Mol. Cells *27*, 657-665, June 30, 2009
DOI/10.1007/s10059-009-0087-y

Molecules and Cells

Thermodynamic Analyses of the Constitutive Splicing Pathway for Ovomucoid Pre-mRNA

Tae Suk Ro-Choi[1,2,3,*], and Yong Chun Choi[1,2,3,*]

The ovomucoid pre-mRNA has been folded into mini-hairpins adaptable for the RNA recognition motif (RRM) protein binding. The number of mini-hairpins were 372 for pre-mRNA and 83-86 for mature mRNA. The spatial arrangements are, in average, 16 nucleotides per mini-hairpin which includes 7 nt in the stem, 5.6 nt in the loop and 3.7 nt in the inter-hairpin spacer. The constitutive splicing system of ovomucoid-pre-mRNA is characterized by preferred order of intron removal of 5/6 > 7/4 > 2/1 > 3. The 5′ splice sites (5′SS), branch point sequences (BPS) and 3′ splice sites (3′SS) were identified and free energies involved have been estimated in 7 splice sites. Thermodynamic barriers for splice sites from the least (| lowest | -Kcal) were 5, 4, 7, 6, 2, 1, and 3; i.e., -18.7 Kcal, -20.2 Kcal, -21.0 Kcal, -24.0 Kcal, -25.4 Kcal, -26.4 Kcal and -28.2 Kcal respectively. These are parallel to the kinetic data of splicing order reported in the literature. As a result, the preferred order of intron removals can be described by a consideration of free energy changes involved in the spliceosomal assembly pathway. This finding is consistent with the validity of hnRNP formation mechanisms in previous reports.

INTRODUCTION

Thermodynamic analysis are needed to explain the mechanism of spliceosomal assembly and pre-mRNA processing. The mechanism of preferred order of intron removals is an unexplored area for multi-intronic pre-mRNA. Introns are removed by co-transcriptional and post transcriptional mechanisms (Beyer and Osheim, 1990; Skoglund et al., 1983). In the case of chicken ovomucoid pre-mRNA, which is composed of eight exons and seven introns, the splicing occur post-transcriptionally. It was demonstrated that the preferred order of intron removal is manifested in the order of the introns 5/6 > 7/4 > 2 > 3/1 (Lewin, 1994) or 5/6 > 7/4 > 2/1 > 3 and sometimes introns 4 or 7 is lost last (Lewin, 2008). This observation suggests that the post-transcriptional preferred order of splicing may results from differential affinities of introns for spliceosome components during the assembly of a single or multiple spliceosomes. Also, the observation suggests that the assembly affinity is closely related to the thermodynamic property of secondary structures of each intron in concert with the regulatory factors involved in the cascades of spliceosomal assembly.

In the present manuscript, an attempt is made to study the preferred order of splicing using thermodynamic approach, although the present approach may be an approximation due to the lack of an accurate database for the secondary structure motifs, especially those involved with RNA binding proteins. However, the study is feasible because basic information is amply available for quantification of thermodynamic parameters. The particle configuration of the ovomucoid pre-mRNP can be visualized in the stabilized form of a protein bound mini-hairpin loops. This model (Ro-Choi and Choi, 2003; 2007) fits with the fundamental structures of 40S pre-mRNP (hnRNP) particles which form the beads (Beyer and Osheim, 1990; Samarina and Krichevskaya, 1981) of the entire ovomucoid pre-mRNP.

When a spliceosome is assembled on the intronic domains of hnRNP, there are cascades of recognition by various snRNPs including U1 snRNP, U2 snRNP and U5·U4/U6 tri-snRNPs. The underlying mechanism of the recognition is RNA-RNA interactions which proceed through cycles of winding (base pairings)/unwinding (Madhani and Guthrie, 1994) leading to construction of a catalytic core. These steps of rearranging the RNA secondary structures are translated into quantifying terms using thermodynamic parameters, especially free energy calculations. By calculation of the free energies involved in spliceosomal assembly steps occurring at three substrate sites (5′SS, BPS, 3′SS), it was possible to find the energy barriers at the rate determining steps at these sites. Comparison of these results with the experimental preferred order of intron removals of 5/6 > 7/4 > 2/1 > 3, it became clear that there is good agreement with the 5 > 7/4 > 2/1 > 3 at the 5′SS hnRNA stem loop structures and a discrepancy of 6 between the two systems. In order to amend this discrepancy, it was recognized that the rate determining steps for introns 5/6 may be modified by the enhancer regulators lowering the higher energy barriers leading to the first priority order of splicing.

MATERIALS AND METHODS

The sequence of chicken ovomucoid gene was obtained with

[1]Medical Research Center for Cancer Molecular Therapy, Dong-A University College of Medicine, Busan 602-714, Korea, [2]Department of Pharmacology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA, [3]Present address: Retirees at 4147 Martinshire Dr., Houston, Texas 77025, USA
*Correspondence: tsrochoi@uspos.net (TSR); ycchoi@uspos.net (YCC)

Received January 28, 2009; revised March 19, 2009; accepted April 22, 2009; published online June 12, 2009

Keywords: co-transcriptional folding, Mini-hairpins, snRNAs; U1, U2, U4, U5, U6, splicing, thermodynamics

the help of Emily Dimmer (Ensemble release 43, http://www.ensemble.org) for 5,587 nt which was subsequently supplemented by additional initiation sites, 85 nt upstream of major initiation site and two poly-A signals in 3′UTR (Gerlinger et al., 1982), totaling 6,067 nt. This includes 138 nt 5′UTR, 633 nt coding sequences, 4,643 nt intronic sequences and 653 nt 3′UTR. The snRNA sequences include U1 snRNA (Branlant et al., 1980; Ro-Choi, 1999; Ro-Choi and Henning, 1977; Ro-Choi et al., 1974), U2 snRNA (Korf and Stumph, 1986; Ro-Choi, 1999) and U6 snRNA (Epstein et al., 1980). The secondary structures of the entire pre-mRNA are constructed in the form of mini-hairpin loops of thermodynamically unfavorable conformations which are stabilized by each hnRNP protein binding. The hnRNA folding in to mini-hairpins are carried out by the methods of Kim and Han (1995) and edited manually (Ro-Choi and Choi, 2003; 2007).

