologies) and 2 nM of template DNA (final concentration) (5:1 molar ratio) were incubated at 37˚C for 15 minutes, in 50 mM Tris-HCl (pH 8), 3 mM magnesium acetate, 0.1 mM EDTA, 0.1 mM DTT, 250 □g/□l BSA and 20 mM NaCl to enable the formation of open complexes (Igarashi and Ishihama, 1991). Open complexes were chased into elongation complexes by simultaneous addition of rNTPs (final concentration of 200 □M ATP, CTP, GTP, 50 □M UTP and 50 □Ci □ [^{32}P] UTP) and rifampicin (final concentration 10 □g/□l) and incubating at 37˚C for 5 min (Kainz and Gourse, 1998). The pre-RNA was purified by size exclusion spin column (TE-100, Clontech).

In vitro splicing

In vitro self-splicing reactions were conducted at 30˚C in splicing buffer (100 mM $(NH_4)_2SO_4$, 50 mM HEPES pH 7.5, 6 mM $MgCl_2$, 1 mM EDTA; Been and Cech 1990) and 0.2 mM GTP (Woodson, 1992). Splicing products of *E. coli* RNAP transcripts were precipitated with three volumes of ethanol, 40 mM NaCl and 30 mM EDTA (final concentrations). Samples were resuspended in 5 □l of 2X formamide loading dye (100% deionized formamide, 2% bromophenol blue, 2% xylene cyanol and 10X TBE). Reactions were heated to 100˚C for 3 min before loading onto a 4% denaturing polyacrylamide gel. The fraction of spliced products (f_{sp}) was determined as shown below:

$$f_{sp} = \frac{\text{cpm (IVS)}}{\text{cpm (total)}} \bullet \frac{\text{\# of nt (pre)}}{\text{\# of nt (IVS)}}$$

The data were fit to a double exponential equation as described in Emerick et al. (1996) and fit parameters shown in Table 3.1. For reannealing experiments, the pre-RNA were heated to 95˚C for 1 min in the absence of $MgCl_2$, then cooled to room temperature in the presence of 6 mM magnesium.

Urea renaturation experiments were conducted by heating the pre-RNA to 95˚C in splicing buffer in the presence of 3 M urea and absence of magnesium salt for 1 min An aliquot of this was added to a mixture of splicing buffer containing 6 mM $MgCl_2$ that was pre-incubated in 30˚C, such that the final concentration of urea was 300 mM. Control reactions contained pre-RNA incubating at 30˚C in splicing buffer with 6 mM $MgCl_2$ and 300 mM urea.

Nondenaturing gel electrophoresis

Equilibrium folding assays were conducted at 30˚C for 30 min with either no magnesium or with 3, 4, 5, 6, 7, 8, 9, 10, 12, 15, 20, 30 or 40 mM $MgCl_2$. The pre-RNA were either purified by spin column chromatography or heat denatured as described above before addition of magnesium. 2 □l aliquots were loaded on a non-denaturing 8% polyacrylamide gel that contained 10 mM $MgCl_2$. Once the RNA enters the matrix of the gel, it doesn't undergo appreciable change in its conformations, allowing separation of pre-RNA according its shape. The gel was run at 10-15W and approximately 10˚C for 6-

8 hrs so as to separate the intermediates and native bands. The intermediates bands appear as a smear on the gel, indicative of the presence of a number of non-native conformations (Emerick and Woodson 1994).

E. coli RNAP pausing assays

Single rounds of transcription elongation were conducted on SWE-21 and TNE-21 templates at 37˚C. Open complexes were formed as described above. Aliquots of elongation complexes at various times (up to 2 min) were quenched with 40 mM NaCl and 30 mM EDTA (Kainz and Gourse, 1998). The samples were precipitated with ethanol as described above. Samples were heated to 100˚C for 3 min before loading onto a 6% polyacrylamide sequencing gel.

To prepare molecular weight markers for each of the pre-RNA, standard PCR reactions were performed using a primer complementary to the T7 promoter and various downstream primers that anneal to different regions of the template (primer sequences in Appendix A). The resulting PCR products were purified by size exclusion spin column (GFX kit, Promega) and cleaned by phenol-chloroform extraction method as described above and used as T7 transcription templates to get RNA bands of varying length and corresponding to different regions on the pre-RNA. These were loaded alongside samples from *E. coli* RNAP pausing experiments. This ladder helped assign position of pause sites to within ± 5 nts for the exon positions and ± 10 nts for the intron positions as shown in Fig 3.4.

The pause-sites were mapped onto the secondary structure of the exon and intron sequence. The most stable structure of the nascent strand corresponding to each pause

site was analyzed by entering the sequences into the Mfold program (Zuker 1989 and Jaeger et al 1990 http://mfold1.wustl.edu/~mfold/rna/form1.cgi).

S30 extract assays

Uniformly labeled Ec-T7 pre-RNA transcribed as described as above was incubated with 1 l of commercially available *E. coli* S30 extract (Promega) and splicing buffer at 30˚C for ten minutes. The S30 extract from *E. coli* contains the bacterial cell lysate excluding the cell wall and membranes. Samples also contained 1 U RNase inhibitor (Promega). 0.2 mM GTP was added after 10 min to initiate splicing reaction. Aliquots from 0-60 min after addition of GTP were taken and samples were extracted with phenol and chloroform and precipitated with ethanol. Samples were analyzed on a 4% denaturing polyacrylamide gel. As a control for this experiment, the pre-RNA was incubated in the S30 extract storage buffer (10mM tris-HCl pH 7.5, 14 mM magnesium acetate, 60 mM potassium acetate and 1 mM DTT). Control experiments were done in the same manner as those in which the pre-RNA was treated with extract. The products were analyzed as described in the in vitro self-splicing section.

RESULTS

Effect of polymerase on in vitro splicing

To see if RNAP elongation rate plays a significant role in enhancing splicing rate in vitro, I compared in vitro self-splicing rates of pre-RNAs transcribed by a slow

elongating (*E. coli*) and a fast elongating (T7) RNAP. The transcripts were incubated in splicing buffer at 30˚C for up to 60 min after addition of 0.2 mM GTP. The fraction of spliced products was plotted versus time (as described in the previous section) and splicing rates were calculated from progress curves.

As can be seen in Fig 3.1, in vitro self-splicing has three kinetic phases. The amplitude of the fast phase (or the burst) represents the RNA population that is correctly folded at the start of the reaction. The amplitude of the slow phase represents the population of pre-RNA that becomes trapped in various misfolded structures but is capable of refolding to the active structure in the duration of the experiment (Emerick et al., 1996). The third phase represents the population of RNA that is incapable of attaining the active structure in the duration of the experiment (60-90 min). The polymerase used to synthesize the pre-RNA affects all these phases and this effect is template specific. Comparing *E. coli* RNAP transcripts, the *E. coli* (Ec) pre-RNA (open squares and dashed lines) has a large burst of 45%, and a high extent of splicing after 60 min (80%). On the other hand, the *Tetrahymena* (Tt) pre-RNA (open circle, dashed line), has poor splicing capacity, with only 10% of the RNA spliced in the burst phase and only 20% showing any splicing activity at all. Denaturing this RNA by heating it to 95˚C in the absence of magnesium salts and allowing it to refold in 6 mM $MgCl_2$, however, restores splicing activity of the pre-RNA to the same extent of the Ec-Eco species (Table 3.1). This suggests that the poor splicing activity is due to the nascent structures formed during transcription rather than some intrinsic defect of this pre-RNA.

Among the T7 transcripts, Tt-T7 pre-RNA (closed circles and continuous line) has same splicing capacity as the Ec-Eco pre-RNA (Fig 3.1), with 42% of the RNA

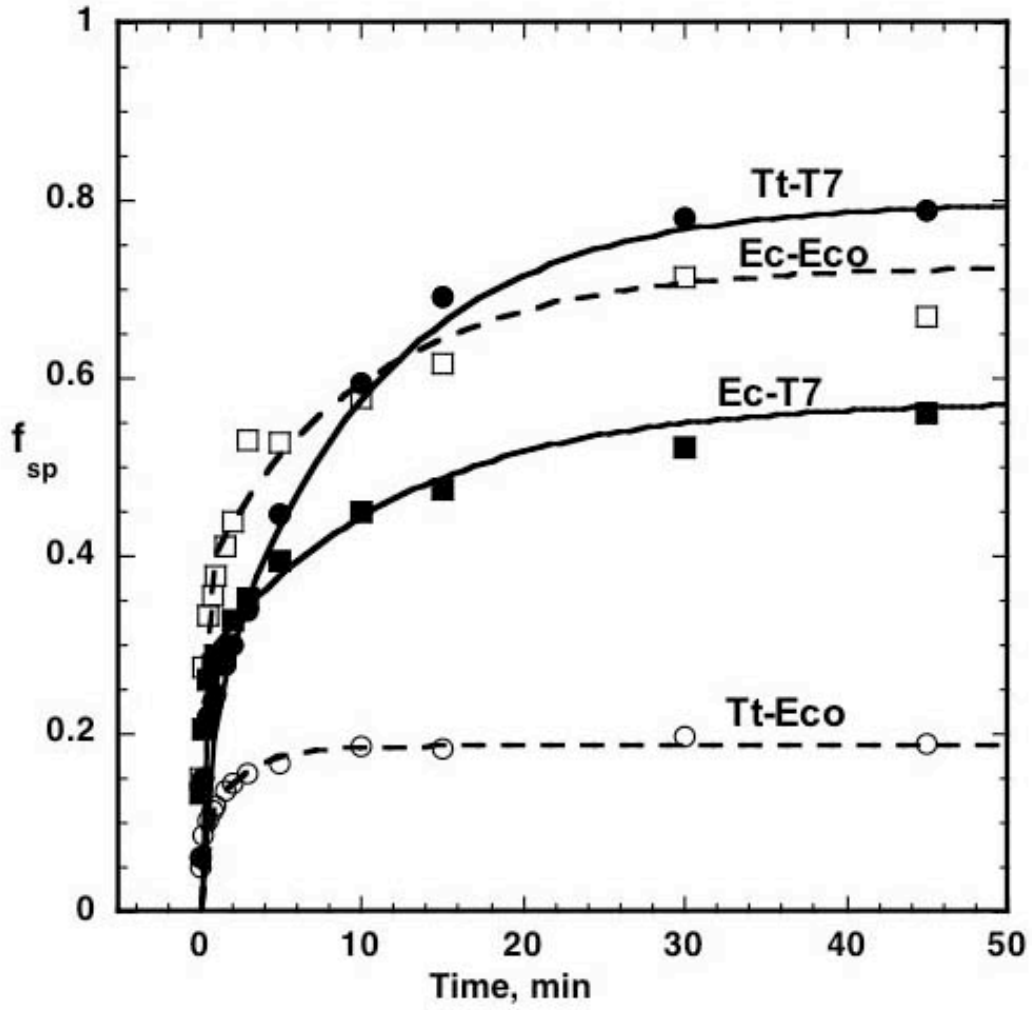


Figure 3.1: Effect of polymerase on non-denatured pre-RNA. Pre-RNA transcribed by either *E. coli* (open symbols) or T7 (closed symbols) RNAP were incubated in splicing buffer and 0.2 mM GTPat 30˚C. Aliquots at various times were run on a denaturing 4% acrylamide gel. Rates and fit parameters are in Table 3.1. Squares indicate Ec pre-RNA and circles Tt pre-RNA.

splicing in the burst phase and 80% of the RNA population splicing to completion in 45 minutes. Ec-T7 pre-RNA (closed squares, continuous line) however, shows lower activity, with less than 30% of the RNA splicing in the burst phase and only 55% of the total RNA population splicing to completion in 60 minutes.

These results show that *E. coli* RNAP enhances self-splicing of only the Ec pre-RNA while the T7 RNAP enhances splicing of the Tt pre-RNA. Thus as in *E. coli*, the effect of polymerases on in vitro folding of the *Tetrahymena* group I intron is template dependent.

Effect of renaturation on T7 transcripts

The temperature-dependence of splicing in vivo (Chapter 2) showed that the *Tetrahymena* rRNA exons form less stable structures than the homologous sequences from *E. coli*. To compare the stability of nascent structures formed by these two exon sequences during in vitro splicing, pre-RNAs were refolded in vitro before self-splicing. Magnesium renaturation drastically reduced the splicing activity of Ec-T7 pre-RNA (Fig 3.2, Table 3.1). The percentage of RNA in the burst phase was reduced by half of that seen for the non-renatured RNA and the total extent spliced after 60 min reduced about 2-fold (Table 3.1). This was unexpected as renaturation is expected to increase splicing activity allowing the RNA to refold into the correct structure.

There were two possible reasons for the decrease in splicing activity. One was that heating to 95˚C is not sufficient to unfold the RNA. The other was that sequential folding during transcription promotes the formation of some native interactions, which are lost during renaturation.

Table 3.1 In vitro self-splicing rates (min^{-1}) of the *Tetrahymena* intron.

Pre-RNA	Non-renatured		Mg^{2+} renatured		urea renatured	
	Rate	Amp.	Rate	Amp.	Rate	Amp.
Ec-T7	3 ± 0.4[a] 0.06 ± 0.01	30% (55%)[b]	N_d 0.02 ± 0.007	15% (30%)	3.6 ± 0.4 0.15 ± 0.02	60% (95%)
Tt-T7	2.4 ± 0.14 0.3 ± 0.1	40% (80%)	1 ± 0.14 0.06	40% (95%)	0.3	60% (95%)
Ec-Eco	2.3 ± 0.5 0.06 ± 0.02	30% (60%)	4.4 ± 1 0.03 ± 0.01	30% (70%)	Nd[c]	Nd
Tt-Eco	1.5 ± 0.7	(25%)	1.5 ± 0.7 0.2 ± 0.07	30% (70%)	Nd	Nd

In vitro self-splicing rates measured as described in Materials and Methods and rates calculated from plots of f_{sp} vs. time fit to the following equation:
$Y= A_1*(1-exp^{-k1t}) + A_2*(1-exp^{-k2t})$; where A_1 and A_2= amplitude of the burst phase and slow phase respectively. k_1 = rate constant of the burst phase and k_2 = rate constant of the slow phase.
[a] Rate constants of self-splicing reactions conducted at 30˚C. Values represent means of three to four trials. Errors shown are standard deviation from the mean.
[b]Amplitudes of self-splicing reactions. Individual fits give values within ± 5%. Values not in parenthesis indicate A_1 Vaule in parenthesis indicate extent of spliced RNA at the end of 60 min at 30˚C.
[c]Nd, Not determined.

