Foreword

Chapter 2 is based on the paper submitted to journal of Molecular and Cellular Biology (Chun Chen, Xinliang Zhao, Ryszard Kierzek and Yi-Tao Yu 2010). Dr. Xinliang Zhao contributed to the data for mapping of the important uridines within the adenovirus pre-mRNA (Figure 2.2C). Dr. Doug Turner provided advice on LNA. Dr. Clara Kielkopf generously provided U2AF$^{65}$ antibody.

Dr. Peter Schattner from University of California Santa Cruz provided the candidates for pseudouridylation in yeast U1 snRNA through snoGPS in Chapter 3. Dr. Chunxing Yang purified the Cbf5p-TAP preparations for in vitro TAP pseudouridylation assay.
Chapter 1

Introduction
1.1.1 Pre-mRNA splicing

Pre-mRNA splicing, which was first discovered 30 years ago, is a reaction that entails the precise removal of introns from an mRNA precursor and the joining of exons together to form a functional mature mRNA (Staley and Guthrie 1998; Burge, Tuschl et al. 1999; Yu, Scharl et al. 1999; Jurica and Moore 2003). Introns are removed by two consecutive transesterification reactions that take place in a large RNA-protein complex named the spliceosome, which is composed of numerous protein components and five small nuclear (sn)RNAs (U1, U2, U4, U5 and U6 snRNA), each functioning as a ribonucleoprotein particle (snRNP). These components assemble onto the pre-mRNA, and are released from the mature mRNA after splicing. During the highly ordered and dynamic assembly process, the snRNPs interact with each other and with the consensus sequences located in the intron to construct a functional spliceosome (Green 1991; Staley and Guthrie 1998; Burge, Tuschl et al. 1999; Yu, Scharl et al. 1999; Reed 2000; Jurica and Moore 2003).

1.1.2 Consensus sequence elements at the splice sites in pre-mRNA

Cis-acting signal sequences are required for the precise removal of introns during pre-mRNA splicing. By comparing sequences at the exon/intron junctions, several conserved sequences in intron were identified at the 5’ splice site (5’SS), the branch point region and the 3’ splice site (3’SS) with an adjacent polypyrimidine tract in higher eukaryotes. And in higher eukaryotes, the 5’ splice site consensus sequence is AG/GURAGU, where R is a purine residue, the slash indicates the 5’ splice site, and the underlined dinucleotide GU is invariant (Breathnach and Chambon 1981).
The more limited 3’ splice site is YAG, where Y is a pyrimidine residue, the slash indicates the 3’ splice site and the underlined AG is invariant (Mount 1982). Additionally, there is a more loosely defined branch site consensus sequence, YNYURAC, where N can be any nucleotide and the underlined adenosine is invariant. The consensus is found near the 3’ ends of the introns, with the adenosine residue serving as the nucleophile in the reaction. In most introns, the branch site is usually located 20-40 nucleotides upstream of the 3’ splice site (Burge, Tuschi et al. 1999). In S. cerevisiae, consensus sequences have also been identified. The sequences are more conserved compared to the loosely defined consensus sequences in higher eukaryotic pre-mRNAs. There is G/GUAUGU at the 5’ splice site, CAG/G at the 3’ splice site and UACUAAC at the branch point recognition region. In most cases, mutations within the invariant nucleotides either greatly reduce or abolish pre-mRNA splicing (Aebi, Hornig et al. 1986).