It is presumed that the stable protein bound hairpin loops (Fig. 1) results from counterbalancing effects on the free energy changes (approximately 5 to 8 Kcal) for initiation of loops (von Heijine, 1987; Xia et al., 2001), which in turn stabilizes the stem by closing base pairs.

For the single strand binding protein, it is generally accepted to have a free energy changes of $-\Delta G$ = -10 to -13 Kcal/mol attributable by protein binding (Draper, 1995), although all the dissociation constants are not widely determined for interactions to hairpin loops of hnRNP. Of these, approximately 5-8 Kcal/mol are needed to overcome the destabilizing loop energy and the remainder is attributed for the conformational changes in molecular reorganization (Schulz and Schirmer, 1985).

The duplex formation of U1 snRNA/5'SS (Fig. 2), U2 snRNA/BPS portion (Fig. 3) and U6 snRNA/5′SS (Fig. 2) are well established.

The free energy calculations were made by the INN-HB (Individual Neighbor-Hydrogen Bonding) model (Freier et al., 1986; Xia et al., 2001) and others (von Heijine, 1987) for Watson-Crick helical regions. The protein bound mini-hairpin loops were calculated by exclusion of initiation of loops. The branch point sequences were screened by ESE finder 3.0 (Cartegni et al., 2003) and a sequence was selected with the highest score within 100 nucleotides from the 3′ splice sites (Fig. 3). Enhancer elements (SF2/ASF, SC35, Srp40 and SRp55) were also screened through the entire sequence including exons and introns by ESE finder 3.0.

RESULTS

Folding pattern

The computer-generated folding patterns edited manually are consistent with previous reports for hnRNP formation (Ro-Choi and Choi, 2003; 2007). The ovomucoid gene transcript (pre-mRNA) of chain length 6,067 nucleotides can be folded into 372 protein-bound mini hairpins. The classical processing of the ovomucoid pre-mRNA (6,067 nt) into its mRNA (1,424 nt) produced 83-86 protein bound mini hairpin loops. The unit RNP structure is one hairpin loop/RRM protein (5 nm fiber), which forms an intermediate aggregate (Tetramer, 10 nm fiber), is assembled into a 40S hnRNP (20-24 nm particle containing 48 protein molecules and 700 $\pm$ 20 nt RNA) and finally matures into the bead form of hnRNP. The average nucleotide number involved in the above spatial arrangements is 16 nt/hairpin loop, including 7.0 nt/stem, 5.6 nt/loop, and 3.7 nt/inter-hairpin spacer. The frequency of mini-hairpins in each exon and intron is summarized in Table 1. There were no differences in the frequency of mini-hair pins between exons and introns, although the distribution is more even in exons than in introns. All of the intronic substrate sites (5′SS, BPS, 3′SS) are located on one or two hairpin loop structures (Fig. 1; Table 2). Therefore, the thermodynamic system of RNA processing commences with the hairpin loop RNP as the initial state rather than naked RNA. The structural maintenance is contributed on loop [2-8 nucleotides binding per RRM (Auweter et al., 2006), 5 to 8 Kcal/mol free energy changes] by a single strand binding protein which in turn stabilizes the stem by closing base pairs. Although all of the dissociation constants are not widely determined for interactions to hairpin loops of hnRNP, it is generally accepted that the total free energy changes ($-\Delta G$ = -10 to -13 Kcal/mol) are assigned to the protein binding to RNA (Draper, 1995). Of these, approximately 5-8 Kcal/mol overcome the destabilizing loop ($-\Delta G$) and the remainder is due to conformational changes in molecular reorganization (Schulz and Schirmer, 1985).

Spliceosome assembly

The 5′ splice site shows fast primary binding (protein-hnRNA) interactions and RNA-RNA interactions (U1RNA-hnRNA) which is a much slower interaction involved through formation of nuclei and double helices, suggesting the presence of an energy barrier by the secondary structure of hnRNA (Buratti and Baralle, 2004; Goguel et al., 1993; Jamison et al., 1995). The 3′ splice site is bound by solely protein-RNA interactions as primary binding product (Shahied-Milam et al., 1998).

To assemble a spliceosomal catalytic core, the three substrate sites (5′SS, BPS, 3′SS), existing in the form of mini-hairpin loop-protein complexes, are designed to be unwound and next recognized through various RNA-RNA interactions by U1 snRNA, U2 snRNA, U6 snRNA, and U5 snRNA respectively. These recognition processes are metabolically integrated into the spliceosomal assembly pathway branched out into three recognition pathways which converge into the final common pathway.

(1) The 5′SS recognition pathway involves:

5′SS hairpin loop-protein complex $\xrightarrow{\text{unwinding}}$
5′SS free strand $\xrightarrow{\text{U1 snRNA}}$
5′SS/U1 snRNA duplex $\xrightarrow{\text{unwinding}}$ 5′SS free strand
$\xrightarrow{\text{U5·U4/U6 snRNA}}$ 5′SS/U5/U6 snRNA duplex.

In the above recognitions, the number of base pairs involved are: 3-11 for unwinding 5′SS RNP complex, 7-10 for 5′SS/U1 snRNA duplex formation and 3-4 for 5′SS/U6 snRNA duplex formation (Fig. 2).