To distinguish between these two possibilities, I renatured the RNA in presence of the denaturant urea. If the loss of splicing activity is due to stable non-native structures, then treating the RNA with urea should increase the splicing activity. If treating with urea caused no difference to the splicing rate, then the loss of splicing activity during renaturation is due to loss of native interactions formed during transcription. This is because urea lowers the stability of both native and non-native interactions.

The pre-RNA was heated to 95˚C in the presence of 3 M urea for 1 min in the absence of magnesium salts. The denatured RNA was added to buffer containing 6 mM $MgCl_2$ at 30˚C and splicing was initiated by addition of GTP (final concentration of urea was 0.3 mM). This restored the splicing activity to its maximum extent (filled triangles in Fig 3.2). This suggested that the nascent structures formed by the Ec-IVS RNA when transcribed by T7 RNA polymerase are very stable and require a strong denaturant like urea to unfold. This is not simply due to urea in the splicing reaction, because addition of 0.3 mM urea without heat renaturation had little effect on the splicing activity (Fig 3.2). However, we cannot rule out the possibility that 0.3 mM urea improves refolding while the RNA is cooled from 95˚C to room temperature.

If the non-native interactions formed by the pre-RNA are indeed stable, we should be able to see them on a non-denaturing gel. To test this I ran the Ec-T7 pre-RNA in a non-denaturing gel. Once the RNA enters the matrix of a non-denaturing gel it does not undergo any appreciable structural reformations (Emerick and Woodson, 1994). As seen in Fig. 3.3, a large proportion of the RNA is trapped in non-native intermediate structures, which do not completely disappear even at high magnesium concentrations.

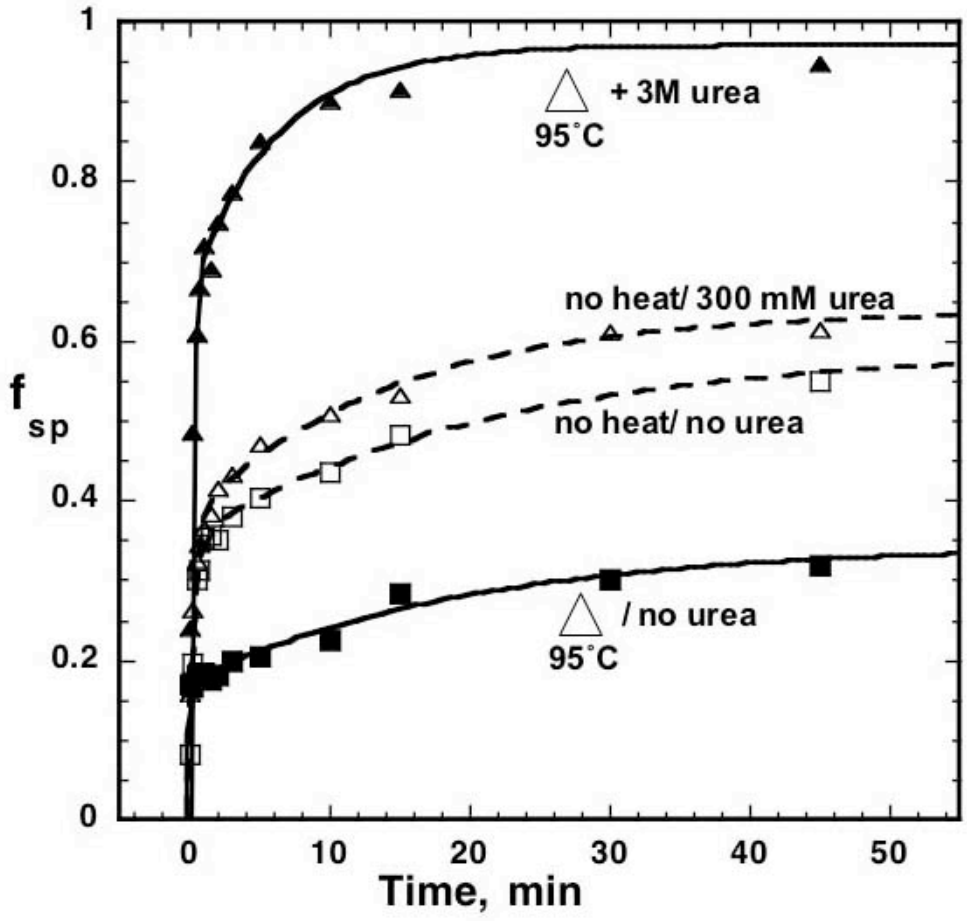


Figure 3.2: Effect of urea on in vitro self-splicing of Ec-T7 pre-RNA at 30˚C. Pre-RNA were heated to 95˚C for 1 min in the presence (filled triangles) or absence (filled squares) of 3 M urea before addition of 1mM GTP to initiate splicing reactions (Materials and Methods). The RNA was cooled to 30˚C in 6 mM $MgCl_2$ and urea diluted 10-fold. Controls include splicing conducted without renaturation of the pre-RNA in the presence (open triangles) or absence (open squares) of 300mM urea. Rates and fit parameters are given in Table 3.1.

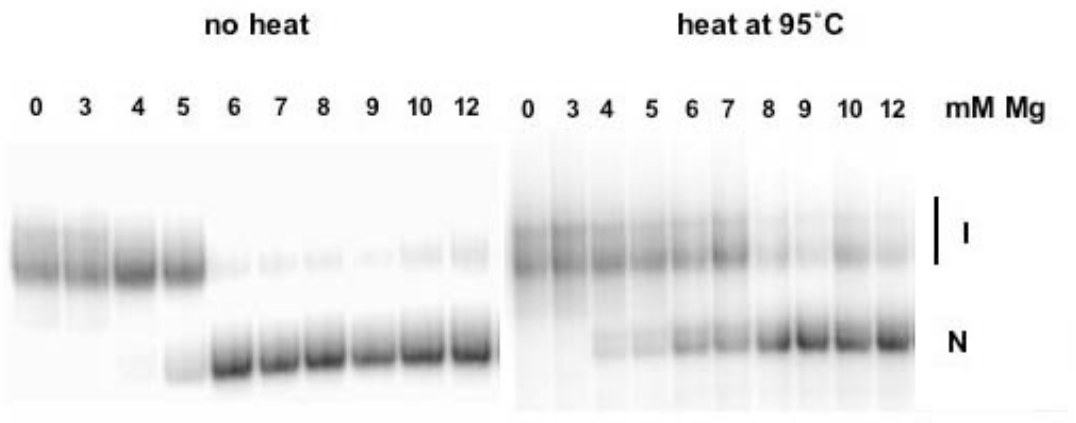


Figure 3.3: Magnesium dependence of folding of Ec-T7 pre-RNA.
Equilibrium folding experiments were conducted on Ec-T7 pre-RNA by incubating the RNA for 30 min at 30˚C in buffer containing varying Mg^{2+} concentrations as indicated above each lane. The samples were run on a non-denaturing 8% gel as seen here (Emerick and Woodson 1994). I, non-native intermediate. N, native RNA conformation (Emerick and Woodson 1994).

Heating the RNA to 95˚C increases the proportion of RNA populating the intermediate states even at high magnesium concentrations. These results corroborate our splicing data.

These two results taken together suggest that T7 transcripts of Ec-IVS RNA form stable interactions during transcription that are predominantly non-native. The pre-RNA with *Tetrahymena* exons however, does not form stable structures. Refolding of both the T7 and Eco transcripts of this RNA results in maximum splicing activity (Table 3.1).

Mapping the pause sites of *E. coli* RNAP

The results so far indicate that transcription by *E. coli* RNAP enhances the splicing activity of only the Ec pre-RNA and not the Tt pre-RNA, indicating that its effects on folding of the *Tetrahymena* group I intron are template dependent. As pausing of the polymerase involves interactions with the template DNA as well as the nascent RNA strand (Artismovitch and Landick 1998), site-specific pausing of the polymerase could be one reason for the template-dependent folding effects. Pan et al. (1999) have shown that pausing of *E. coli* RNAP enhances the folding of the RNase P RNA over refolding in the presence of magnesium salts. I thus decided to map the pause sites of *E. coli* RNAP on the known secondary structure of the exon and intron sequences.

Open complexes were initiated by incubating template DNA and polymerase (1:5 molar ratio) at 37˚C for 15 min. A single round of transcription was ensured by the addition of rifampicin. Elongation was initiated by the addition of NTPs and stopped at various times. The samples were processed as described in Materials and Methods and the results are shown in Figs 3.4 and 3.5.

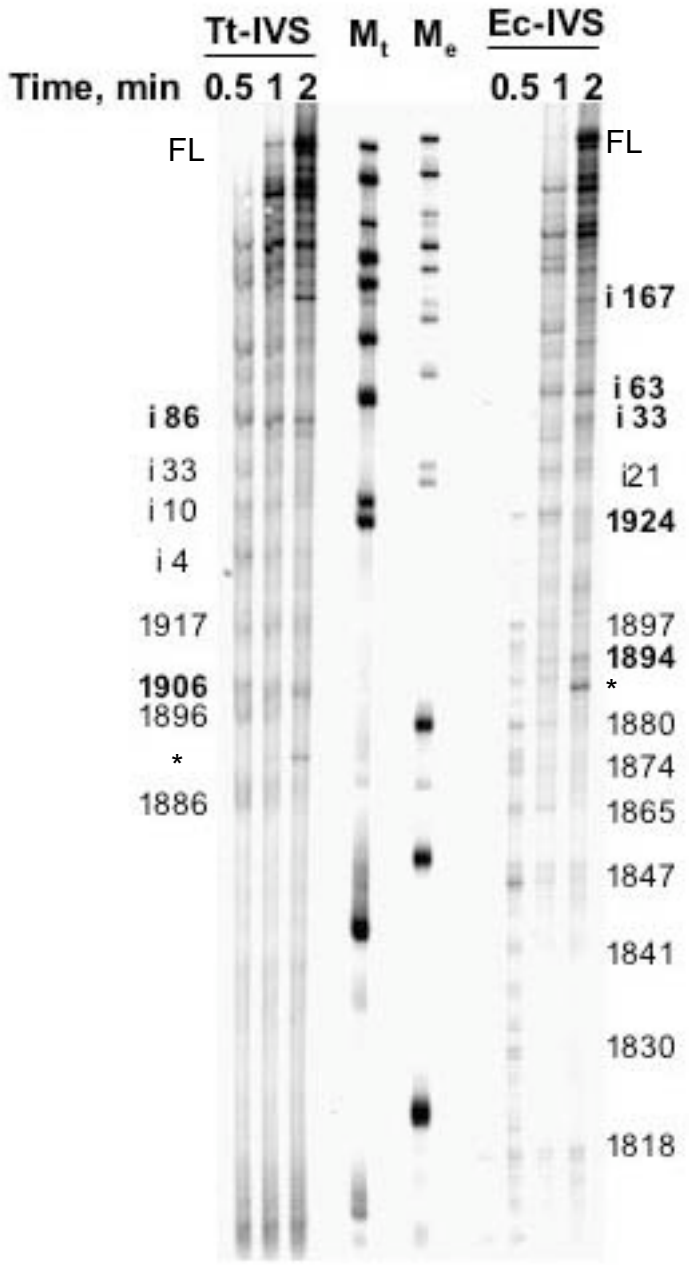


Figure 3.4: Pause sites of *E. coli* RNAP. Tt-IVS; pre-RNA with *Tetrahymena* rRNA exons. Ec-IVS; pre-RNA with *E. coli* rRNA exons. M_t and M_e are markers for Tt-IVS and Ec-IVS respectively. Bold numbering indicates strong pauses and plain numbering indicates mild pauses. Numbers are all according to the system used to number *E. coli* 23S rRNA. Numbers starting with "i" indicates nucleotide position in the intron. FL, full-length product; *, splicing products. Transcription of Ec pre-RNA starts at nt 1760 and Tt at nt 1861.

From Fig. 3.4, it is clear that there are two sets of pause sites. One set is chased away at longer times of elongation, these I call mild pauses. The other set of pauses are those that persist even after 2 min of elongation; these I call strong pauses. The polymerase pauses upstream of the intron more often on the *E. coli* rDNA template than on the *Tetrahymena* rDNA template. The intron however, has same number of pause sites in both the pre-RNAs. These results indicated that the different exon sequences cause the polymerase to pause differently on the two templates.