1.1.3 The polypyrimidine tract in pre-mRNA

Close to the 3’ splice site, there is a region called the polypyrimidine tract that promotes assembly of the spliceosome. The polypyrimidine tract is rich with pyrimidine nucleotides, especially uridine. It is usually 15-20 nucleotides long and located about 5-40 nucleotides upstream of the 3’ splice site. A number of protein factors bind to or associate with the polypyrimidine tract, including the essential splicing factor U2 snRNP auxiliary factor (U2AF). U2AF recognizes the consensus 3’ splice site in the pre-mRNA and coordinates the early step of spliceosome assembly. Heterogeneous ribonucleoprotein protein C (hnRNP C), polypyrimidine tract binding protein
(PTB), and intron-binding proteins, are also bound to the polypyrimidine tract (Choi, Grabowski et al. 1986; Bothwell, Ballard et al. 1991; Zamore, Patton et al. 1992). Interestingly, it is apparent that there is great flexibility in the specific sequence of a given tract. The tract is usually long and variable in human introns, yet generally missing in yeast. However, many yeast introns are enriched for uridines adjacent to the 3’ splice site, particularly at the position 9 nucleotides upstream of the 3’ splice site (Parker and Patterson 1987). Many studies have functionally analyzed the polypyrimidine tract during pre-mRNA splicing process and shown that the polypyrimidine tract is essential for efficient early spliceosome assembly and splicing of many introns in higher eukaryotes. The distance between the branch point sequence and the polypyrimidine tract is critical for efficient lariat structure formation and that progressive deletions of the polypyrimidine tract dramatically reduce pre-mRNA splicing efficiency (Reed 1989) (Wieringa, Hofer et al. 1984). Not only can the polypyrimidine tract increase the efficiency of branch point utilization, it can also alter 3’ splice site selection, thus it is probably involved in the regulation of splice site selection in alternative splicing (Sosnowski, Belote et al. 1989). By mutational analysis, Norton found that in some pre-mRNA substrates, introduction of purines into the polypyrimidine tract is detrimental to splicing only when the length of the tract is shortened and when there is a reduced number of consecutive uridine residues (Norton 1994). The effect is more detrimental when the introduction of purines is immediately downstream of the branch point sequence rather than close to the 3’ splice site (Reed 1989). Furthermore, it was reported that uridine and cytidine
function differently within the polypyrimidine tract (Bouck, Fu et al. 1995). In 1997, Patton’s group utilized a cis-competition assay and determined that uridines are the preferred pyrimidine within the polypyrimidine tract, and that the nucleotide composition of the polypyrimidine tract can affect the efficiency of the first step pre-mRNA splicing (Coolidge, Seely et al. 1997).

1.1.4 A two-step trans-esterification pathway of pre-mRNA splicing

Splicing of mRNA precursors was discovered over 20 years ago, but not until the early 1980s was a two consecutive transesterification reaction pathway proposed. The two splicing intermediates, the free 5’ exon intermediate and the lariat intron/3’ exon intermediate were discovered first (Padgett, Konarska et al. 1984) (Ruskin, Krainer et al. 1984), followed by the identification of the lariat structure (Wallace and Edmonds 1983). In 1985, the free 5’ exon intermediate and the lariat intron/3’ exon intermediate were also unveiled in yeast (Lin, Newman et al. 1985). The two-step chemical reaction is catalyzed by the spliceosome (Figure 1.1). In the first transesterification, the 2’hydroxyl group (2-OH) of a particular adenosine (A) residue of the intron (branch point) attacks the phosphate at the 5’ exon/intron junction, resulting in cleavage of the phosphodiester bond between the 5’ exon and intron and subsequent formation of a new 5’-2’ phosphodiester bond between the 5’ end of the intron and the branch point adenosine residue. In the first step reaction, a free 5’ exon and a lariat inton/3’ exon intermediate are produced. In the second transesterification, the 3’-OH group of the 5’ exon attacks the phosphate at the
intron/3’ exon junction to release the lariat intron and generates the spliced mature mRNA by joining the free 5’ and 3’ exons together.