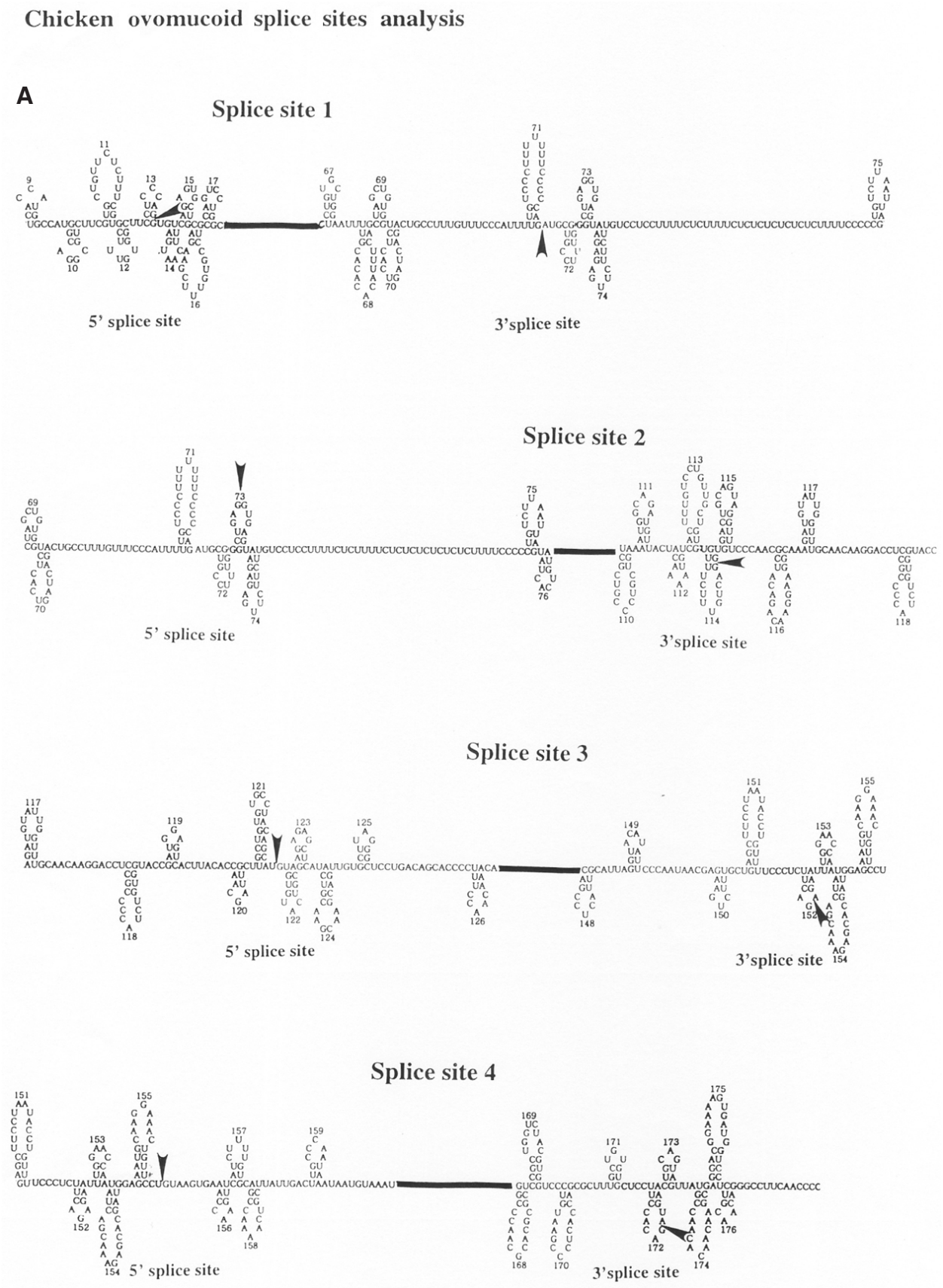
(2) The BPS recognition pathway involves:

BPS hairpin loop-protein complex $\xrightarrow{\text{unwinding}}$ BPS free strand
$\xrightarrow{\text{U2 snRNA}}$ consensus BPS/U2 snRNA complex formation
$\longrightarrow$ expanded non-Watson-Crick BPS/U2 snRNA duplex formation (Fig. 3).

In the above recognitions, the numbers of base pairs involved are 3-6 for unwinding BPS complex and 3-6 for BPS/U2 snRNA duplex formation.

(3) The 3′SS recognition pathway involves:

3′SS hairpin loop-protein complex $\xrightarrow{\text{unwinding}}$ 3′SS free strand
$\xrightarrow{\text{U2 snRNA}}$ inferred 3′SS/U2 snRNA duplex formation.



(Fig. 1, continued)