To see if this differential pausing of the polymerase is significant with respect to the folding of the pre-RNA, I mapped the pause sites on to the secondary structure of the pre-RNA (Fig 3.5). The Ec pre-RNA has a number of mild pauses along the P(-2) stem (nt. 1835 to 1905) and upstream at nt. 1818 and 1830. Apart from this it also has two significant strong pauses, one midway through P(-2) stem (at nt.1894) and one at the base of P(-1) (at nt.1923) (Fig 3.6). During transcription of the Tt pre-RNA, however there are only three mild pauses, one at nt 1917, upstream of the intron insertion site in the P(-1) helix, and two in the P(-2) stem (1886 and 1896). The only strong pause site on the *Tetrahymena* pre-RNA is at the base of the stem of P(-1) (nt. position 1906). The pause sites of the polymerase on the intron region of both these pre-RNA were identical within the error range (± 10 nt) except for the position i63 on the Ec pre-RNA and i83 on Tt pre-RNA (Fig 3.5)

Effect of S30 extract on in vitro self-splicing

As mentioned in Chapter 1, although there is no evidence for the presence of any species-specific protein factor in the in vivo facilitation of splicing, this could be

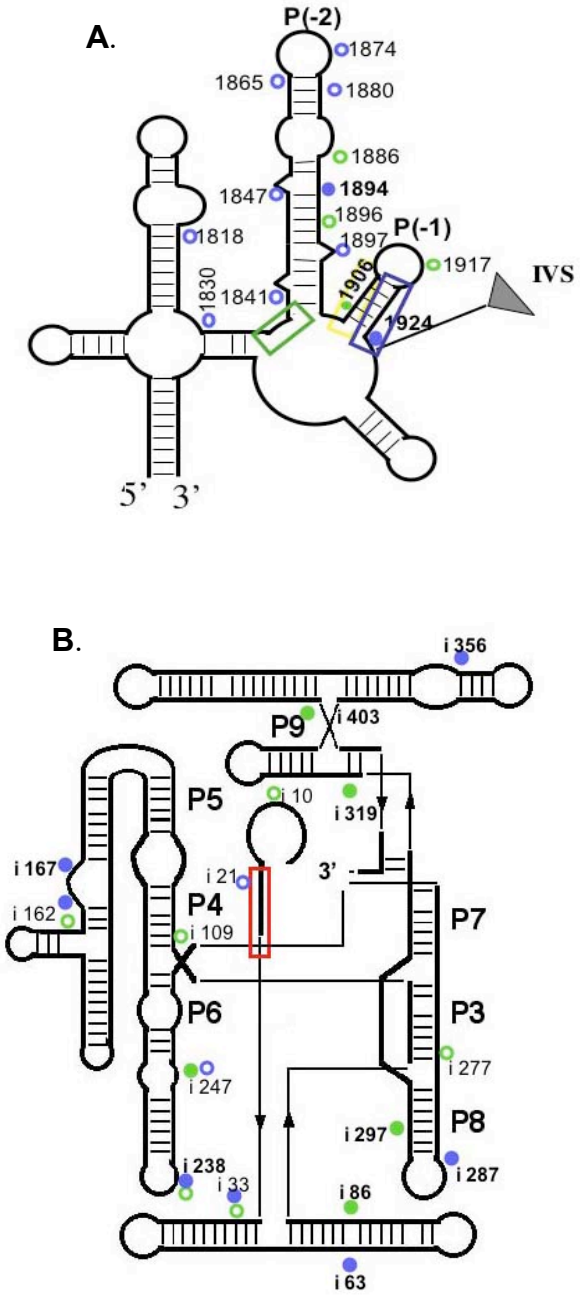


Figure 3.5: Map of pause sites of *E. coli* RNAP. Open circles indicate mild pauses and filled circles indicate strong pauses. Blue circles represent pausing on Ec pre-RNA and green circles those on Tt pre-RNA. Colored boxes indicate alternate pairings as shown in Fig 3.7. A. Secondary structure of exons (Noller 1991).B. Secondary structure of *Tetrahymena* intron (Cech et al 1994).

affected by generic proteins like ribosomal proteins. To verify this hypothesis, we studied in vitro self-splicing in the presence of commercial *E. coli* S30 extracts. These extracts includes ribosomal proteins, some of which have been reported to have chaperone activity (Zhang et al 1995b). In vitro self-splicing assays were conducted by incubating Ec-T7 pre-RNA with either S30 extracts or just buffer at 30˚C for 10 min (Fig 3.6). S30 extracts enhanced splicing in vitro by two-fold, but 50% of this enhancement was due to the buffer alone (Fig 3.6). Hence the addition of S30 extract does not account for the difference between in vitro and in vivo splicing rates.

DISCUSSION

As described before, Heilman-Miller and Woodson (2003) showed that cotranscriptional folding of the intron alone is 2-fold faster than refolding in vitro, establishing that folding during transcription is important. As folding during transcription can be manipulated by the nature of polymerase used, I measured the self-splicing activity of the *Tetrahymena* group I intron as a reporter of its folding process in this study. For this I use the full-length intron flanked by domain IV rRNA exons.

Facilitation of in vitro folding by *E. coli* RNAP

From my results it is clear that the polymerase used to transcribe the pre-RNA changes the splicing activity. For the hybrid pre-RNA with *E. coli* rRNA exons, the *E. coli*

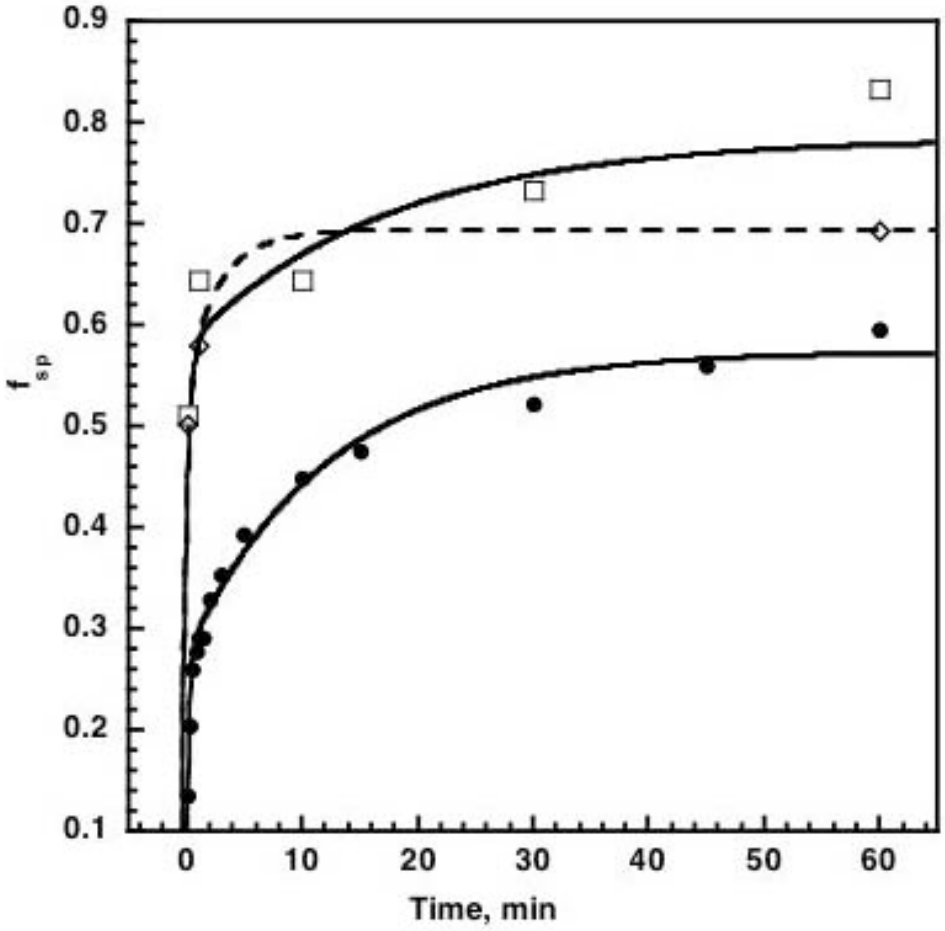


Figure 3.6: Effect of S30 extracts on in vitro self-splicing of Ec-T7 pre-RNA. Uniformly labeled RNA was incubated at 30˚C with S30 extract (open squares), extract buffer (open diamonds) or splicing buffer (filled circles) as described in Materials and Methods. Data are fit to double exponential equations with the same parameters as shown in Table 3.1. Burst sizes for Ec-T7 at different conditions are: 25% in splicing buffer, 50% in S30 storage buffer and 65% in S30 extract.

RNAP transcripts splice better than T7 transcripts (Fig 3.1). For the pre-RNA with *Tetrahymena* exons, however, this facilitatory effect is lost (Fig 3.1), with the *E. coli* transcripts splicing poorly than the T7 transcripts. Thus the facilitatory effect of *E. coli* RNAP is template dependent. This can be explained by the differential pausing of the *E. coli* RNA polymerase on the two different templates.

Pausing of *E. coli* RNAP is known to involve interactions of the polymerase with both the template DNA and the nascent RNA strand (Artismovitch and Landick, 1998). It has also been shown that pausing of *E. coli* RNAP helps in the folding of the RNase P RNA (Pan et al., 1999). They have shown that addition of the elongation factor NusA during transcription enhances folding of a circularly permuted variant of this RNA, suggesting that enhanced pausing due to NusA allows more time for rearrangement of nascent structures before further sequences are transcribed.

E. coli RNAP paused on distinct sites along the Ec and Tt pre-RNAs (Fig 3.4 and 3.5). This suggested to us that this differential pausing could be a reason for the differential folding facilitatory effect of the polymerase. To see if this differential pausing translates to a difference in the formation of native interactions, we used the MFOLD program to analyze the stability of the predicted nascent structures formed when the RNAP is paused at each of the strong pause sites these two pre-RNAs.

When the active site was at nucleotide 1924 on the Ec template, two out of the three most stable predicted structures contained the PX and P(-2) helices. These conformations favor 5' splice-site recognitions (Cao and Woodson 1998). The inhibitory P(-1) stem cannot form when the polymerase is paused at this nucleotide, because the nucleotides that form part of that stem will still be making contacts with the polymerase

and hence not be available to fold. To see if the P(-1) stem can form when all the nucleotides necessary for formation of this stem are available to fold, I analyzed the predicted structures formed when the polymerase is paused at the next strong pause site for this template, which is at the intron nucleotide position 33. When paused at this site, pre-RNA residues up to nucleotide number 23 in the intron are available to fold. Among the predicted structures the most stable conformer contained the PX helix and not P(-1). However, P(-1) helix was formed in 50% of the remaining structures predicted. These results indicate that pausing pattern by *E. coli* RNAP predisposes the Ec pre-RNA to fold into the active structure

Similar analysis of the structures predicted to form during transcription of the Tt pre-RNA by MFOLD gave the following results. When the polymerase was paused at intron position 33 (mild pause i33 in Fig 3.5), most of the stable structures contained P(-1) and none formed PX. When the polymerase was paused at the single strong pause site on the exon (position 1906, Fig 3.5), the predicted structure formed most of P(-2). PX cannot form in this structure, as the nucleotides needed to form this helix are not yet synthesized. Thus pausing of *E. coli* RNAP on the Tt pre-RNA predisposes the RNA to mostly form inactive structures.

The differential pausing of the polymerase on the exon region of the two pre-RNAs promotes differential pausing in the intron region also. This is because the intron conformation is partially coupled to the conformation of the exons. Thus changes in exon conformation leads to the difference in the pausing pattern observed in the intron region of these two pre-RNAs. One such pause site is at position 63 on the Ec template and position 83 on the Tt template (Fig 3.5). Analysis of the predicted structures

obtained by MFOLD gave the following results. Of the four structures with the highest predicted stability for Ec transcripts ending at nt 63, three of them had the PX stem, two of which had the native P(1) stem along with the PX stem. The remaining structure had only the P(-2) and the non-native P(-1) stem. Thus for the Ec pre-RNA, site-specific pausing of the polymerase on both the intron and exon regions favors formation of native interactions over non-native ones.

In the case of the Tt sequence, pausing at intron position 83 resulted only in three structures as predicted by MFOLD. Two of these (lowest energy) had only inactive stem P(-1) and no PX. The other predicted structure (higher energy) had the active P1 stem but still no PX. These results indicate that pausing on the intron region of the Tt template, like those on the exons of the same pre-RNA, favors formation of non-native interactions over native ones.

These results show that the observed pausing pattern of the *E. coli* RNAP could alter the balance between active and inactive conformations. As explained in the introduction chapter, it has been shown that the P1 stem formed by the IGS of the intron and the 3' region of the exon competes with the P(-1) stem. P(-1) sequesters the exon sequences that pair with the IGS, thus inhibiting 5' splice site recognition and splicing (Woodson and Cech, 1991). It has also been shown that long range interaction between sequences 82 nucleotides upstream of the splice site and sequences 3' of P(-2) which comprise the PX stem, enables the formation of P1 (Woodson and Emerick, 1993), (Fig 3.7 A.). This has led us to propose a mechanism for the way template dependent pausing of the *E. coli* RNAP affects in vitro folding. As shown in Fig. 3.7 B the strong pauses of the polymerase on the *E. coli* template promotes the formation of nascent structures that

enables formation of the correct native structure. However, pausing of the polymerase on the *Tetrahymena* template promotes the formation of inactive nascent structures resulting in RNA getting trapped in non-native conformations.

E. coli exon sequences form stable nascent structures

The previous paragraphs present a model of how the nascent secondary structures play a significant role in the formation of active pre-RNA. The stability of the nascent structures is also important because inactive intermediates need to unfold before active ones form. The stronger the nascent structures, the harder it is to denature them and hence slower the folding process. Comparison of the T7 transcripts helps us get an idea about this aspect of the pre-RNA. T7 RNAP has an average elongation rate of about 200-400 nt per second (Golomb and Chamberlin, 1974). Thus, essentially the whole pre-RNA is synthesized within two seconds. This means that the RNA can collapse into a compact structure as soon as the whole RNA is formed.

The splicing rate of the slow phase of Tt-T7 pre-RNA is faster than the corresponding rate for the Ec-T7 pre-RNA (Table 3.1). As this phase is limited by the refolding of the intermediates to the native structure (Emerick et al 1996), the observed rate increase suggests that rearrangement of inactive structures is faster for the Tt pre-RNA than the Ec pre-RNA.

This observation is strengthened by renaturation experiments, when the fraction of spliced RNA falls to 30% from 55% (Table 3.1, Fig 3.1). It takes a strong denaturant like urea to help the RNA regain full splicing activity (Fig 3.3), suggesting that the nascent structures formed by the Ec pre-RNA are very stable. Fig 3.4 shows that these nascent structures are predominantly non-native intermediates.