1.1.5 Spliceosome assembly pathway

Excision of the intron from pre-mRNA by the spliceosome is conserved from yeast to humans (Burge, Tuschl et al. 1999). The spliceosome is a large ribonucleoprotein complex, which contains five uridine-rich small nuclear ribonucleoprotein particles (snRNPs). Each snRNP is composed of one short RNA molecule (U1, U2, U4, U5 and U6 snRNA) and a set of Sm or Sm-like proteins with numerous specific proteins (Will and Luhrmann 2001). The interactions of the snRNPs with each other and the intron are highly dynamic and the ordered assembly process is driven by DExD/H-box RNA-dependent ATPases. The sequential spliceosome assembly process, which is also remarkably conserved from yeast to humans, ensures the accuracy of splicing by multiple intron-recognition events. During which complexes E, A, B, and C are formed on the pre-mRNA in humans (Figure 1.1). Formation of a nonspecific H complex initiates the assembly reaction following addition of the pre-mRNA to the splicing extracts (Konarska and Sharp 1987). Association of U1snRNP and various protein factors, including U2 auxiliary factor (U2AF, Mud2p in yeast) and splicing factor 1 (SF1, BBP/Ms15p in yeast), turns the H complex into the early (E) complex in humans or the commitment complex (CC complex) in yeast and commits the intron to be spliced out (Legrain, Seraphin et al. 1988) (Bindereif and Green 1987). U2 snRNP then joins the commitment complex in an ATP-dependent manner, and U2AF associates with the
Figure 1.1 Spliceosome assembly and catalysis of pre-mRNA splicing. The thick lines represent the intron, and the boxes are exons. The 5′ splice site (5′SS), the 3′ splice site (3′SS) with the polypyrimidine tract (Yn) and the branch point adenosine (BP) are indicated in the pre-mRNA. First, U1 snRNA (red) assembles onto the pre-RNA (in the H complex) to form a commitment complex (CC). Then U2 snRNA (green) joins the complex to convert it to complex A. Subsequently, tri-snRNP U4/U5/U6 (blue/brown/purple respectively) forms the major spliceosome (complex B1). The newly assembled spliceosome then undergoes a dynamic rearrangement of RNA-RNA interactions, resulting in the release of U1 and U4 snRNAs and formation of complex B2. Formation of complex B2 induces the first catalytic step of pre-mRNA splicing to form the free 5′ exon and the lariat structure, thus leading to the formation of complex C1. Additional conformational changes taken place, the second step of splicing occurs: joining of 5′ and 3′ exons and formation of complex C2. Mature mRNA is then released, complex I is formed. Finally, the lariat intron is released and the spliceosome is disassembled and recycled.
polypyrimidine tract and 3’ splice site; meanwhile, SF1 associates with the branch region. Stable association of U2 snRNP with the pre-mRNA converts the E complex to the pre-spliceosome (A complex) (Das, Zhou et al. 2000). Following the formation of complex A, the association of the preformed U4/U5/U6 tri-snRNP with the pre-mRNA yields the B1 complex. Conformational changes ensue to convert B1 complex in to the B2 complex. In order to form the catalytically active complex, U4 and U6 snRNAs, which are extensively base-paired in the tri-snRNP, are unwound. U6 displaces U1 snRNA at the 5’ splice site, and U2 and U6 snRNAs form extensive base-pair interactions. The multiple conformational rearrangements ensure that the proper splicing signals are recognized and that splicing occurs accurately. The spliceosome is now catalytically ready for the first step of splicing. Before the second step of splicing can take place, the B2 complex undergoes several rearrangements to form complex C1 (Umen and Guthrie 1995). These changes include conformational rearrangements in U2 snRNA (Hilliker, Mefford et al. 2007), a perturbation in the U2-branch region interaction (Smith, Query et al. 2007) and association of U5 snRNA with the 5’ and 3’ splice sites (Newman 1997). Complex C2, which contains the spliced product, forms after the lariat intron is spliced out. Finally, the mature mRNA is released and C2 complex converts to complex I. After the spliced out lariat intron is released, complex I dissociates into its components, which subsequently enter a new cycle of spliceosome assembly (Moore, Query et al. 1993) (Ruby and Abelson 1991).