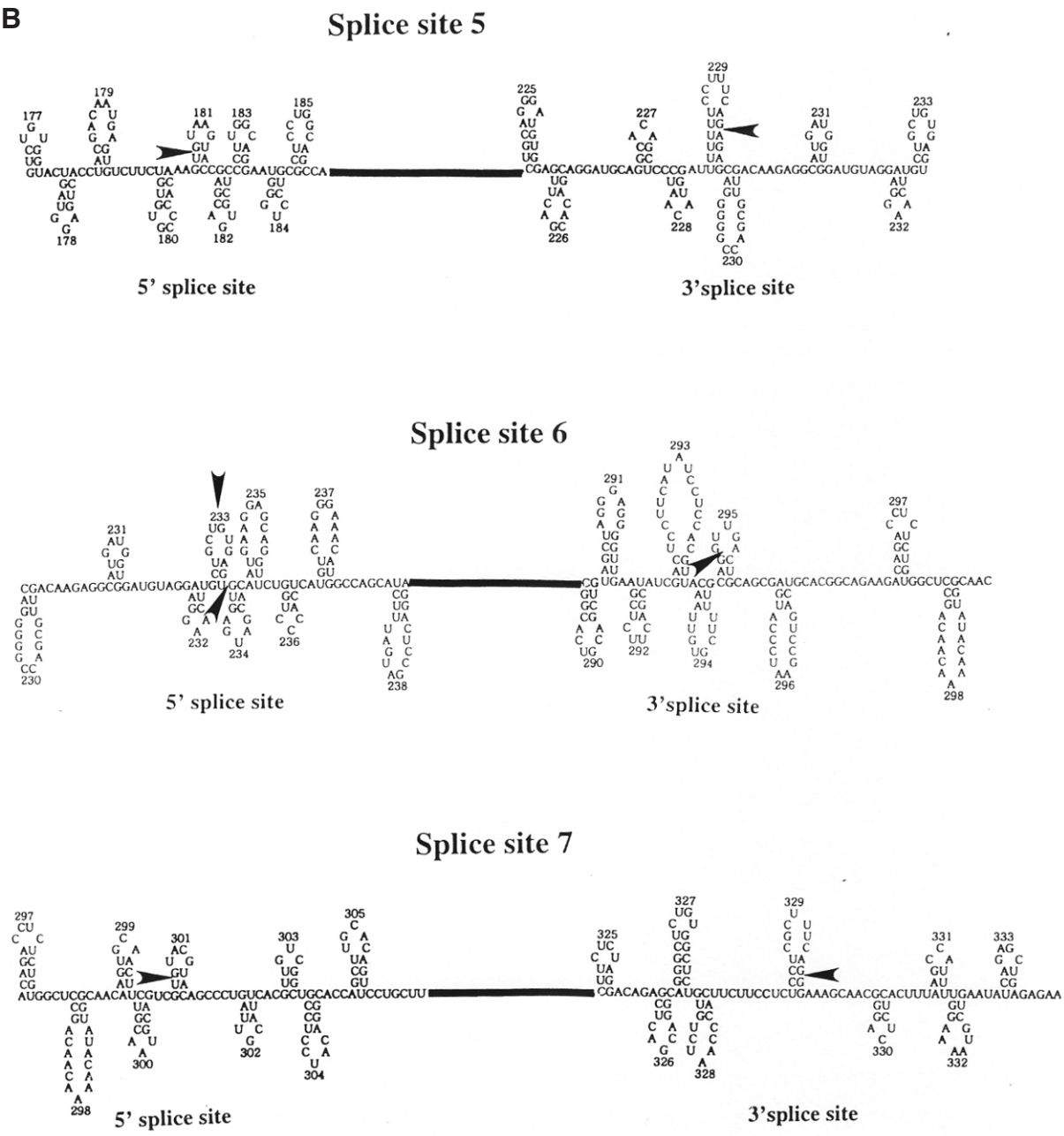


Fig. 1. The ovomucoid pre-mRNA mini-hairpins were constructed using the method of Kim and Han (1995) and manually edited Ro-Choi and Choi (2003; 2007). Each mini-hairpin (a unit of stem-loop) constituted from the first base paired to the last inter-space nucleotide (Ro-Choi and Choi, 2003). Those splice sites (5′SS and 3′SS) containing mini-hairpins and surrounding mini-hairpins (4 mini-hairpins upstream and 4 mini-hairpins down stream) were included in the illustration. The splice sites are marked by arrow heads. The numbers at the stem-loops are the numbers of stem-loops in entire ovomucoid pre-mRNA which has 372 mini-hairpins. The involved hair-pins for the calculation of free energies in Table 2 are consistent with the numbers in this figure.

In the above recognitions, the number of base pairs involved are 3-6 for unwinding 3′SS complex and 3 for inferred 3′SS/U2 snRNA duplex formation.

The converging common pathway involves the construction of a multi-branch loop resulting from duplex formation between the U2 and U6 snRNAs, which are loaded onto substrate sites (5′SS, BPS, 3′SS) and U5 snRNA.

To examine the above three recognition pathways of each intron, computational analyses are necessary to calculate free energy changes, location of rate determining steps and their values of energy barriers.

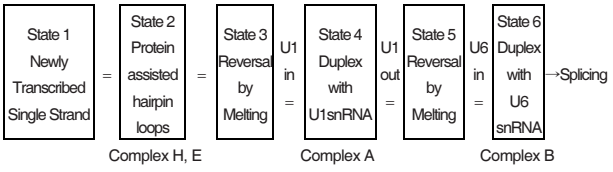
As mentioned above, it is assumed in the present work, that the free energy changes are workable estimates for the protein assisted hairpin loop. Among the stages of complex H, E, A, B1, B2, C1, and C2, the stages of complex H, E are the stage of ATP independent pre-catalytic spliceosomal assembly, which proceeds as a reversible and pre-commitment step. After this

Table 1. Summary of stem-loop analysis

Exons			Junctions			Introns		
(Nt)	S-L	Frq	(Nt)	S-L	Frq	(Nt)	S-L	Frq
1 (181)	12	15.1	186	12	15.5	1 (937)	58	16.2
2 (11)	1	11.0				2 (700)	40	17.5
3 (116)	6	19.3				3 (496)	30	16.5
4 (32)	2	16.0				4 (249)	16	15.6
5 (130)	8	16.3				5 (717)	46	15.5
6 (81)	5	16.2				6 (1,031)	61	16.9
7 (95)	5	19.0				7 (409)	27	15.2
8 (689)	43	16.0						
Total (1,335)	82	16.3				Total (4,539)	278	16.3

Total nt = 1,335 +186 +4,539+7 (5′ Overhang) = 6,067
Minimal contraction rate; 3.9, maximal contraction rate; 8.11
The number of stem-loops (S-L) in each exons, junctions and introns were manually counted.
The frequncy (Frq) were calculated dividing total number of nucleotides by number of stem-loops.
The number of nucleotides was 1,335 in pure exons, 186 in the junctions and 4,539 in pure introns. There were 7 nucleotides at the 5′ end which is not in stem-loop structure.
There were no difference in frequency between exons and introns although there were more uneven distributions in the introns. The frequency of 16.3 are consistent with other hnRNA folding patterns (Insulin gene hnRNA, 25-hydroxyvitamin D3 1-α-hydroxylase gene hnRNA and FMR1 gene hnRNA, (2,3).
The contraction rates of 3.9 for minimal and 8.11 for maximal are also in same range with the previous findings.

stage, U1 snRNA forms a duplex with the 5′-splice site by covering 7-11 nucleotides around the GU site (Freund et al., 2003). Subsequently, U1 snRNA is released and replaced by U6 snRNA. These time-dependent transitions at the 5′ splice sites are described in the following diagram.



At the transition state 2 to state 3 (hairpin to single strand), one or two hairpin loops are converging to yield the ensuing state 4 (U1 duplex) (Fig. 1).

The calculation of free energy changes for the formation of protein assisted hairpin loops is described in Table 2. The values of free energy changes involved in the protein assisted hairpin loops localized at 5′ splice sites 1-7 show a range of -ΔG = -2.5 to -16.7 Kcal/mol. Among all the sites, stem loops at the 5′SS in splice site 3 is the most stable (-16.7 Kcal/mol) and the only stem loop (stem-loop 121) that may be maintained without bound proteins. A comparison of -ΔG′s among the protein assisted hairpin loops at 5′SS show an increasing order of 5,7,4, 1,6,2,3 in their stability. If an allowance of 1-2 Kcal/mol (20% deviation) is given for the calculation of protein assisted hairpin loops, all the 5' splice sites (hairpin loops) except the splice site 3 show lower free energy changes (-ΔG) than those involved in U1-duplex formation (stage 4) (Table 2).