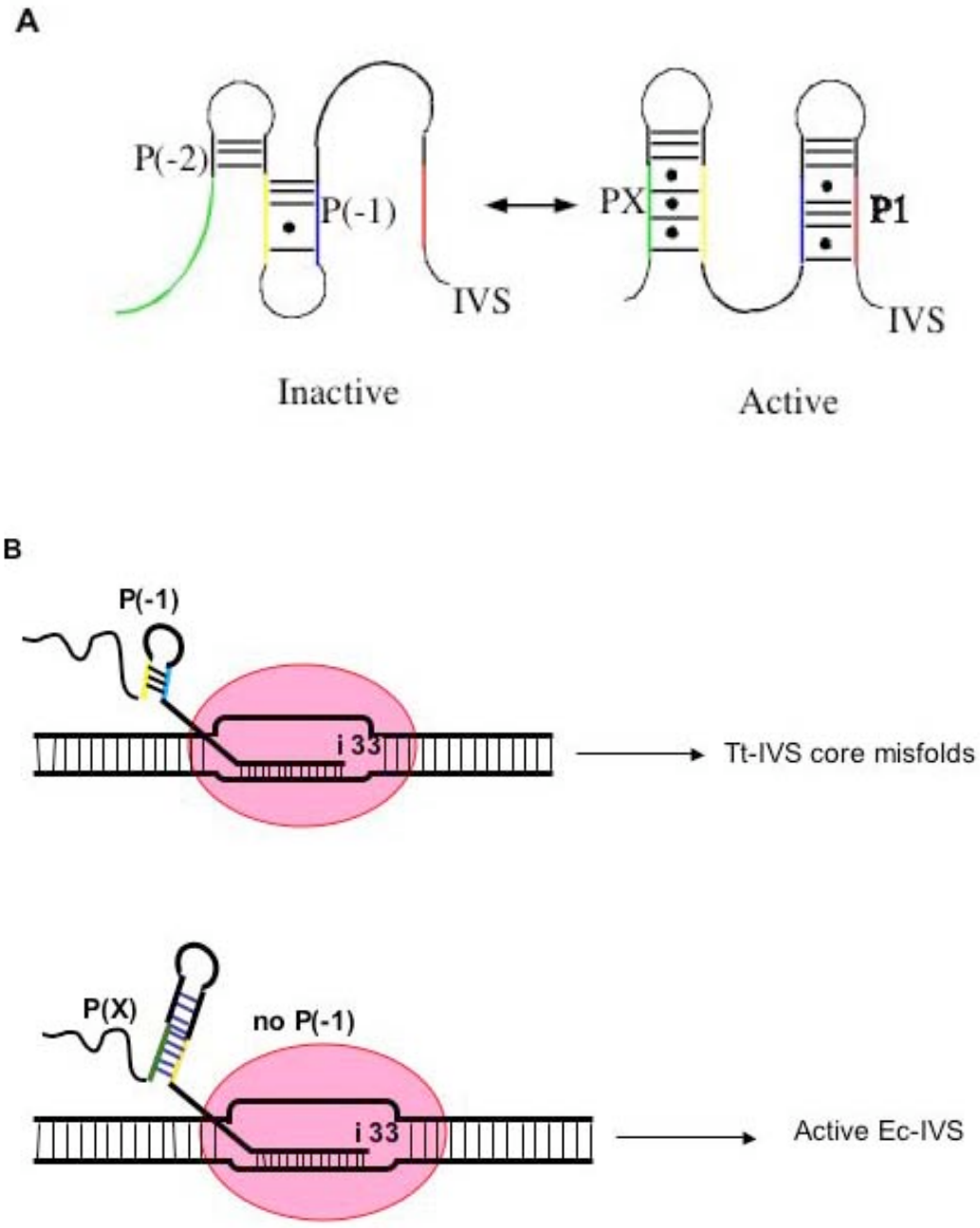


Figure 3.7: A. Alternative structures in the rRNA. Formation of P(-2) and P(-1) helices leads to the misfolding of the intron core resulting in loss of splicing activity. Rearrangement of these sequences to form PX and P1 helices leads to the formation of the active intron (Woodson and Emerick 1993). B. Model for pausing facilitated folding of the intron.

Although renaturation can increase the fraction of active RNAs in vitro, in vivo this does not occur. Thus the fraction of total RNA molecules that do not attain active structure (the third phase in vitro) would get degraded rapidly in vivo. Thus although some RNA molecules may fold fast they may still not be active as they could be trapped in misfolded structures that refold very slowly.

In conclusion, the results can be summarized as follows. The strategic pausing of the *E. coli* RNAP along the Ec pre-RNA promotes the formation of nascent structures in the 5' exon that are conducive to the formation of the active intron. These nascent structures are also quite stable and hence do not easily rearrange to the alternate, inactive structures. The pausing of the *E. coli* RNAP on the *Tetrahymena* pre-RNA however, leads to the formation of inactive nascent structures. These structures are however not very stable and can be destroyed by heating to 95˚C. This low stability of nascent structures is an asset when transcribed by T7 RNAP as they help the RNA scan through alternate structures of low stability until the native structure is achieved.

Chapter 4

EFFECT OF TRANSCRIPTIONAL PARAMETERS ON THE FOLDING OF THE *TETRAHYMENA* GROUP I INTRON IN YEAST

INTRODUCTION

In Chapter 2, I have shown that the in vivo folding facilitation of the *Tetrahymena* group I intron in *E. coli* is partly due to the combined effect of the polymerase and exon sequences. The in vitro splicing experiments presented in Chapter 3 suggest that site-specific pausing of *E. coli* RNA polymerase alters the balance between native and non-native interactions, and this pausing pattern varies with the exon sequence.

However, other factors may also contribute to the in vivo activity of group I introns. *Tetrahymena thermophila* is a eukaryote with a more complicated cellular environment than bacteria. Eukaryotes have three different RNA polymerases and these transcripts undergo different processing mechanisms. Interactions of the nascent transcript with the different proteins that are involved in these processing mechanisms could affect folding and hence splicing of the Tetrahymena group I intron.

Early studies in the Cech laboratory measured the rate of splicing and decay of this intron in its natural environment (Brehm and Cech 1983). Cech and Rio (1979) showed that splicing of this intron occurs early in pre-rRNA processing. More recently, Lin and Vogt (1998) have shown that this intron is spliced efficiently from the pol I transcripts of the rDNA repeats in *S. cerevisiae*. However, nothing is known about the

influence of various polymerases and exon sequences on the splicing of this intron in a eukaryote. Hence, we decided to study splicing of the *Tetrahymena* intron when transcribed by different eukaryotic polymerases and when flanked by different exons in yeast.

All the yeast experiments were done in collaboration with Scott Jackson. He did all the cloning and both of us did the RNA steady-state and decay analyses. The two of us also did the RNA extraction, northern gels and data analysis together.

MATERIALS AND METHODS

Plasmids

The *Tetrahymena* ribozyme was inserted into the homologous position of every rDNA repeat of *S. cerevisiae* strain INVSC2/TtLSU1 (Lin and Vogt 1998). We obtained this clone from the Vogt lab to measure splicing rates of pol I transcripts. The pNOY102 plasmid, containing the entire 35S rRNA operon placed under the control of a galactose inducible promoter in a 2 micron plasmid with a ura3 marker gene (Nogi et al 1993) was a kind gift from M. Nomura. The Tetrahymena group I intron sequences were inserted into this clone after position 1925 (E. coli numbering) by Scott Jackson, to produce the pNOY-IVS plasmid (map in Appendix B). These two systems were used to compare the effects of yeast polymerases I and II on the splicing of this intron in yeast. The *Tetrahymena* ribozyme flanked by 145 nt of domain IV 26S rRNA exon sequence at the 5' end and 86 nt of the same at the 3' was cloned into a 2-micron plasmid under a galactose inducible promoter by S. Jackson and called pSW015. This precursor is similar

to the Tt-IVS pre-RNA used in my *E. coli* and in vitro study. A similar plasmid containing the homologous sequences of yeast rDNA domain IV flanking the *Tetrahymena* intron was also cloned by S. Jackson and called pSJ015. S. Jackson made the clone pSJ831, which contains the *Tetrahymena* group I intron (with no rDNA exons) placed in front of sequences encoding the GFP mRNA. All these clones are shown in Table 4.1 and a detailed plasmid map is shown in Appendix B.

Steady state and intron decay

All clones containing a pol II promoter were grown in 200 ml SC media containing 2% galactose at 30˚C overnight. The yeast strains BY4733 containing pNOY102-IVS, pSJ015 or pSW015 were grown in SC media supplemented with 0.3 mM histidine. The SJ831 clone was grown in minimal media supplemented with 2 mM of leucine. Steady-state levels of plasmid expression were analyzed by harvesting 10 ml aliquots at 18 and 20 hours (OD_{600}~1.0). The cells were pelleted by spinning at maximum speed in a clinical centrifuge and frozen in dry ice. To measure decay of the intron, the cells were grown at 30˚C for 18 hrs and harvested at 5000 rpm (Beckman rotor JS13.1) at 4˚C for 5 min. after taking steady-state time points. The cells were then resuspended in 16 ml minimal media with the appropriate amino acid supplements and 2% glucose. 2 ml aliquots were removed at different times starting from immediately after resuspension in glucose-enriched media to 420 min. Cells were pelleted in a microfuge for a few seconds at maximum speed, and the pellets snap-frozen in dry ice. The INVSC2/TtLSU1 strain was grown in 500 ml of YPD (Bio 101 inc.) at 30˚C overnight with shaking. Cells were harvested after steady-state growth was reached as described above.

Yeast total RNA extraction

Total RNA from yeast cells were extracted by the acidic phenol method described in http://www.sanger.ac.uk/PostGenomic/S_pombe/docs/rnaextraction_website.pdf with slight modifications as described here. The pelleted cells were resuspended in 500 µl of TES containing 10% of sodium dodecylsulphate (SDS). An equal volume of acidic phenol was added and the cells vortexed vigorously for 2 min. The cells were then incubated for one hour at 65˚C with vortexing every 5 min for one minute. The cells were then placed in ice for 5 min and then centrifuged in a microcentrifuge at r_{max} for 5 min at 4˚C. The aqueous phase was extracted with acidic phenol again as described above but without the one-hour incubation. The freshly separated aqueous phase was then treated with a mixture of chloroform and isoamylalcohol (1:24) and vortexed vigorously for 10 seconds. The samples were centrifuged for 2 min. Equal volumes of chloroform was added to the aqueous phase and vortexed and centrifuged as above. The RNA was precipitated by addition of 40 µl 3 M sodium acetate (pH 5.3) and three volumes of ethanol to the aqueous phase from the previous step. The amount of total RNA extracted was measured by reading the absorbance at A_{260}.

Northern Analysis

Northern assays were conducted as described in Nikolcheva and Woodson (1999), except that 10 µg of total RNA was used instead of 5 µg. The membranes were hybridized with ^{32}P labeled probe for the intron and yeast actin mRNA, where the latter served as a control for loading. To obtain the ratio of intron to unspliced pre-rRNA in INVSC2/TtLSUI or NOY-IVS strains serial dilutions of total RNA (1-20 µg) were

blotted, and the amount of precursor or spliced intron RNA extrapolated from concentrations which produced a linear correlation between the northern blot signal and concentration.

RESULTS

As a precedent for studying folding of autocatalytic RNAs in vivo, folding of the hammerhead ribozyme has been studied previously by measuring its catalytic activity in yeast, which was found to correlate well with in vitro activity (Fedor 2001). As in Chapter 2, the splicing activity of the Tetrahymena group I intron was used to compare the effects of polymerases, exons and RNA processing events on the folding of the intron in yeast. Splicing rates in yeast were measured in the same way as they were measured in *E. coli*, as the product of the steady-state ratio of intron to precursor levels and the decay rate of the free intron.

At steady-state, in the presence of galactose, transcripts accumulated to levels comparable to the endogenous actin mRNA (Fig 4.1). The ratios of spliced intron to precursor were measured from northern blots. The amount of radioactivity in the spliced intron band and the unspliced precursor band were normalized to the actin signal to account for differences in cell growth or RNA recovery. Total RNA from cells harvested at various times after switching the carbon source to glucose, were used to measure intron decay rate. The decrease in the intensity of the intron band in northern blots was plotted versus time and fit to a single exponential decay equation to give the intron decay rate, which was found to be 0.02 min^{-1}.

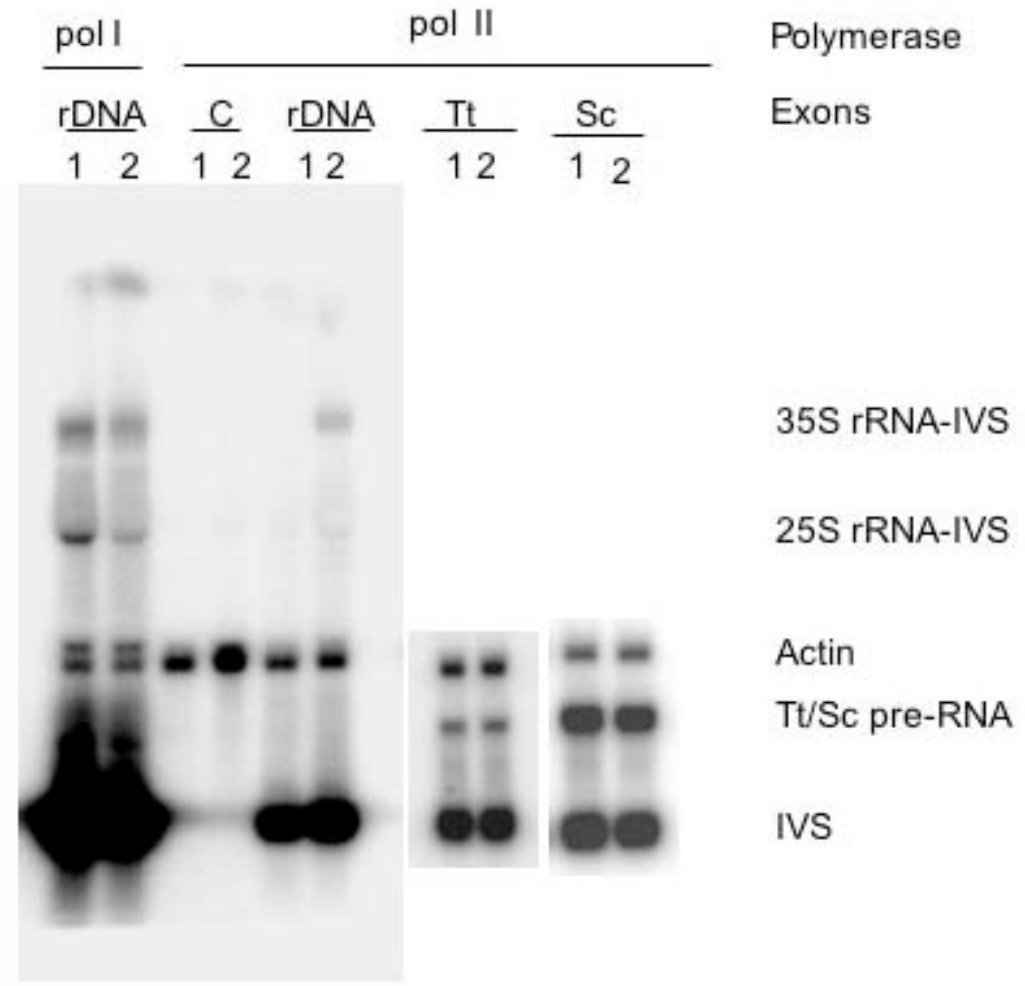


Figure 4.1: Northern analysis of in vivo splicing. 10 g total yeast RNA from cells harvested after expression in galactose reached steady-state at 18 (1) and 20 (2) hrs. These were probed with intron and actin probes. rDNA-pol I, pol I transcripts of. INVSC2/TtLSU1; rDNA-pol II, pol II transcripts of pNOY-IVS; C, control RNA that does not contain the intron in the rDNA operon; Tt-pol II, pol II transcripts of pSW015; Sc-pol II: pol II transcripts of pSJ015.