1.1.6 The splicing factor U2 snRNP auxiliary factor U2AF
U2 small nuclear ribonucleoprotein auxiliary factor (U2AF) is an essential heterodimeric human splicing factor of two subunits that binds to the polypyrimidine tract, the 3’ splice site and various splicing factors within the machinery (Zamore and Green 1989). The 35 kDa small subunit, U2AF\textsubscript{35} (U2AF1), directly binds to the AG dinucleotide at the 3’ splice site (Wu, Romfo et al. 1999). The 65 kDa large subunit, U2AF\textsubscript{65} (U2AF2), contacts the essential polypyrimidine tract that is dominantly composed of uridines (Zamore, Patton et al. 1992). In both yeast and humans, assembly of U2AF with the pre-mRNA is crucial for recruitment of the U2 snRNP to the spliceosome, which requires an interaction with the U1 snRNP (Cote, Beaudoin et al. 1995) (Kent, Ritchie et al. 2005). U2AF is highly conserved from yeast to humans. Mud2p is the yeast homolog of U2AF\textsubscript{65} (Zamore and Green 1989). The homolog of U2AF\textsubscript{35} has been shown to function in splicing in \textit{S. pombe} (Webb, Lakhe-Reddy et al. 2005), however, there is no apparent U2AF\textsubscript{35} homolog in \textit{S. cerevisiae}, likely because of no requirement of the 3’SS for the first catalytic step of splicing (Rymond and Rosbash 1985) (Zamore and Green 1989). U2AF\textsubscript{65} contains three RNA-recognition motifs (RRMs). RRM1 and RRM2 are the two central motifs and are responsible for the recognition of the polypyrimidine tract in the pre-mRNA. The third motif (UHM) is specialized for protein-protein interactions and is highly conserved from yeast to humans. Interaction of the UHM in U2AF\textsubscript{65} with SF1 strengthens the binding to the polypyrimidine tract. Further strengthening the binding to the polypyrimidine tract requires the association of the N-terminus of U2AF\textsubscript{65} and U2AF\textsubscript{35} (Kielkopf, Lucke et al. 2004). U2AF\textsubscript{65} recognizes a polypyrimidine tract
sequence universally. To better understand the nature of U2AF$^{65}$-RNA binding, using a complementary mutational analysis and X-ray structure determinations, Kielkopf’s group has revealed a network of hydrogen bond interactions between U2AF$^{65}$ and the polypyrimidine tract (Sickmier, Frato et al. 2006).

1.2 Pseudouridine: post-transcriptional modification of cellular RNAs

Pseudouridylation is one of the most abundant post-transcriptional internal modifications found in rRNA, tRNA, snRNA and possibly mRNA (Ofengand and Gournier 1998) (Grosjean, Sprinzl et al. 1995) (Massenent, Mougin et al. 1998) (Yu, Terns et al. 2005). Pseudouridines are also found in U3, U8, snR4 and snR8 snoRNAs (Yu, Terns et al. 2005), although the functional importance of these pseudouridines have not been identified.

Formation of pseudouridines is divided into two classes: RNA-dependent or RNA-independent. With a few exceptions, pseudouridylation is catalyzed RNA-dependently in eukaryotes. Box H/ACA RNPs are responsible for the RNA-dependent pseudouridylation. Each RNP contains a unique box H/ACA RNA and four core proteins (Yu, Terns et al. 2005). The Box H/ACA RNA serves as a guide by base pairing with the target RNA substrate and the proteins facilitate the modifications at the specific site (Kiss 2001). Pseudouridines are usually concentrated in functionally important regions in RNAs and are highly conserved across species (Stephenson, Karijolish et al. 2008). Although their biosynthesis and biological roles have remained poorly understood, studies have shown that at least some pseudouridines in rRNA are necessary for protein synthesis (Piekna-Przybylska,
Przybylski et al. 2008) and some in U2 snRNA are required for pre-mRNA splicing (Zhao and Yu 2004; Zhao and Yu 2007). Pseudouridine has the ability to coordinate a water molecule through its free N1-H, resulting in a rigidifying effect on the nearby sugar–phosphate backbone. Pseudouridine can also enhance local RNA base stacking (Charette and Gray 2000). These effects may explain the functional importance of pseudouridine residues in RNAs.

1.2.1 Pseudouridine in RNA: what and where?

Pseudouridine is the 5’-ribosyl isomer of uridine (Figure 1.2) with its identification as a “fifth nucleotide”. It comprises approximately 4% of the nucleotides in yeast tRNA (Davis and Allen 1957). Formation of pseudouridine is an isomerization reaction. The resulting modified base features the free N1-H in pseudouridine. The free N1-H acts as an additional hydrogen bond donor (Davis 1998), and allows for the participation of pseudouridine in novel pairing interactions in RNAs. The structural properties of pseudouridine enable its effect on enhancing local RNA stacking by favoring a C3’-endo conformation of the ribose (Davis 1998).