The stage 3 is a transition state between protein assisted hairpin loop state (state 2) and U1-5′ splice site duplex state (state 4). The state 4 is an intermediate process which serves as an energy barrier for transition to state 5 (Freund et al., 2005; Liu, 2002; Staley and Guthrie, 1999). From state 3 to state 4, where the hairpin loop structures anneal to the 5′ end portion of U1 snRNA, 7 canonical base pairs and, some times, extended complementarity into upstream exonic region (1-4 nt), up to 11 base pairs are involved (Freund, 2003; Lund et al., 2002). Table 2 column 3 shows the free energy changes with U1 snRNA at splice sites 1-7 annealed. These show a considerable range of free energy changes; i.e., -ΔG = -6 to -12 Kcal/mol. A close inspection of the free energy changes (U1 duplexes) shows two distinct ranking interval value; those above -10 Kcal/mol (splice sites 3, 4, 5, 6, and 7) and those below -10 Kcal/mol (splice sites 1 and 2), suggesting a clustering similar to those demonstrated in earlier works. Furthermore, when the time course dependent free energy changes, both in the state 2 (protein assisted hairpin loops) and state 4 (U1 snRNA duplex with 5′ splice sites), are correlated together for splicing efficiency, it is clearly demonstrated that the mechanism of rate determining steps is operative for the 5′ splice site dynamics, because the larger the free energy changes, the slower the splicing rate (splice sites 7, 5, 4, and 6 precede the splice sites 1, 2, and 3).

Accordingly, the rate determining steps exist one in state 2 for the 5′ splice site 3 and the others in the state 4 for the splice sites 1, 2, 4, 5, 6, and 7 which are sequestered to contain higher -ΔG′s. The experimental clustering in the form of splicing pathway shows the order of 5/6, 7/4, 2/1 and 3 (Lewin, 2008) or 5/6, 7/4, 2, 3/1 (Lewin, 1994). This clustering is similarly recognized in the thermodynamic order of 5, 4/7, 6, 2/1 and 3 in the combined data sets by two methods (von Heijin, 1987; Xia et al., 2001) presented in Table 2. These are the sum of energy barriers to overcome before the irreversible state of the spliceosomal complex is achieved. Therefore, the thermodynamic prediction is found to be in good agreement with the experimental demonstration except for intron 6 (Lewin, 2008). It is noteworthy that exon 6 has the highest density of SF2/ASF (1 ESE per 13 nt with scores 2.0-2.5) and intron 6 has the highest scoring enhancer elements of SF2/ASF (CACACGU; score 6.2), and SRp55 (UGCAUC; score 5.5). In addition, intron 6 contains a polymorphic site where two 5′SS′s exist with six nucleotides apart (Figs. 1 and 2; Table 2).

Energy dynamics at the level of 5′SS/U1 hybrids

There are conflicting reports on U1 snRNP function in the formation of E and A spliceosomal complex. The 5′SS can be recognized in the absence of U1 5′ oligonuceotides (Lund and

Table 2. Free energy estimates around the splice sites (Kcal/mol).

Splice site	5′SS (pre-mRNA)	5′SS/U1 hybrid separation	BPS (pre-mRNA)	3′SS (pre-mRNA)	Sum of all 4 sites
1	-5.2	-11.8.	-7.6	-3.7	-28.3
	-7.5	-8.7	-5.8	-2.4	-24.4
	-6.4	-10.3	-6.7	-3.1	-26.4
2	-10.0	-11.3	-4.9	-0.8	-27.0
	-9.2	-8.7	-5.2	-0.6	-23.7
	-9.6	-10,0	-5.1	-0.7	-25.4
3	-16.4	-6.6	-2.8	-2.8	-28.6
	-16.7	-5.3	-2.4	-3.3	-27.7
	-16.6	-6.0	-2.6	-3.1	-28.2
4	-3.9	-5.5	-5.4	-3.6	-18.4
	-3.9	-10.6	-4.2	-3.3	-22.0
	-3.9	-8.1	-4.8	-3.5	-20.2
5	-2.5	-6.2	-2.7	-3.8	-15.2
	-2.4	-11.1	-7.2	-1.5	-22.2
	-2.5	-8.7	-5.0	-2.7	-18.7
6	-9.5	-5.9,-7.2	-4.5	-6.6	-26.5, -27.8
	-7.6	-7.2,-8.7	-2.4	-4.2	-21.4, -22.9
	-8.6	-6.6,-8.0	-3.5	-5.4	-24.0, -25.4
7	-2.6	-5.8	-5.4	-6.5	-20.3
	-2.4	-6.8	-4.2	-8.2	-21.6
	-2.5	-6.3	-4.8	-7.4	-21.0

Orders from the least stable (Current data)	Splicing order in the published (Lewin, 2008)
5′SS only, 5, 7, 4, 1, 6, 2, 3 [5, 7/4, (6) 1/2, 3]	5/6, 7/4, 2/1, 3 or 5/6, 7/4, 2, 3/1
All 4 sites	
Combined methods 5, 4, 7, 6, 2, 1, 3 [5, 4/7, 6, 2/1, 3]	

The free energies were calculated from 1-2 mini-hairpins involved in 5′SS (5′ Splice Site), 5′SS/U1 snRNA hybrids (Fig. 2), branch point sequence mini-hairpins and 3'SS mini-hairpins. The free energies were calculated by two different methods, black color (Xia et al., 2001) and red color (von Heijine, 1987). The average of two sets of data were expressed in green color. The branch sites (Fig. 3) were screened within 100 nucleotides from 3′SS (3′ Splice Site) by ESE finder 3.0 (Cartegni et al., 2003) and selected the highest score ones for the calculation of free energies involved.