Splicing of the Group I Intron from pol I Transcripts in yeast

The intron is spliced very efficiently from the pre-rRNA transcript of rDNA repeats from the yeast chromosome XII. As can be seen in Fig 4.1, at steady state most of the (98%) precursor RNA is spliced and the intron band is overexposed while the precursor (35S rRNA) band is just visible. In order to get an accurate measure of the steady-state ratio of intron to precursor levels, it was necessary do a standard curve where different amounts of total RNA were loaded on the gel and the membrane was exposed for different lengths of times. The values were obtained by linear extrapolation from the optimal exposure intensity for the intron and precursor band. From this value we calculate a splicing rate of 3.9 min^{-1} (Table 4.1) for this transcript, which is the highest splicing rate we measure. This indicates that splicing is the best in the natural environment.

Effect of Polymerase on Splicing in yeast

In *E. coli* I showed that *E. coli* RNAP facilitates splicing of the RNA when flanked by *E. coli* exon sequences, compared to T7 RNAP. Therefore, I wanted to see if eukaryotic polymerases also affect splicing in yeast. To do this I compared the splicing of the pre-RNA that contains the intron flanked by whole chromosomal rDNA that is transcribed by either pol I or pol II. We find that when the intron is transcribed by pol I, it splices 20-fold better than when it is transcribed by pol II (Table 4.1). This suggests that the splicing of the intron is the best when transcribed by the native RNAP.

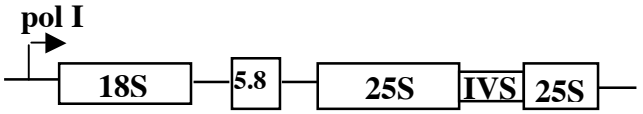
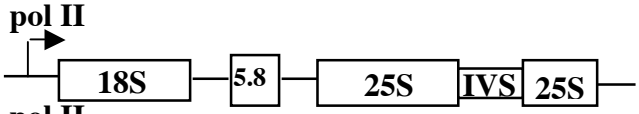
Effect of Exons on Splicing in yeast

In *E. coli* and in vitro, I find that the domain IV of the 23S rRNA sequences of *E. coli* conferred more stability to the nascent pre-RNA structures than the corresponding sequences from *Tetrahymena* rRNA. Combined with the site-specific pausing of the *E. coli* RNAP, this selectively enhanced splicing of the Ec-IVS pre-RNA. To see if different exon sequences in yeast had a similar differential effect, we tested the splicing of the *Tetrahymena* intron when flanked by *S. cerevisiae* domain IV rRNA, *Tetrahymena* domain IV rRNA or generic mRNA (GFP) sequences. All these pre-RNAs are transcribed by pol II and are polyadenylated (S.J., personal communication). We find that pre-RNAs with *Tetrahymena* domain IV 26S rRNA exons (Tt-IVS) have three-fold higher splicing than those with the yeast domain IV rRNA exons (Sc-IVS). Sc-IVS and GFP-IVS, however, have similar splicing rates (Table 4.1). Thus natural sequences enhance activity slightly. As these exon effects are subtle, we wanted to see if splicing of mutant intron accentuates this difference.

Splicing of mutant introns in yeast

Pre-RNA with a wild type intron core showed only about 2-fold difference in splicing rate when transcribed by phage or prokaryotic polymerase in *E. coli* (Table 2.1, Chapter 2). The difference between the two transcripts was accentuated when I compared splicing of the pre-RNA with a mutant intron core. To see if folding mutations cause similar retardation in splicing of the pol II transcript RNAs, splicing rates of three different mutants was measured in yeast. The mutation G100C destabilizes the native P3

Table 4.1. In vivo splicing rates (min^{-1}) of the *Tetrahymena* pre-RNA.

Clone	Splicing rate, min^{-1} [a]
Wild Type Intron Core	
	3.9 ± 0.621
	0.19 ± 0.061
	0.023 ± 0.005
	0.061 ± 0.038
	0.030 ± 0.016
Mutant Intron Core	
G100C[b]	0.010 ± 0.001
C260G[b]	0.00
G264A[b]	0.00

In vivo splicing rates were calculated from northerns. Variations are standard deviations of the mean of 2-4 trials. Sc-IVS: *Tetrahymena* intron flanked by domain IV of yeast 25S rRNA. Tt-IVS: *Tetrahymena* intron flanked by domain IV of *Tetrahymena* 26S rRNA.

[a]. In vivo splicing rates calculated using decay rate of intron, which is 0.02 min^{-1}. [b]. Mutant introns in the Sc-IVS background

helix in the intron core while favoring the formation of an alternate helix P(-1) leading to an inactive pre-RNA (Pan and Woodson 1998). The mutant pre-RNA was found to barely fold in 6 mM $MgCl_2$ in vitro. This mutation was introduced into the pre-RNA with the domain IV Sc exon sequences. Its splicing in yeast was reduced by only 2-fold at 30˚C (Table 4.1), suggesting that splicing of this structural mutant is rescued in yeast.

The mutation G264A disrupts the G-binding site in the P7 helix and was shown to severely inhibit the second step of splicing in vitro (Been and Perrotta 1991). In yeast, this mutation completely abolishes splicing of the Sc-IVS pre-RNA (Table 4.1) suggesting that splicing of the G-site mutation is not rescued in yeast.

The mutation C260G that destabilizes the triple helix at the intron core was already described in Chapter 2. This mutation reduced splicing by 50-100-fold for the Tt-Eco pre-RNA (Table 2.1) in *E. coli*. This mutation completely abolished splicing of the Sc-IVS pre-RNA in yeast (Table 4.1). This again suggests that mutations that reduce the stability of the native structure are not rescued in yeast pol II transcripts.

In conclusion, splicing in yeast is best when transcribed by pol I. Splicing from the pre-RNA with the entire rRNA operon as exons is better than from those with short rRNA exons. Lastly, mutations in the intron core are more deleterious than exon-related misfolds. These all suggest that splicing in yeast is much more complicated in than that in E. coli. These are discussed below.

DISCUSSION

Role of Polymerases

As described in the introduction, the aim of this study is to try to understand the cellular mechanisms that enable efficient splicing of the *Tetrahymena* intron by promoting correct folding. In *E. coli*, I found that a large portion of this folding facilitation occurs during transcription of the pre-RNA when interactions between the polymerase and nascent RNA strand translate to a disruption in the balance between formation of active and inactive structures. In yeast, the transcription mechanism is more complicated. Not only are there three kinds of polymerases used to transcribe different kinds of RNA, but these transcripts interact with different sets of proteins and undergo different processing mechanisms that are necessary for the function of the RNA (Howes 2002, Smith and Steitz 1997). Thus differences in splicing activity could also reflect differences in the RNA packaging mechanisms.

Our results show that splicing of pol I transcripts are 20-fold faster than those of pol II. The pol II transcribed pre-rRNA undergoes normal pre-rRNA processing mechanisms (Nogi et al 1991). Moreover, in cases when the source of ribosomal RNA is solely pol II derived, cell growth is retarded only slightly (~20-40%), suggesting that rRNA processing events are not fatally compromised (Oakes et al 1998). However, a study of the nucleolar morphology in strains containing the pol II derived rRNA (without the intron), show that pol I and its interaction with pol I-specific transcriptional machinery is required for the normal crescent shape and perinuclear localization of the nucleolus (Oakes et al 1998). Proper nucleolar morphology has been shown to be

associated with correct biogenesis of the ribosome (reviewed in Shaw and Jordan 1995). Thus the lower splicing rate of the pol II transcribed pre-RNA could be either a reflection of the differential interaction of pol II transcriptional machinery with the rDNA compared to the interaction of the pol I transcriptional machinery with the rDNA, or be due to the difference in the polymerase elongation rates or a combination of both.

Transcription of rDNA repeats by pol I yields high amounts of pre-RNA. There is a possibility that these rRNA copies titrate out a group I intron splicing-specific inhibitor that might be present in yeast. However, it should be noted there is no evidence in literature for the presence of such a species-specific splicing inhibitor in yeast.

In addition to a perturbation of nucleolar structure, the properties of the polymerase elongation complex may also affect the folding of nascent transcripts. The pol I polymerase has an elongation rate of about 60 nt/sec (French et al 2003) while that of pol II is 5-25 nt/sec (Kadesch and Chamberlin 1982, Manley 1984 and Marzluff and Huang 1984). Moreover, there is no direct evidence for pausing of pol I RNAP during elongation. The pausing of pol II RNAP, however, has been shown to affect alternative splice-site selection (Roberts et al 1998; de la Mata et al 2003).

Role of exon structure

Comparing the splicing of pol II transcripts we find that the pre-RNA with the *Tetrahymena* domain IV rRNA exons splices 3 times faster than those with the yeast exons. As these two pre-RNAs are processed in a similar manner, this difference has to be due to the nature of the exon sequences. The primary sequence of these two rRNAs is highly conserved (Noller 1991), as is their secondary structure. Thus it can be concluded

that the difference in splicing rates is a reflection of a difference in the stability of the structures adopted by these two exon sequences. In *E. coli*, I have shown that the nascent structures formed by the *E. coli* rRNA exons are more stable than those formed by the *Tetrahymena* exons. Thus it is possible that the yeast rRNA exons form less stable structures while it is being transcribed than the *Tetrahymena* rRNA exons and hence decay faster than the latter. Measuring the decay rate of the precursor RNA of non-splicing variants of these two pre-RNAs could help answer this question.

The pre-rRNA with the full length yeast rRNA exons splices 10-fold faster than those with the minimal domain IV rRNA exons. This difference could be due to association of proteins during transcription. While the pol II mRNAs are bound by various pre-mRNA processing proteins like hnRNPs (Howe 2002), the pol I transcripts are associated with ribosomal proteins and other factors involved with ribosome biogenesis (Smith and Steitz 1997). Another explanation for this difference in splicing rate could be the presence of the extra rRNA sequences. It has been shown in vitro and in *E. coli* that longer rRNA exons facilitates splicing of this intron possibly by disrupting formation of alternative structures that are inactive (Woodson 1992 and Nikolcheva and Woodson 1999).

To distinguish between these two possibilities, we compared the splicing of mutant pre-RNA. We find that when flanked by the minimal domain IV RNA, the G100C mutant pre-RNA splices 2-fold slower than the wild type intron (experiments and data processing done by me and Scott Jackson). When flanked by the full length rRNA, however, this mutant splices 10-fold slower than its corresponding wild type precursor (experiments and data processing done by Scott Jackson), although the actual

rate is the same as the Sc-IVS-G100C mutant. This might suggest that the splicing facilitation seen between the wild-type precursors is probably more due to RNA processing events. Splicing of folding mutants, C260G and G264A, however, are negligible when flanked by either the entire rRNA or by the minimal rRNA exon sequences. This suggests that whatever the facilitatory mechanism, folding mutants are not rescued in yeast.

Chapter 5

DISCUSSION

The goal of my thesis is to understand the mechanism of in vivo folding facilitation of the *Tetrahymena* group I intron and to compare the in vivo folding pathway with the well characterized in vitro folding pathway. There are two main differences between in vivo and in vitro folding (as measured by splicing activity) of this RNA. One is the 20-50 fold rate enhancement seen in vivo and the other is the distinct triphasic nature of self-splicing, which is not reported in vivo. As it has already been established that the facilitatory mechanism is not due to any species-specific protein factor such as a maturase (Zhang et al 1995a), we wanted to see if generic cellular mechanisms could account for these differences. One such mechanism is transcription of the RNA, where differential rate of elongation and differential pausing of the polymerase have been shown to alter the folding of the nascent transcript (Kramer and Mills 1981, Monforte et al 1990, Roberts et al 1998).

Traditionally, in vitro folding experiments have been conducted on RNA transcribed by phage T7 RNAP while in vivo experiments used RNA transcribed by the host RNA polymerase, *E. coli* RNAP or *Tetrahymena* pol I (Emerick and Woodson 1997, Nikolcheva and Woodson 1999 and Brehm and Cech 1983). As there are considerable differences in the elongation rate and pausing characteristics of these polymerases we decided to see if the in vivo folding facilitation was due to these attributes of the polymerase.

COMPARISON OF SPLICING IN VITRO AND IN *E. COLI*.