Pseudouridine was initially characterized in tRNA, and contributes to stabilization of the specific structural motifs (Charette and Gray 2000). In the ribosomal RNAs, pseudouridines have been mapped to areas of functional importance involved in protein synthesis (Ofengand and Gournier 1998). In higher eukaryotes, pseudouridines are found in the major spliceosomal snRNAs (U1, U2, U4, U5 and U6 snRNA) and the minor vertebrate variants responsible for AU/AC intron splicing (U12, U4atac, and U6atac) (Figure 1.3A). However, only a few pseudouridines are
Figure 1.2 Chemical differences between Uridine (U) and pseudouridine (Ψ). The uracil base in uridine (left) is linked through its N-1 position (arrowhead) to the C-1’ position of the ribose. The base in uridine possesses one hydrogen-bond acceptor and one donor (dashed arrows; a and d, respectively). Isomerization reaction occurs when the uracil base is rotated 180°+ degrees through an N3-C6 diagonal axis (circular arrow). In pseudouridine (right), the C-5 position of uracil (arrowhead) is linked to the C-1’ position of the sugar ring, resulting in an increase in hydrogen bonding capacity (to one acceptor and two donors) compared with that in uridine (modified from (Charette and Gray 2000)).
found in *S. cerevisiae* spliceosome snRNAs, which are located in U1, U2 and U5 snRNAs (Figure 1.3B). Despite the number of pseudouridines found in higher eukaryotes and *S. cerevisiae* are quite different, these pseudouridines are all located in functionally important regions engaged in the assembly and functioning of the spliceosome through inter or intramolecular interactions (Massenent, Mougin et al. 1998). For example, all the pseudouridine residues in the branch site recognition region of U2 snRNA have been verified to be critical for pre-mRNA splicing both in yeast and higher eukaryotes (Zhao and Yu 2004; Yang, McPheeters et al. 2005).

### 1.2.2 Structural properties of pseudouridine

The availability of powerful approaches, such as nuclear magnetic resonance and X-ray crystallography, offered a great opportunity to analyze the structural properties of pseudouridine. Pseudouridine has been found in the *anti* configuration, and confers rigidity rather than flexibility on RNA structure (Davis 1998) (Auffinger 1998). The conclusion is based on one of the special properties of pseudouridine compared to its parental uridine that pseudouridine has an additional potential hydrogen bond donor. This additional hydrogen bond engages pseudouridine to coordinate a water molecule between its N1-H and the 5’ phosphates of both pseudouridine and its preceding residue (Figure 1.4) (Auffinger 1998). Such a water bridge restricts base conformation and mobility of the backbone 5’ to the site of pseudouridylation (Davis 1995). In addition to its effect on RNA structural rigidity, pseudouridine enhances local RNA stacking by favoring a C3’-endo conformation of the ribose (Davis, Veltri et al. 1998). Here the structured water molecule replaces a
Figure 1.3 Pseudouridines of vertebrate major spliceosomal U snRNAs (A) and S. cerevisiae spliceosomal U snRNAs (B). (A) The 5’ caps (2, 2, 7, trimethylated guanosine cap for U1, U2, U4 and U5 snRNA, and γ-methylated guanosine cap for U6 snRNA) are shown. The pseudouridines are boxed. (B) The interaction between the 5’ and 3’ splice sites and the branch site (BS) consensus sequences with U1 and U2 snRNAs are shown in scheme I of panels A and B. UsnRNA-UsnRNA and UsnRNA-pre-mRNA interactions are the catalytic center of the spliceosome as shown in scheme II of panel A and B. The pseudouridine residues are boxed (adopted from Massenet, Motorin et al. 1999).
weak C5-H….Ow (Ow: oxygen atom of the water molecule) interaction in uridine with a stronger N1-H….Ow interaction in pseudouridine (Auffinger 1998; Davis 1998), which reduces conformational flexibility of the pseudouridine moiety to render the phosphodiester backbone in its vicinity more rigid. Consequently, increased stacking of neighboring nucleotides propagates throughout the adjacent region, which greatly contributes pseudouridine on RNA structural stabilization (Davis 1998). For example, a highly conserved pseudouridine in the region of U2 snRNA that pairs with the intron has been reported to induce a modest increase in thermal stability compared with its unmodified uridine analogue (Newby and Greenbaum 2001).