The stem-loops (see Fig. 1) involved in calculations are

Splice site 1: 5′SS; stem-loops 13, and 14
BPS; stem-loops 69 and 70
3′SS; stem-loop 71

Splice site 2: 5′SS; stem-loops 73 and 74
BPS; stem-loops 110 and 111
3′SS; stem-loop 114

Splice site 3: 5'SS; stem-loops 121 and 122
BPS; stem loop 148
3′SS; stem loop 152

Splice site 4: 5′SS; stem-loop 155
BPS; stem-loop 170
3′SS; stem-loop 172

Splice site 5: 5′SS; stem-loops 181 and 182
BPS; stem-loops 227 and 228
3′SS; stem-loop 229

Splice site 6: 5′SS; stem-loops 233 and 234
BPS; stem-loop 293
3′SS; stem-loop 295

Splice site 7: 5′SS; stem-loop 301
BPS; stem-loop 328
3′SS; stem-loop 329

Kjems, 2002; Rossi et al., 1996) and in the presence of a U1 5′ oligonucleotide, U1 snRNP makes a more specific and stable complex prior to U2 snRNP and U5•U4/U6 tri-snRNP entry into the complex and converts to B and C complexes in an ATP-dependent manner. On the other hand, extended base pair complementarity with hyperstability between U1 snRNA and the 5′ splice site has been reported to become a temperature-sensitive splicing repression in yeast. This step requires a DEAD box helicase protein Prp28p in yeast (Staley and Guthrie, 1999) and p68 RNA helicase in human (Liu, 2002). On the other hand, in mammalian system, increased complementarity increases 5′splice site recognition for the specificity in the 5′SS sites (Freund et al., 2005; Roca et al., 2005). Figure 2 shows the complementarity of 5′SS consensus sequences of chicken ovomucoid with U1 snRNA and U6 snRNA for 7 splice sites. Splice site 1 has 4 nt extended complementarity and splice site 2 has 2 nt extended complementarity which may increase the stability of complexes to -22.5 Kcal/mol and -15.5 Kcal/mol respectively from -8.7 Kcal/mol of canonical 7 nucleotides complementarity. Since the U1 secondary structure has only 9-11 free bases at 5′ end for pairing to complementary sequences, it is difficult to estimate how much these extended complementarity will contribute to the thermodynamics of splicing.

The heterogeneity of 5′SS at splice site 6 provides stability of -7.2 Kcal/mol (short mRNA) and -8.7 Kcal/mol (long mRNA, 2 more amino acids, Val, Ser) and the ratio of products is 25% and 75% (Catterall et al., 1980; Stein et al., 1980). This proportionality is in good agreement with the equation ln3 = 1.1 = K [-8.7-(-7.2)] (Isaacs,1995) which gives K = 0.73 and 73% for the longer one. It appears to produce more final products in stable complex in spite of energy barriers which may slow down the reaction.

DISCUSSION

The present thermodynamic study complements the kinetic study on the preferred order of splicing pathway. Although this study method is an approximation, it is an initial attempt to study splicing thermodynamics of RNA-RNA interactions and protein-RNA interactions. It should be noted that an application of conventional RNA-RNA thermodynamics alone could provide a good but partial agreement. However, more predictable agreement was obtained by including system with duplex altering proteins and stabilizing enhancer regulator proteins. Based upon these combined systems above, it was possible to define the initial state of splicing, known as the pre-spliceosomal transition state, the hairpin loop-protein complexes of hnRNP. As an outgrowth of present work, three new issues arose from the parameter calculations because of some uncertainty in the accuracy of quantitative interpretation. These issues as discussed below should be examined in future.

(1) The pattern of splicing:

Although the splicing system was not characterized, it was assumed that in higher eukaryotes, the system is driven by the exon definition model (Berget, 1995; Brow, 2002). However, it is also possible that both the exon and intron definitions are operative. The patterns of splicing introns for seven introns show that there is one case of a single intron removal (intron 3) and three cases of simultaneous splicing of two introns at a time. These are side by side arrangements (5/6, 2/1), a case with a side distance arrangement (7/4) and a case of those with side by side arrangement loaded with enhancer regulator proteins such as SF2/ASF and SRp55. The candidate introns, spliceable through intron definition mechanism, would be introns 3, 4, and 7. The simultaneous splicing of introns 4 and 7

Fig. 2. The 5′SS and complementarity to U1 5′ end and U6 RNA are illustrated. The chicken U1 5′ end sequence is from Branlant et al. (9) The free energie were calculated by the method of von Heijine (16) for canonical 7 nucleotide complementarity in (-Kcal) and extended complementarity in [-Kcal].

may be a synchronized splicing of introns 4 and 7 independently. The single splicing of intron 3 may also be by the intron definition mechanism.

(2) The free energy contribution of RNA binding to stability of duplex; There are two classes of RNA binding proteins which contribute Tm increasing or Tm decreasing properties to RNA duplexes. These are found in hnRNP and enhancer regulator proteins. The hnRNP proteins are single stranded RNA binding proteins with multiple RRM motifs which provide high affinity structures (~8 to 13 Kcal/mol). They can bind the unstable mini-hairpin loops and stabilize them with input free energy derived from conformational changes, and neutralize the destabilizing initiation free energy of the loop (~5 to 8 Kcal/mol). In place of the general estimate, it is desirable to establish that the thermodynamic stability parameter be available for types of proteins to types of duplex motifs.

(3) The details of the 3′SS recognition pathway have not been established. The importance of this is related to the structure of the catalytic center for the second transesterification. Instead of the mechanistic approach, phylogenic analogy has been taken to visualize the partner RNA molecule of the 3′SS. Using the model of the phylogenic relative, self-splicing group II intron characterized by domains V(d5) and VI(d6) (Brow, 2002; Seetharaman et al., 2008; Toor et al., 2008), it was possible to assign a tentative partner site at the 5′ end stem portion of U2 snRNA duplexed with U6 snRNA. This site is located on one of four foldable arms of the multi-branch loop model. In this foldable model, the 3′SS can be readily accessible to both Mg++ platform and 5′SS for the second transesterification. It is the hope that catalytic center in eukaryotes will be defined more clearly in future.

ACKNOWLEDGMENTS

Authors wish to express their sincere gratitude to Professor Mark O.J. Olson and Professor Lynn C. Yeoman for their contribution to skillful improvement of the manuscript.