Facilitation of in vivo splicing by polymerase

If transcription by *E. coli* RNA polymerase was the sole reason for the rate enhancement seen in vivo, then self-splicing activity of in vitro transcripts of this polymerase should be comparable to the splicing activity in vivo. As a corollary, the splicing activity of T7 transcripts of this intron in *E. coli* must be comparable to the self-splicing activity of the same in vitro. This is true for the Tt-IVS pre-RNA (comparing fast phase rates of self-splicing and splicing rates in *E. coli* at 30˚C shown in Tables 3.1 and 2.1). The approximate two-fold higher self-splicing rates of these transcripts could be due to the high error in measuring rates faster than 1 min^{-1} in vitro. This suggests that for this RNA the difference between in vivo and in vitro splicing rates observed in earlier studies were due to the difference in the polymerase used to transcribe it.

For the Ec-IVS pre-RNA, although the in vivo splicing rate of the T7 transcript is only 3-fold faster than the fast phase of self-splicing in vitro, the in vivo splicing rate for the Eco transcripts is 10-fold faster than in vitro self-splicing (Table 2.1 and fast phase rates in Table 3.1). However, both these rate differences are still smaller than previous data (20-50 fold difference, Brehm and Cech 1983 and Zhang et al 1995).

One possible explanation for the polymerase specific effects between in vitro and in vivo activity for the Ec pre-RNA, is the preferential binding of *E. coli* ribosomal proteins to the *E. coli* RNAP transcript. Binding of *E. coli* ribosomal proteins could promote correct folding of RNA by enhancing the stability of native interactions in the nascent transcript. For example, the binding site for the ribosomal protein L2 is in

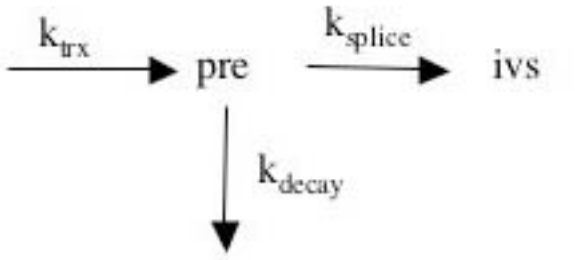
domain IV of the 23S rRNA that forms the 5'exon of my Ec pre-RNA (Beauclerk and Cundliffe, 1988). If L2 binds the transcript, the enhanced splicing of the Eco transcript compared to the T7 transcript of Ec-IVS may be due to stable misfolded structures formed during transcription by T7 RNAP, resulting in reduced binding of the proteins.

There is some evidence that T7 transcripts fold poorly in *E. coli*. It has been shown that T7 transcripts of the 23S rRNA do not form functional 70S ribosomes in *E. coli* (Lewicki et al 1993). This could explain why I do not find any significant increase in self-splicing when I add *E. coli* S30 extracts to T7 transcripts of Ec-IVS pre-RNA (Fig 3.6, chapter 3).

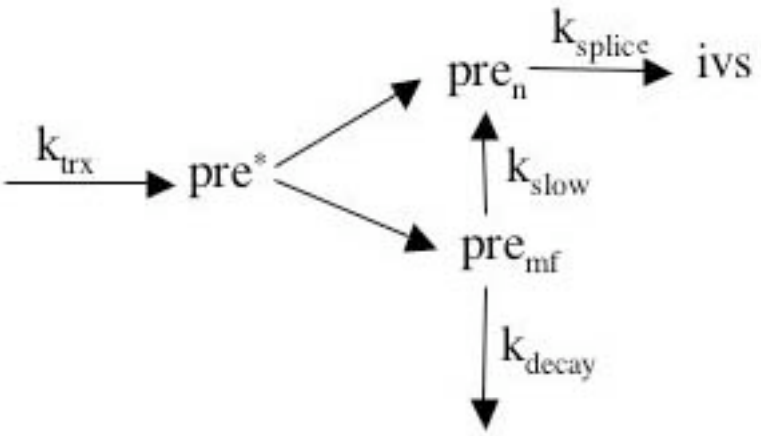
Triphasic splicing in vivo

Another difference between in vivo and in vitro splicing of the *Tetrahymena* intron is the distinct triphasic nature of in vitro splicing that is not observed in vivo (Emerick et al 1996). This could either mean that the entire nascent RNA population folds into the correct structure in vivo without forming misfolded intermediates or these misfolded intermediates are short lived. As the slow splicing phase represents the splicing of the slow folding RNA population in vitro, it is possible that in vivo the misfolded structures of low stability are degraded faster than they can splice. This could be the reason for the apparent non-triphasic nature of in vivo splicing.

The fate of the pre-RNA in vivo can be summarized as shown in the scheme below:

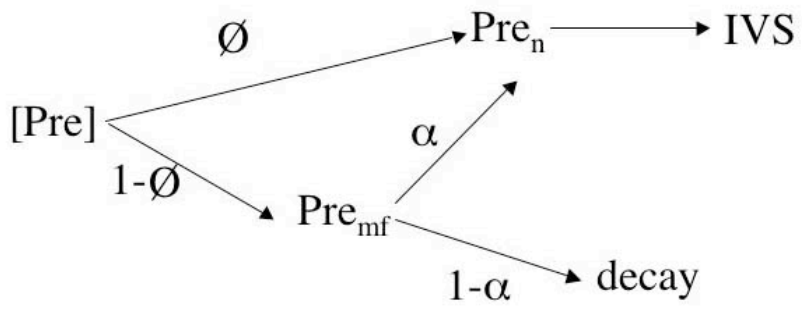


Thus the average decay rate of the pre-RNA reflects the sum of k_{splice} and k_{decay} values. As explained in Chapter 2, the low values of the observed average decay rate of the pre-RNA in *E. coli* (Table 2.2), suggests that the decay rate we measure is that of the unfolded RNA population. Thus the above scheme can be modified as:



where, pre^* is the entire population of nascent pre-RNA that can either fold to the native (pre_n) structure which splices fast or can misfold (pre_{mf}). The misfolded pre-RNA refolds very slowly (k_{slow}) to the native structure or is degraded (k_{decay}) quickly. Thus folding of the RNA in vivo also takes place via multiple pathways.

As the precursor decay data indicate that the precursor RNA gets partitioned into misfolded and native RNA populations, we need to revisit the splicing rates reported in Table 2.1. These values were taken to represent rate constants for splicing, but in reality it is more probable that these represent relative fractions of active RNA. In order to get an idea of the absolute splicing rate we will have to take into account the partitioning factor as well as the rate of refolding while measuring ivs/pre ratios at steady state. This is summarized by the following scheme:



Thus $k_{sp} = \frac{[ivs]}{[pre]} \cdot \frac{k_d}{Ø + \square(1-ø)}$

Where, ø is partitioning factor between active and inactive pre-RNA, and □ is the fraction of inactive pre-RNA that refolds before it degrades,

Although we cannot measure the absolute splicing rates with accuracy from the data reported in the present study, one can nevertheless explore the limits for these rates by making certain assumptions. It is unlikely that the refolding rate is slower in vivo than in vitro, thus we can estimate a lowest value for alpha from the in vitro slow folding rates. The in vitro folding rate (k_{fold}) of a short precursor is 0.14 min^{-1} (Pan and

Woodson, 1998) and the decay rate of the non-splicing (G264) mutant precursor (k_d pre) is 0.12 min^{-1} (Zhang et al 1995). Thus,

$$\square = k_{fold}/ k_{fold}+k_{decay} \sim 0.5$$

If we assume that the partitioning factor (ø) is nearly 100% for highly active RNA like Ec-Eco, then we can estimate ø value for a less active RNA for example, Tt-Eco.

Let $\frac{[ivs]}{[pre]} = \frac{ø}{(1- ø)}$ or $\frac{ø + \square(1- ø)}{(1- ø)(1-\square)}$

If ø for Ec-Eco is 0.95, then

$$\frac{Ø'}{1- Ø'} = \frac{k_{sp}'}{23\ min^{-1}} \bullet \frac{95}{5}$$; for Tt-Eco, Ø'=0.76

These ø values are higher than those observed for pre-RNA folded in vitro without renaturation (0.13; Pan et al 1997).

The map of transcription pause sites presented in Chapter 3 suggest that the enhanced splicing in vivo can be attributed to differential interaction of nascent RNA with the polymerase. Because the conformation of the intron is coupled to that of the rRNA, co-transcriptional folding of the exons has a measurable effect on splicing activity. Second, the apparent lack of triphasic kinetics of splicing in vivo can be attributed to a greater yield of active RNA and to the degradation of the misfolded structures rather than the absence of misfolding. These results suggest that bacterial cells (and possibly yeast too) solve the RNA folding problem by exercising an active surveillance over nascent transcripts and actively degrading those that are poorly folded.

APPENDIX A

List of primers

I. Primers used for cloning in E. coli.

1) TN112BAM 5'E(31): 5' GCA CGG ATC CAG CTG GCT GCA ACT GTT TAT T 3'

2) TN112BAM 3'E(28): 5' GCA CGG ATC CTG GGG CTC TGA GTC ACT T 3'

Upstream and downstream primers for making EC-IVS. Contains BamH I site at the 5' ends.

3) SW012BAM 5'E(31): 5' GCA GGG ATC CGG TAA TCC GAC TGT TTA ATA A 3'

4). SW012BAM 3'E(28): 5' GCA CGG ATC CTA GTA GAT AGG GAC AGT G 3'

Upstream and downstream primers for making Tt-IVS. Contains BamH I sites at the 5' ends.

II. Primers used for generating templates for in vitro transcription.

KPS ECO-pET(74): 5' **TGT TGA CAA TTA ATC ATC CGG CTC GTA TAA TGT** GTG GAA TTG GGG AAT TGT GAG CGG ATA ACA ATT CCC CTC TA 3'

Used to make ECO templates in vitro. Contains the trc promoter at the 5'end shown in bold characters.

III. Downstream primers used for making molecular weight markers for pausing assays.

IP006(18): 5’ CCA AAG GTA AAT ATT GCT 3’

Also used for primer extension reactions conducted on E. coli total RNA.

DPCPA433(27): 5’ TGA TAA CTT TTC CCT CCA ACG AGT ACT 3’

DPCPA477(27): 5’ GAT GCA ATC TAT TGG TTT AAA GAC TAG 3’

DPCPA485(27): 5’ TTT TAA ACC GAT GCA ATC TAT TGG TTT 3’

DPCPA505(27): 5’ CGC AAT TTG ACG GTC TTG CCT TTT AAA 3’

DPCPA543(27): 5’ CTG AGA CTT GGT ACT GAA CGG CTG TTG 3’

DPCPA589(27): 5’ GTC AGC TTA TTA CCA TAC CCT TTG CAA 3’

DPCPA645(27): 5’ ATA TCA ACA GAA GAT CTG TTG ACT TAG 3’

DPCPA673(27): 5’ GAC CGA CAT TTA GTC TGT GAA CTG CAT 3’

DPCPA697(24): 5’ TTA TGA GAA TAC ATC TTC CCC GAC 3’

DPCPA725(27): 5’ TCC CAT TAA GGA GAG GTC CGA CTA TAT 3’

DPCPA(WT)414(27): 5’CGA GTA CTC CAA AAC TAA TCA ATA TAC 3’

All these bind to the intron positions mentioned in the primer name.

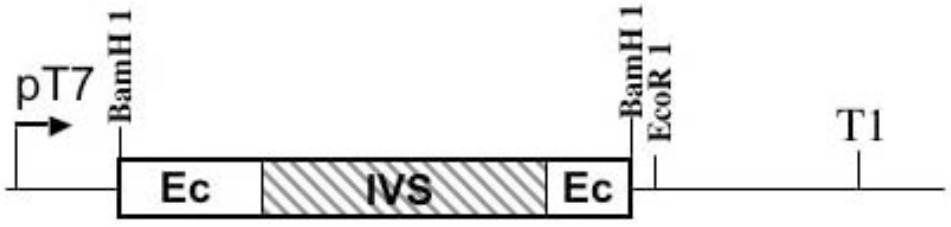
APPENDIX B

PLASMIDS USED IN E. COLI STUDY

pTN21: Size of plasmid: 6.0kb; vector pET 21

176nt 5'exon, 90nt 3'exon, exon: *E. coli* domain IV rRNA

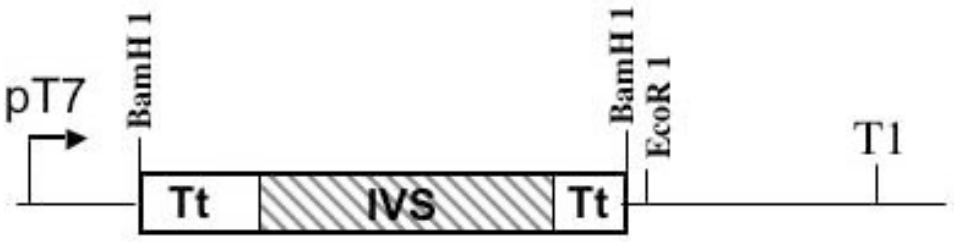
30nt 5' vector, 165nt 3'vector; T1: T7 terminator



pSW21: Size of plasmid: 6.0kb; vector pET 21

165nt 5' exon, 86nt 3' exon; exon: *Tetrahymena* domain IV rRNA

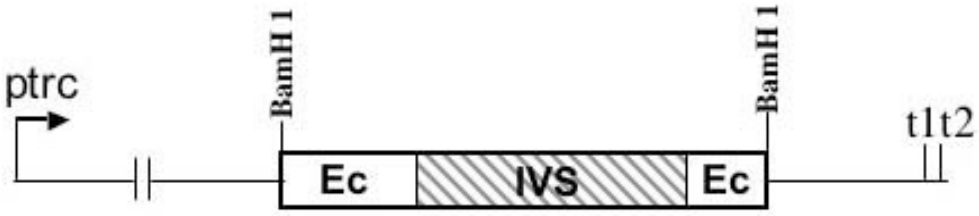
30nt 5' vector, 165nt 3' vector; T1: T7 terminator



pTN380: Size of plasmid: 5.1kb; vector pSE380

176nt 5'exon, 90nt 3'exon, exon: *E. coli* domain IV rRNA

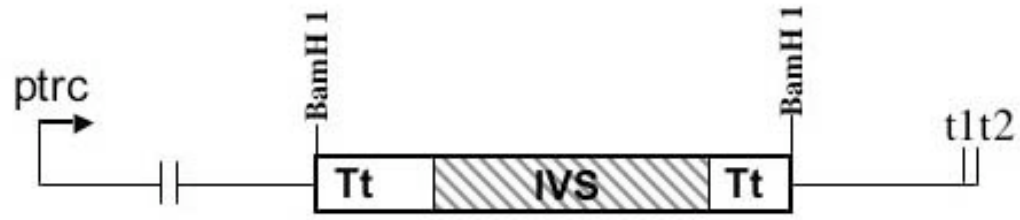
57nt 5' vector, 250nt 3' vector, t1t2: terminator



pSW380: Size of plasmid: 5.1kb; vector pSE380

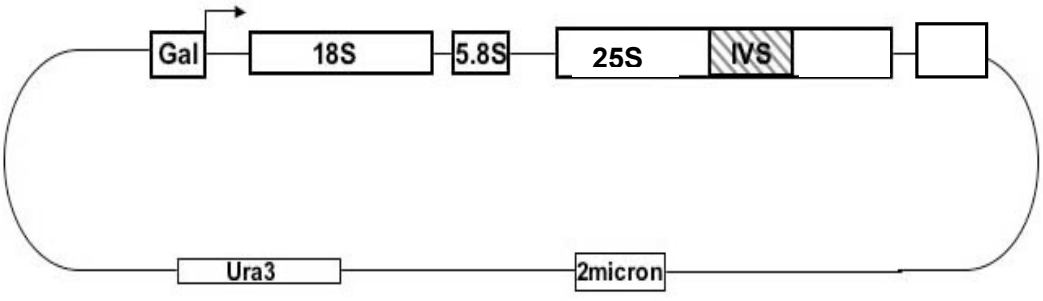
165nt 5' exon, 86nt 3' exon; exon: *T. thermophila* domain IV rRNA