1.2.3 Formation of pseudouridines

Pseudouridine is synthesized by pseudouridine synthase, which catalyzes the isomerization reaction at specific uridine sites in RNAs. To date, it has been known that pseudouridine can be introduced by two different mechanisms: RNA-dependent and RNA-independent.

1.2.3.1 RNA-independent mechanism

In the RNA-independent mechanism, pseudouridine synthase is a single protein or a protein complex. Pseudouridine synthases specifically select uridine residues for isomerization to pseudouridines by standard protein-based recognition of the structural context of the target uridines. This mechanism is widely used in eubacterial rRNA and tRNA. All known pseudouridine synthases can be categorized into five families, TruA, TruB, TruD, RluA and RsuA, which are named after the E. coli enzymes (Gustafsson, Reid et al. 1996; Koonin 1996; Charette and Gray 2000)
Figure 1.4 Three-dimensional model of the hydrogen-bond-mediated water bridge formed by $\Psi$ in RNA (modified from (Charette and Gray 2000). The arrows indicate hydrogen bonding interactions. The coordination of water strongly favors the C3'-endo configuration of the ribose ring.
One of the most significant features of pseudouridine synthases is that they recognize their substrate RNA(s) specifically. Some of the pseudouridine synthases modify a single position in a particular RNA species. For example, RsuA isomerizes U516 in the small-subunit rRNA in *E. coli* exclusively (Conrad, Niu et al. 1999). Some enzymes select their substrates in structurally equivalent positions in multiple related RNAs. TruB, which is responsible for the universally conserved $\psi55$ of the T$\psi$C loop of tRNA, exhibits this type of specificity. Some pseudouridine synthases have dual-specificity to act on two structurally distinct RNA species that yet share local sequence and structural similarity. The *E. coli* pseudouridine synthase RluA is an example to modify 23S rRNA and certain tRNAs. Other pseudouridine synthases, either from the TruA or TruD family, are capable of modifying several nearby sites on a specific RNA. In budding yeast, the nomenclature of pseudouridine synthases have been numbered sequentially, developed independently of the bacterial system. For example, Pus1p and Pus7p are the *S. cerevisiae* orthologs of TruA and TruD from *E. coli*, respectively. Pus1p acts at position 34, 36 in tRN-Ile, position 27 or 28 in several tRNAs (Motorin, Keith et al. 1998). Pus7p is able to isomerize uridine at position 35 (Ma, Zhao et al. 2003), and 93 (unpublished data) in U2 snRNA.

1.2.3.2 RNA-dependent mechanism

Relatively few pseudouridines in eubacterial rRNAs and tRNAs are formed by site-specific pseudouridine synthases. In eukaryotic RNA species, protein-H/ACA RNA complexes termed ribonucleoprotein complexes (H/ACA RNPs) direct site-
specific pseudouridylation. Each H/ACA RNP comprises one small box H/ACA RNA and four core proteins (dyskerin/Cbf5p, Gar1p, Nhp2p/L7aep and Nop10p) (Ni, Tien et al. 1997; Kiss 2001). The H/ACA RNA is involved in substrate recognition through base-pairing interactions with the target RNA, and it also provides a scaffold for assembly of the associated proteins (figure 1.5).