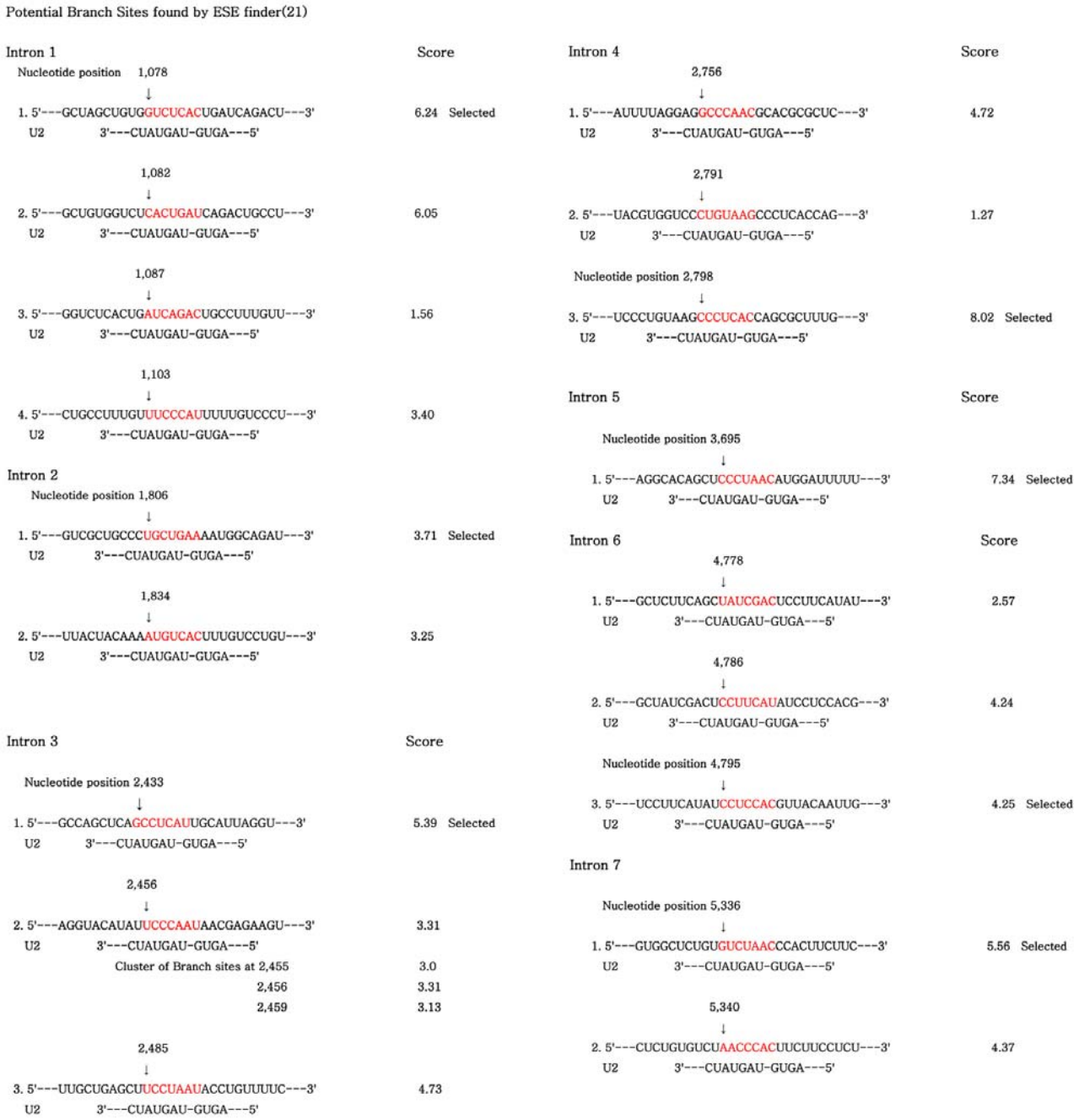


Fig. 3. The potential branch sites were screened within 100 nt from the 3′SS by ESE finder 3.0 (21) and the sites with highest score were selected. The complementarity to U2 branch sites were illustrated and the involved mini-hairpins were selected for the free energy calculations.

REFERENCES

Auweter, S.D., Oberstrass, F.C., and Allain, F.H.-T. (2006). Sequence-specific binding of single-stranded RNA: is there a code for recognition? Nucleic Acids Res. *34*, 4943-4959.

Berget, S.M. (1995). Exon recognition in vertebrate splicing. J. Biol. Chem. *270*, 2411-2414

Beyer, A.L., and Osheim, Y.N. (1990). Ultrastructural analysis of the ribonucleoprotein substrate for pre-mRNA processing. In The Eukaryotic Nucleus; Molecular Biochemistry and Macromolecular Assemblies, Vol. 2, P.R. Strauss, and S.H. Wilson, eds. (The Telford Press, Chapter 18), pp. 431-451.

Branlant, C., Krol, A., Ebel, J.-P., Lazar, E., Gallinaro, H., Jacob, M., Sri-Widada, J., and Jeanteur, P. (1980). Nucleotide sequences of nuclear U1A RNAs from chicken, rat and man. Nucleic Acids. Res. *8*, 4143-4154.

Brow, D.A. (2002). Allosteric cascade of spliceosome activation. Annu. Rev. Genet. *36*, 333-360.

Buratti, E., and Baralle, F.E. (2004). Influence of RNA Secondary Structure on the Pre-mRNA Splicing Process. Mol. Cell. Biol. *24*, 10505-10514.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., and Krainer, A.R. (2003). http://rulai.cshl.edu/tools/ESE.

Catterall, J.F., Stein, J.P., Kristo, P., Means, A.R., and O'Malley, B.W. (1980). Primary sequence of ovomucoid messenger RNA