57nt 5' vector, 250nt 3' vector, t1t2: terminator



PLASMIDS USED IN YEAST STUDY

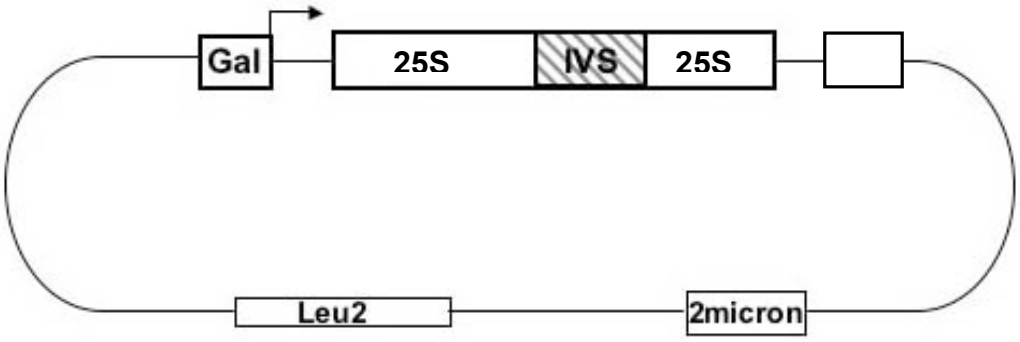
pNOY-IVS: Size of plasmid: 16.6kb; CYC1, transcription terminator



pSJ015: Size of plasmid: 7.1 kb; vector p425-Gal1

145nt 5' exon, 86nt 3'exon; exon: *S. cerevisiae* 25S domain IV rRNA

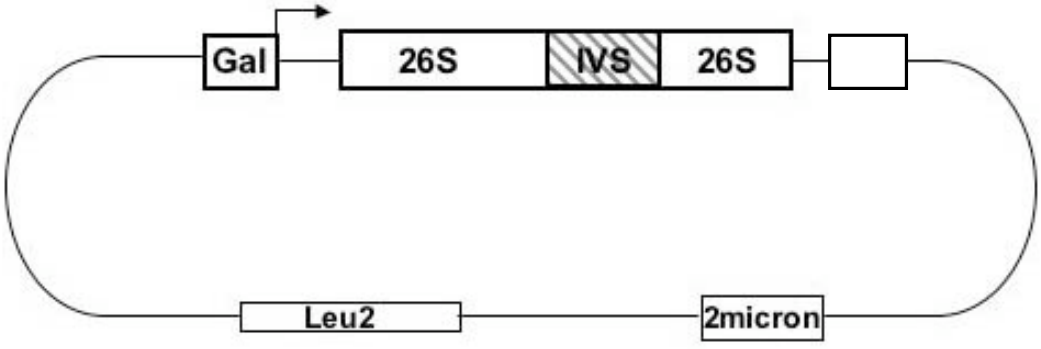
CYC1: transcription terminator



pSW015: Size of plasmid: 7.1kb; vector p425-Gal1

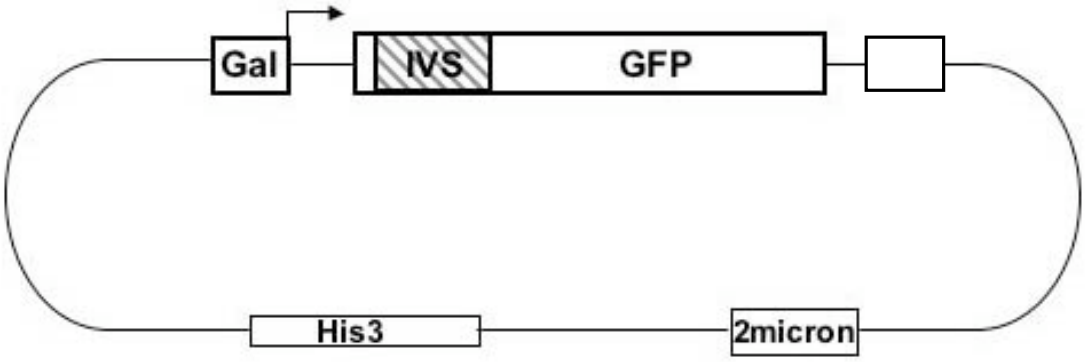
145nt 5' exon, 86nt 3' exon; exon: *T. thermophila* 26S domain IV rRNA

CYC1: transcription terminator



pSJ831: Size of plasmid: 7.5kb; vector p425-Gal1

CYC1: transcription terminator



REFERENCES

Alain F.H.-T. and Varani G. *Nucl. Acids Res.* **23**, 341-350 (1995).

Artismovitch I. and Landick R. *Genes Dev* **12**, 3110-3122 (1998).

Banerjee A.R., Jaeger J.A. and Turner D.H. *Biochemistry* **32**, 153-163 (1993).

Banerjee A.R. and Turner D.H. *Biochemistry* **34**, 6504-6512 (1995).

Barford E.T. and Cech T.R. *Mol Cell Biol*. **9**, 3657-3666 (1989).

Bass B.L. and Cech T.R. *Nature* **308**, 820-826 (1984)

Beauclerk A.A. and Cundliffe E. *EMBO J* **7**, 3589-3594 (1988).

Beaudry A.A. and Joyce G.F. *Biochemistry* **29**, 6534-6539 (1990).

Bechhofer D. In Belasco JG, Brawerman G, eds. *Control of messenger RNA stability*.

New York: Academic Press; 31-52 (1993).

Been M. and Cech T.R. *Cell* **47**: 207-216 (1986).

Been M. and Perrotta A. *Science* **252**, 434-437 (1991).

Bevilacqua P.C., Sugimoto N. and Turner D.H. *Biochemistry* **35,** 648-658 (1996).

Boyle J., Robillard J.T. and Kim S.H. *J. Mol. Biol.* **139**, 601-625 (1980).

Brehm S.L. and Cech T.R. *Biochemistry* **22,** 2390-2397 (1983).

Brion P. and Westhof E. *Ann. Rev. Biophys. Biomol. Struct*. **26**, 113-137 (1997).

Burke J.M., Irvine K.D., Kaneko K.J., Kerker B.J., Oettgen A.B., Tierney W.M.,

Williamson C.L., Zaug A.J. and Cech T.R. *Cell* **45**: 167-176 (1986).

Burke J.M., Belfort M., Cech T.R., Davies R.W. Schweyen R.J., Shaub D.A. and Szostak J.W., Tabak H.F. *Nucl. Acids Res.* **15**, 7217-7221 (1987).

Cao Y. and Woodson S.A. *RNA* **4**, 901-914 (1998).

Caprara M.G., Mohr G. and Lambowitz A.M. *J. Mol. Biol.* **257**, 512-531 (1996a).

Caprara M.G. Lehnert V., Lambowitz A.M. and Westhof E. *Cell* **87**: 1135-1145 (1996b).

Caprara M.G. Myers C.A. and Lambowitz A.M. *J. Mol. Biol.* **308**, 165-190 (2001).

Cate J.H. and Doudna J.A. *Structure* **4**, 1221-1230 (1996).

Cate J.H., Gooding A.R., Podell E., Zhou K., Golden B.L., Kundrot C.E. Cech T.R. and Doudna J.A. *Science* **273**, 1678-1685 (1996).

Cate J.H., Hanna R.L. and Doudna J.A. *Nature Struct. Biol.* **4**, 553-558 (1997).

Cech T.R. and Rio D.C. *Proc. Natl. Acad. Sci.* **76**, 5051-5055 (1979).

Cech T.R., Damberger S.H., and Gutell R.R. *Nat Struct Biol*. **1**, 273-280 (1994).

Cech T.R. *Gene* **73**, 259-271 (1988).

Cech T.R. *Ann. Rev. Biochem.* **59**, 543-568 (1990).

Cech T.R. Cech T.R. *Gene*. **15**, 33-36 (1993).

Chamberlin M.J., Nierman W.C., Wiggs J. and Neff N. *J. Biol. Chem.* **254**, 10061-10069 (1979).

Chevalier B.S. and Stoddard B.L. *Nucleic Acids Res* **29**, 3757-3774 (2001).

Clodi E. Semrad K. and Shroeder R. *EMBO J* **18**, 3776-3782 (1999).

Coetzee T., Herschlag D. and Belfort M. *Genes Dev* **8**, 1575-1588 (1994).

Colmenarajo G. and Tinoco I. Jr *J. Mol. Biol.* **290**, 119-135 (1999).

Couture S., Ellington A.D., Gerber A.S., Cherry J.M., Doudna J.A., Green R., Hanna M., Pace U., Rajagopal J., Szostak J.W. *J. Mol. Biol.* **215**, 345-358 (1990).

Crothers D.M., Cole P.E., Hilbers C.W. and Shulman R.G. *J. Mol. Biol.* **87**, 63-88 (1974).

Davies R.W., Waring R.B., Ray J.A., Brown T.A. and Scazzocchio C. *Nature* **300**, 719-724 (1982).

de la Mata M., Alonso C.R., Kadener S., Fededa J.P., Blaustein M., Pelisch F., Cramer P., Bentley D., Kornblihtt A.R. *Mol Cell*. **12**, 525-532 (2003).

Din N., Engberg J., Kaffenberger W. and Eckert W. *Cell* **18**: 525-532 (1979).

Doherty E.A. and Doudna J.A. *Biochemistry* **36**, 3159-3169 (1997).

Doherty E.A., Herschlag D. and Doudna J.A. *Biochemistry* **38**, 2982-2990 (1999).

Donahue C. P. and Fedor M.J. *RNA* **3**, 961-973 (1997).

Donahue C.P., Yadava R.S, Nesbitt S.M. and Fedor M.J. *J. Mol. Biol.* **295**, 693-707 (2000).

Doudna J.A., Grosshans C., Gooding A. and Kundrot C.E. *Proc. Natl. Acad. Sci.* **90**, 7829-7833 (1993).

Doudna J.A. and Cech T.R. *RNA* **1**, 36-45 (1995).

Downs W.D. and Cech T.R. *RNA* **2**, 718-732 (1996).

Downs W.D. and Cech T.R. *Genes and Dev*. **8**, 1198-1211 (1994).

Draper D.E. *Nature Struct. Biol*. **3**, 397-400 (1996).

Emerick V.L. and Woodson S.A. *Biochemistry* **32**, 14062-14067 (1993).

Emerick V.L, Pan J. and Woodson S.A. *Biochemistry* **35**, 13469-13477 (1996).

Engelhardt M.A., Doherty E.A., Knitt D.S., Doudna J.A. and Herschlag D. *Biochemistry* **39**, 2639-2651 (2000).

Fedor M.J., *Biochem. Soc. Trans*. **30**, 1109-1115 (2001).

French S.L., Osheim Y.N., Cioci F., Nomura M., and Beyer A.L. *Mol. Cell. Biol.* **23**, 1558-1568 (2003).

Friedman D.I. and Court D.L. *Mol. Micro*. **18**, 191-200 (1995).

Gabus C., Derrington E., Leblanc P., Chnaiderman J., Dormont D., Swietnicki W., Morillas M., Surewicz W.K., Marc D., Nandi P. and Darlix J.L. *J. Biol. Chem.* **276**, 19301-19309 (2001).

Geese W.J. and Waring R.B. *J. Mol. Biol.* **308**, 609-622 (2001).

Gerdes K, Gultyaev A.P., Franch T. Pedersen K. and Mikkelsen N.D. *Annu. Rev. Genet*. **31**, 1-31 (1997).

Gesteland R.F. and Atkins J.F. Editors of *The RNA World*, 2nd edit., Cold Spring Harbor, Laboratory Press, Cold Spring Harbor, NY.

Gesteland R.F., Cech T.R. and Atkins J.F. Editors of *The RNA World*, 2nd edit., Cold Spring Harbor, Laboratory Press, Cold Spring Harbor, NY.

Golden B.L and Cech T.R. *Biochemistry* **35**, 3754-3763 (1996).

Golden B.L., Gooding A.R., Podell E.R. and Cech T.R. *Science* **282**, 259-264 (1998).