1.2.3.2.1 H/ACA RNAs (guide RNAs)

Each H/ACA RNP is defined by its unique RNA component, the H/ACA RNA. Each H/ACA RNA exhibits a hairpin-hinge-hairpin-tail secondary structure and contains the conserved sequence of ANANNA (box H, the hinge region) and the ACA motif in the tail. The ACA box is located three nucleotides upstream of the 3’ end of the box H/ACA RNA (Kiss 2001). Site selection for pseudouridylation occurs by base-pairing interactions between antisense sequences in the box H/ACA RNA and sequences in the substrate RNA on both sides of the target uridine. The target uridine is usually located 14-16 nucleotides upstream of either box H or box ACA. This complementary base pairing produces a pseudouridylation pocket in which the target uridine is unpaired and can be reached by its synthase. Not all guide RNAs posses the hairpin-hinge-hairpin-tail structure. In trypanosomes and archaea, RNAs involved in guiding pseudouridylation consist of a single hairpin followed by an ACA box (Liang, Liu et al. 2001). In addition, Huttenhofer’s group found that in archaea, guide RNAs contain one to three stem-loops with less strictly conserved ACA box elements at both sites and the apical stem of these RNAs may contain K-turn motifs that are not detected in eukaryotic H/ACA RNAs (Rozhdestvensky, Tang et al. 2003).
Figure 1.5 H/ACA small ribonucleoprotein particles. Each H/ACA RNP comprises a target-site-specific H/ACA RNA and four core proteins: Nop10p, Nhp2p, Gar1p, and the pseudouridylation synthase Cbf5p. H/ACA RNAs form hairpin-hinge-hairpin-tail secondary structures. Each hairpin contains one large internal loop (the pseudouridylation pocket), which decides the specific sites by base-pairing with the flanking region of target uridine. The target uridine is usually 14 to 16 nucleotides away from the conserved box H (ANANNA) or box ACA. The H/ACA RNA provides a scaffold for the binding of the proteins and the proteins especially Cbf5p catalyze the reaction. Orthologs of yeast Cbf5p have been identified in rats (Nap57), in minifly (Mfl), in drosophila (Nop60b) and in humans (Dyskerin). The proteins pre-assemble into a protein complex that interacts specifically with box H/ACA RNAs. The stoichiometry and organization of the proteins within the complex is not known.
The H/ACA RNAs that direct modifications of ribosomal RNAs and Pol III transcribed spliceosomal U6 snRNA accumulate within the nucleolus (so called snoRNAs) (Tycowski, You et al. 1998). In contrast, the H/ACA RNAs (so called scaRNAs) involved in the modification of the Pol II transcribed spliceosomal snRNAs (U1, U2, U4 and U5 snRNA), are found within the Cajal bodies (Darzacq, Jady et al. 2002). In vertebrate, H/ACA RNAs are encoded within introns and are not independently transcribed but processed from the pre-mRNA introns (Richard, Kiss et al. 2006). However, in yeast, only a few of them are intronic and most of them are derived from independent mono-, di, or polycistronic RNA precursors. The correct processing and subnuclear localization of the mature H/ACA RNAs are directed by conserved sequences through their associated proteins and polymerase II-associated factors. To date, a large number of H/ACA RNAs have been found in humans and have been predicted to be guides for nearly all pseudouridines in human rRNAs, but only few of them are verified by experiments due to technical limitations. In S. cerevisiae, 41 of the 44 known pseudouridine modifications in rRNA have been linked with a verified H/ACA RNA (Schattner, Decatur et al. 2004). However, H/ACA RNAs accounting for the pseudouridylation sites in yeast spliceosomal snRNAs are less well identified.

1.2.3.2.2 The core proteins of H/ACA RNPs

The four core H/ACA proteins named in yeast are Nhp2p, Gar1p, Nop10p and Cbf5p (Henras, Henry et al. 1998; Watkins, Gottschalk et al. 1998; Wang and Meier 2004). In the absence of the H/ACA RNA, the four proteins can form a four-
subunit complex in eukaryotes (Wang and Meier 2004), or a maximal Cbf5-Nop10-Gar1 three-subunit complex in archaea (Baker, Youssef et al. 2005). The four core proteins are required for enzymatic activity, stability and localization of the H/ACA RNP.