as determined from cloned complementary DNA. J. Cell Biol. *87*, 480-487.
Draper, D.E. (1995). Protein-RNA recognition. Annu. Rev. Biochem. *64*, 593-620.
Epstein, P., Reddy, R., Henning, D., and Busch, H. (1980). The nucleotide sequence of Nuclear U6 (4.7 S) RNA. J. Biol. Chem. *255*, 8901-8906.
Freier, S., Kierzek, R., Jaeger, J.A., Sugimoto, N., Caruthers, M.H., Neilson, T., and Turner, D.H. (1986). Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA *83*, 9373-9377.
Freund, M., Asang, C., Kammler, S., Konermann, C., Krummheuer, J., Hipp, M., Meyer, I., Gierling, W., Theiss, S., Preuss, T., et al. (2003). A novel approach to describe a U1 snRNA binding site. Nucleic Acids Res. *31*, 6963-6975.
Freund, M., Hicks, M.J., Konermann, C., Otte, M., Hertel, K.J., and Schaal, H. (2005). Extended base pair complementarity between U1 snRNA and the 5′ splice site does not inhibit splicing in higher eukaryotes, but rather increases 5′ splicing site recognition. Nucleic Acids Res. *33*, 5112-5119.
Gerlinger, P., Krust, A., LeMeur, M., Perrin, F., Cochet, M., Gannon, F., Dupret, D., and Chambon, P. (1982). Multiple initiation and polyadenylation sites for the chicken ovomucoid transcription unit. J. Mol. Biol. *162*, 345-364.
Goguel, V., Wang, Y., and Rosbash, M. (1993). Short artificial hairpins sequester splicing signals and inhibit yeast pre-mRNA splicing. Mol. Cell. Biol. *13*, 6841-6848.
Isaacs, N. (1995). Kinetics and thermodynamics. In the book "Physical Organic Chemistry" Chapter 2, (England: Longman Scietific and Technical, Longman House, Burnt Mill, Harlow Essex CM20 2JE), pp. 87-128.
Jamison, S.F., Pasman, Z., Wang, J., Will, C., Lührmann, R., Manley, J.L., and Garcia-Blanco, M.A. (1995). U1 snRNP-ASF/SF2 interaction and 5′ splice site recognition: characterization of required elements. Nucleic Acids Res. *23*, 3260-3267.
Kim, H.-J., and Han, K. (1995). Automated modeling of the RNA folding process. Mol. Cells *5*, 406-412.
Korf, G.M., and Stumph, W.E. (1986). Chicken U2 and U1 RNA genes are found in very different genomic environment but have similar promoter structures. Biochemistry *25*, 2041-2047.
Lewin, B. (1994; 2008). RNA splicing and processing. Genes IX, Chapter 26, (Pearson Prentice Hall, Pearson Education, Inc.), pp. 667-705.
Liu, Z.-R. (2002). p68 RNA Helicase is an essential human splicing factor that acts at the U1 snRNA-5′ splice site duplex. Mol. Cell. Biol. *22*, 5443-5450.
Lund, M., and Kjems, J. (2002). Defining a 5′ splice site by functional selection in the presence and absence of U1 snRNA 5′ end. RNA *8*, 166-179.
Madhani, H.D., and Guthrie, C. (1994). Dynamic RNA-RNA interactions in the spliceosome. Ann. Rev. Genet. *28*, 1-26.
Roca, X., Sachidanandam, R., and Krainer, A. (2005). Determinants of the inherent strength of human 5′ splice sites. RNA *11*, 683-698.
Ro-Choi, T.S. (1999). Nuclear snRNA and nuclear function (Discovery of 5′ cap structures in RNA). Crit. Rev. Eukaryot. Gene Expr. *9*, 107-158.
Ro-Choi, T.S., and Choi, Y.C. (2003). Structural elements of dynamic RNA strings. Mol. Cells *16*, 201-210.
Ro-Choi, T.S., and Choi, Y.C. (2007). A modeling study of co-transcriptional metabolism of hnRNP Using FMR1 gene. Mol. Cells *23*, 228-238.
Ro-Choi, T.S., and Henning, D. (1977). Sequence of 5′ oligonucleotide of U1 RNA from Novikoff hepatoma cells. J. Biol. Chem. *252*, 3814-3820.
Ro-Choi, T.S., Reddy, R., Choi, Y.C., Raj, N.B., and Henning, D. (1974). Primary sequence of U1 nuclear RNA and unusual feature of 5′ end structure of LMWN RNA. Fed. Proc. Fed. Am. Soc. Exp. Biol. *33*, 1548.
Rossi, F., Forne, T., Antoine, E., Tazi, J., Brunel, C., and Cathala, G. (1996). Involvement of U1 small nuclear ribonucleoproteins (snRNP). in 5′ splice site-U1 snRNP interaction. J. Biol. Chem. *271*, 23985-23991.
Samarina, O.P., and Krichevskaya, A.A. (1981). Nuclear 30S RNP Particles. In the book "The Cell Nucleus; Nuclear Particles", Part B., H. Busch, ed., (Academic Press, Inc.), pp. 1-48,
Schulz, G.E., and Schirmer, R.H. (1985). Protein-Ligand Interactions. Principles of Protein Structure, Chapter 10, (New York, Berlin, Heidelberg, Tokyo: Springer-Verlag).
Seetharaman, M., Eldho, N.V., Padgett, R.A., and Dayie, K.T. (2008). Structure of a self-splicing group II intron catalytic effector domain 5: Parallels with spliceosomal U6 RNA. RNA *12*, 235-247.
Shahied-Milam, L., Soltaninassab, S.R., Iyer, G.V., and LeStourgeon, W.M. (1998). The heterogenous nuclear ribonucleoprotein C protein tetramer binds U1, U2, and U6 snRNAs through its high affinity RNA bindinf domain (the bZIP-like motif). J. Biol. Chem. *273*, 21359-21367.
Skoglund, U., Andersson, K., Björkroth, B., Lamb, M.M., and Daneholt, B. (1983). Visualization of the formation and transport of a specific hnRNP particle. Cell *34*, 847-855.
Staley, J., and Guthrie, C. (1999). An RNA switch at the 5′ splice site requires ATP and the DEAD box protein Prp28p. Mol. Cell *3*, 55-64.
Stein, J.P., Catterall, J.F., Kristo, P., Means, A.R., and O'Malley, B.W. (1980). Ovomucoid intervening sequences specify functional domains and generate protein polymorphism. Cell *21*, 681-687.
Toor, N., Keating, K.S., Taylor, S.D., and Pyle, A.M. (2008). Crystal structure of a self-spliced group II intron. Science *320*, 77-82.
von Heijine, G. (1987). Sequence analysis in molecular biology, Chapter 4, Nucleotide sequences: what you can do with your sequence once you have it (Academic Press), pp. 19-80.
Xia, T., Mathews, D.H., and Turner, D.H. (2001). Thermodynamics of RNA secondary structure formation. In RNA, Chapter 2, D. Söll, S. Nishimura, and P.B. Moore, eds., pp. 21-48.