Golomb M. and Chamberlin M. *J. Biol. Chem.* **249**, 2855-2863 (1974).

Gonzalez R.L. and Tinoco I. Jr *J. Mol. Biol.* **289**, 1267-1282 (1999).

Grabowski P.J., Zaug A.J. and Cech T.R. *Cell* **23**: 467-476 (1981)

Gultyaev A.P., Franch T. and Gerdes K. *J. Mol. Biol.* **273**, 26-37 (1997).

Heilman-Miller S.L. and Woodson S.A. *J. Mol. Biol.* **306**, 1157-1166 (2001).

Heilman-Miller S.L. and Woodson S.A. *RNA* **9**, 722-733 (2003).

Heilman-Miller S.L. and Woodson S.A. *J. Mol. Biol.* **328**, 385-394 (2003b).

Herschlag D. and Cech T.R. *Biochemistry* **29**, 10159-10171 (1990).

Herschlag D. and Khosla M. *Biochemistry* **33**, 5291-5297 (1994).

Herschlag D., Khosla M., Tsuchihashi Z. and Karpel R.L. *EMBO J* **13**, 2913-2924 (1994).

Herschlag D. *J. Biol. Chem.* **270**, 20871-20874 (1995).

Heuer T.S., Chandry P.S., Belfort M., Celander D.W. and Cech T.R. *Proc. Natl. Acad. Sci.* **88**, 11105-11109 (1991).

Higgins C.F., Causton H.C., Dance G.S.C. and Mudd E.A. In Belasco JG, Brawerman G, eds. *Control of messenger RNA stability*. New York: Academic Press; 13-30 (1993).

Ho Y. and Waring R.B. *J. Mol. Biol.* **292**, 987-1001 (1999).

Ho Y., Kim S.J. and Waring R.B. *Proc. Natl. Acad. Sci.* **94**, 8994-8999 (1997)

Howe K.J., *Biochim. Biophy. Acta* **1577**, 308-324 (2002).

Howe K.J., Kane C.M., Ares M. Jr *RNA* **9**, 993-1006 (2003).

Jaeger J.A., Turner D.H. and Zuker M. *Methods Enzymol*. **180**, 212-226 (1990)

Joyce G.F., van der Horst G. and Inoue T. *Nucl. Acids Res.* **17**, 7879-7889 (1989).

Jurica M.S. and Stoddard B.L. *Cell Mol Life Sci* **55**, 1304-1326 (1999).

Kadesch T.R. and Chamberlin M. J. *J. Biol. Chem.* **257**, 5286-5295 (1982).

Keift J.S. and Tinoco I. Jr *Structure* **5**, 713-721 (1997).

Kim S.H. and Cech T.R. *Proc. Natl. Acad. Sci.* **84**, 8788-8792 (1987).

Kjems J., Olesen S.O. and Garrett R.A. *Biochemistry* **24**, 241-250 (1985).

Kramer F.R. and Mills D.R. *Nucl. Acids Res.* **9**, 5109-5124 (1981).

Kruger K., Grabowski P.J., Zaug A.J., Sands J., Gottschling D.E. and Cech T.R. *Cell* **31**: 147-157 (1982).

Lacatena R.M. and Cesareni G. *Nature* **294**, 623-626(1981)

Laggerbauer B., Murphy F.L. and Cech T.R. *EMBO J.* **13**, 2669-2676 (1994).

Latham J.R. and Cech T.R. *Science* **245**, 276-282 (1989).

Lehnert V., Jaeger L, Michel F. and Westhof E. *Chem. Biol.* **3**, 993-1009 (1996).

Lewicki B.T.U., Margus T., Remme J. and Nierhaus K.H. *J. Mol. Biol.* **231**, 581-593 (1993).

Lin J. and Vogt V.M. *Mol Cell Biol*. **18**, 5809-5817 (1998).

Luebeke K.J., Landry S.M. and Tinoco I. Jr *Biochemistry* **36**, 10246-10255 (1997).

Manley J.L. in Hames BD, Higgins SJ, eds. *Transcription and Translation- a practical approach*. Oxford, UK, Washington, DC: IRL Press. 89-129 (1984).

Mannella C.A., Collins R.A., Green M.R. and Lambowitz A.M. *Proc. Natl. Acad. Sci.* **76**, 2635-2639 (1979).

Marzluff W.F. and Huang R.C.C. in Hames BD, Higgins SJ, eds. *Transcription and Translation- a practical approach*. Oxford, UK, Washington, DC: IRL Press. 89-129 (1984).

Masukata H. and Tomizawa J. *Cell* **44**: 125-136 (1986).

McConnell T.S. and Cech T.R. *Biochemistry* **34**, 4056-4067 (1994).

McGraw P. and Tzagoloff A. *J. Biol. Chem.* **258**, 9459-9468 (1983).

Michel F. and Dujon B. *EMBO J.* **2**, 33-38 (1983).

Michel F., Jacquier A and Dujon B. *Biochimie* **64**, 867-881 (1982).

Michel F., Hanna M., Green R., Bartel D. and Szostak J.W. *Nature* **342**, 391-395(1989)

Michel F., Ellington A.D., Couture S. and Szostak J.W. *Nature* **347**, 578-580 (1990)

Michel F. and Westhof E. *J. Mol. Biol.* **216**, 585-610 (1990).

Misra V.K. and Draper D.E. *Biopolymers* **48**, 113-135 (1998).

Moller-Jensen J., Franch T. and Gerdes K. *J. Biol. Chem.* **276**, 35707-35713 (2001).

Monforte J.A., Kahn J.D. and Hearst J.E. *Biochemistry* **29**, 7882-7890 (1990).

Murphy F.L. and Cech T.R. *Biochemistry* **32**, 5291-5300 (1993).

Nagel J.H., Gultyaev A.P., Gerdes K. and Pleij C.W. *RNA* **5**, 1408-1418 (1999).

Nedbal W., Frey M., Willemann B., Zentgraf H. and Sczakiel G. *J. Mol. Biol.* **266**, 677-687 (1997).

Nikolcheva T and Woodson S.A. *J. Mol. Biol.* **292**, 557-567 (1999).

Nikolcheva T and Woodson S.A. *RNA* **3**, 1016-1027 (1997)

Nogi Y., Yano R. and Nomura M. *Proc. Natl. Acad. Sci.* **88**, 3962-3966 (1991).

Nogi Y., Yano R., Dodd J. Carles C. and Nomura M. *Mol Cell Biol*. **13**, 114-122 (1998).

Noller H.F.. *Ann. Rev. Biochem.* **60**, 191-227 (1991).

Oakes M, Aris J.P., Brockenbrough J.S., Wai H., Vu L. and Nomura M. *J Cell Biol*. **143**, 23-34 (1998).

Pan J, Thirumalai D and Woodson S.A. *J. Mol. Biol.* **273**, 7-13 (1997).

Pan J and Woodson S.A. *J. Mol. Biol.* **280**, 597-609 (1998).

Pan T, Artismovitch I, Fang X.W., Landick R. and Sosnick T.R. *Proc. Natl. Acad. Sci.* **96**, 9545-9550 (1999).

Pan J., Thirumalai D. and Woodson S.A. *Proc. Natl. Acad. Sci.* **96**, 6149-6154 (1999).

Pan J., Deras M.L. and Woodson S.A. *J. Mol. Biol.* **296**, 133-144 (2000).

Pan T. and Sosnick T.R. *Nature Struct. Biol.* **4**, 931-938 (1997).

Peyman A. *Nucleic Acids Res.* **22**, 1383-1388 (1994).

Polisky B., Zhang X-Y. and Fitzwater T. *EMBO J.* **9**, 295-304 (1990).

Price J.V. and Cech T.R. *Science* **228**, 719-722 (1985).

Price J.V., Engberg J. and Cech T.R. *J. Mol. Biol.* **196**, 49-60 (1987).

Pyle A.M., Murphy F.L. and Cech T.R. *Nature* **358**, 123-128(1992)

Pyle A.M. *Science* **261**, 709-714 (1993).

Regnier P. and Arraiano C.M. *BioEssays* **22**, 235-244 (2000).

Riesner D. in *The Viroids* (Diener, T.O., Ed.) pp 63-98, Plenum, New York.

Roberts G.C., Gooding C., Mak H.Y., Proudfoot N.J. and Smith C.W.J. *Nucl. Acids Res.* **26**, 5568-5572 (1998).

Rook M.S., Treiber D.K. and Williamson J.R. *J. Mol. Biol.* **281**, 609-620 (1998).

Rook M.S. Treiber D.K. and Williamson J.R. *Proc. Natl. Acad. Sci.* **96**, 12471-12476 (1999).

Saldanha R.J., Patel S.S., Surendran R., Lee J.C. and Lambowitz A.M. *Biochemistry* **34**, 1275-1287 (1995).

Saldanha R.J., Ellington A. and Lambowitz A.M. *J. Mol. Biol.* **261**, 23-42 (1998).

Schroeder R., Grossberger R., Pichler A. and Waldsich C. *Curr. Opin. Struct. Biol.* **12**, 296-300 (2002).

Sclavi B., Woodson S.A., Sullivan M., Chance M.R. and Brenowitz M. *J. Mol. Biol.* **266**, 144-159 (1997).

Sclavi B. Sullivan M., Chance M.R., Brenowitz, M. and Woodson S.A. *Science* **279**, 1940-1943 (1998).

Semrad K. and Schroeder R. *Genes Dev* **12**, 1327-1337 (1998).

Shaw P.J. and Jordan E.G. *Annu. Rev. Cell Dev. Biol*. **11**, 93-121 (1995).

Shaw L.C. and Lewin A.S. *J. Biol. Chem.* **270**, 21552-21562 (1995).

Smith C.M. and Steitz J.A. *Cell* **89**, 669-672 (1997).

Solem A., Chatterjee P. and Caprara M.G. *RNA* **8**, 412-425 (2002).

Stein A. and Crothers D.M. *Biochemistry* **15**, 160-168 (1976).

Strobel S.A., Cech T.R., Usman N. and Beigelman L. *Science* **267**, 675-678 (1995).

Strobel S.A. and Shetty K. *Proc. Natl. Acad. Sci.* **94**, 2903-2908 (1997).

Thirumalai D. and Woodson S.A. *Acc. Chem. Res.* **29**, 433-439 (1996).

Tinoco I. Jr and Bustamante C. *J. Mol. Biol.* **293**, 271-281 (1999).

Tomizawa J. and Itoh T. *Proc. Natl. Acad. Sci.* **78**, 6096-6100 (1981).

Treiber D.K., Rook M.S., Zarrinkar P.P. and Williamson J.R. *Science* **279**, 1943-1946 (1998).

Treiber D.K. and Williamson J.R. *Curr. Opin. Struct. Biol*. **9**, 339-345 (1999).

Tsuchihashi Z., Khosla M. and Herschlag D. *Science* **262**, 99-102 (1993).

van der Horst G., Christian A. and Inoue T. *Proc. Natl. Acad. Sci.* **88**, 184-188 (1991).

Waldisch C., Masquida B., Westhof E. and Schroeder R. *EMBO J.* **21**, 5281-5291 (2002).

Wallwebber G.J., Mohr S, Rennard R., Caprara M.G. and Lambowitz A.M. *RNA* **2**, 114-132 (1997).

Wang M.D., Schnitzer M.J., Yin H., Landick R., Gelles J. and Block S.M. *Science* **282**, 902-907 (1998).

Ward D.F. and Gottesman M.E. *Nature* **292**, 212-215 (1981).

Waring R.B., Ray J.A., Edwards S.W., Scazzocchio, C. and Davies R.W. *Cell* **40**: 371-380 (1985).

Waring R.B., Towner P., Minter S.J. and Davies R.W. *Nature* **321**, 133-139 (1986).

Weeks K.M. and Cech T.R. *Biochemistry* **34**, 7728-7738 (1995a).

Weeks K.M. and Cech T.R. *Cell* **82**: 221-230 (1995b).

Weeks K.M. and Cech T.R. *Science* **271**, 345-348 (1996).

Weeks K.M. *Curr Opin Struct Biol* **7**, 336-342 (1997).

Wild M.A. and Gall J.G. *Cell* **16**: 565-573 (1979).

Wong E.M. and Polisky B. *Cell* **42**: 959-956 (1985).

Woodson S.A and Cech T.R. *Biochemistry* **30**, 2042-2050 (1991).

Woodson S.A. *Nucl.Acids Res.* **20**, 4027-4032 (1992).

Woodson S.A. *Nat. Struct. Biol.* **7**, 349-352 (2000).

Woodson S.A. and Emerick V.L. *Mol. And Cell. Biol.* **13**, 1137-1145 (1993).

Wu M., Tinoco I. Jr *Proc. Natl. Acad. Sci.* **95**, 11555-11560 (1998).

Yadava R.S., Choi A.J., Lebruska L.L. and Fedor M.J. *J. Mol. Biol.* **309**, 893-902 (2001).

Zarrinkar P.P. and Williamson J.R. *Science* **265**, 918-924 (1994).

Zarrinkar P.P. and Williamson J.R. *Nature Struct. Biol.* **3**, 432-438 (1996).

Zarrinkar P.P., Wang J and Williamson J.R. *RNA.* **2**, 564-573 (1996).

Zaug A.J. and Cech T.R. *Cell* **19**: 331-338 (1980).

Zaug A.J. and Cech T.R. *Nucleic Acids Res.* **10**, 2823-2838 (1982).

Zaug A.J., Grabowski P.J. and Cech T.R. *Nature* **301**, 578-583 (1983).

Zhang F, Ramsay E.S. and Woodson S.A. *RNA* **1**, 284-292 (1995).

Zhang A., Derbyshire V., Salvo J.L. and Belfort M. *RNA* **1**, 783-793 (1995b).

Zhang A., Rimsky S., Raeban M.E., Buc H. and Belford M. *EMBO J.* **15**, 1340-1349 (1996).

Zhang A. Wasserman K.M., Ortega J. Steven A.C. and Storz G. *Mol. Cell* **9**, 11-22 (2002).

Zuker M. *Science* **244**, 48-52 (1989)