Cbf5p, namely dyskerin in human, NAP57 in rat, minifly or Nop60b in Drosophila, is a generic pseudouridine synthase, and belongs to the pseudouridine synthase of TruB family (Massenet, Ansmant et al. 1999). Both Cbf5p and TruB contain two domains: a catalytic core domain that is common to all pseudouridine synthases and a carboxyl-terminal pseudouridine synthase and archaeosine tranglycosylase (PUA) domain (Koonin 1996; Zebarjadian, King et al. 1999). Cbf5p is responsible for the modification carried out by its particle. When Cbf5p was mutated, there were defects in H/ACA RNP activity. Point mutations in the predicted Cbf5p synthase motif abolished pseudouridylation in rRNAs in yeast (Zebarjadian, King et al. 1999). In humans, dyskerin mutation has been found in patients with X-linked dyskeratosis congenital (DC) (Heiss, Knight et al. 1998), a disease characterized by bone marrow failure, skin abnormalities, and increased susceptibility to cancer (Meier 2003). Interestingly, the archaeal pseudouridine synthase Cbf5 has been reported to be able to catalyze pseudouridylation at position 55 of archaeal tRNA in vitro (Roovers, Hale et al. 2006). However, this “stand alone” activity of Cbf5p is greatly enhanced in the presence of the other core H/ACA RNP proteins.

Although much has been learned about Cbf5p, the functional importance of the other three proteins remains largely unclear. Genetic deletion studies have shown
that all the core proteins except Gar1p are required for the stabilization of the H/ACA RNP complex (Bousquet-Antonelli, Henry et al. 1997; Henras, Henry et al. 1998; Watkins, Gottschalk et al. 1998; Dez, Henras et al. 2001). A crystal structure of a complete archael box H/ACA RNP complex shows that Gar1p does not bind to the guide RNA but instead to Cbf5p, suggesting that Gar1p most likely plays role in regulating substrate loading and release (Li and Ye 2006). However, in yeast, the core domain of Gar1p was proposed to be responsible for binding the H/ACA RNA, and indeed the mammalian Gar1 cross-linked to RNA near the catalytic site of the assembled RNP (Bagni and Lapeyre 1998; Rashid, Liang et al. 2006), suggesting that the interaction of Gar1p and the RNA may only occur in eukaryotes or may be transient (Rashid, Liang et al. 2006). Nhp2p has a putative RNA-binding domain, but it binds to the H/ACA RNA secondary structures nonspecifically (Henras, Henry et al. 1998). Unlike Nhp2p, its archaeal ortholog L7Ae binds to a kink-turn motif in RNA specifically. Nop10p contains a structurally well-defined N-terminal region composed of a β-hairpin that directly contacts the box H/ACA RNA. The rest of the protein lacks a globular structure. This unstructured region of Nop10p likely lines the catalytic domain of Cbf5p in the RNP complex, stabilizing it and providing a docking site for Nhp2p (Wang and Meier 2004; Reichow and Varani 2008) (Henras, Henry et al. 1998). The interaction between Nop10p, Nhp2p and the H/ACA guide RNA may serve to position the catalytic center of Cbf5p to the target uridine (Li and Ye 2006).

1.2.4 Functions of pseudouridines
Particular pseudouridine residues have been found to be physiologically essential in various organisms including bacteria, yeast, and mammals (Gutgsell, Englund et al. 2000; Bykhovskaya, Casas et al. 2004; Fernandez-Vizarra, Berardinelli et al. 2007). However, a molecular function for pseudouridine remains unknown. In both spliceosomal snRNAs and rRNAs, pseudouridines are located at similar positions across species. In the five spliceosomal snRNAs, pseudouridines are virtually concentrated in the 5’ half of each RNA molecule, and are clustered in important regions for pre-mRNA splicing among various organisms (Yu, Scharl et al. 1999). snRNA pseudouridylation is important for U snRNP biogenesis, spliceosome assembly and catalytic process of pre-mRNA splicing. Using a Xenopus oocyte reconstitution system, the function of pseudouridines in U2 snRNA in pre-mRNA splicing has been studied. The 6 pseudouridines in the branch site recognition region and the 3 pseudouridines at the 5’ half of U2 snRNA, are all required for snRNP assembly and pre-mRNA splicing (Yu, Shu et al. 1998; Donmez, Hartmuth et al. 2004; Zhao and Yu 2004). The NMR structural studies showed that the presence of pseudouridine at position 35 in yeast U2 snRNA branch site recognition region induced a dramatically altered architectural landscape compared with that of its unmodified counterpart, which maintain the bulge of the branch point nucleotide adenosine for nucleophilic attack in the first step of splicing reaction (Newby and Greenbaum 2001). Furthermore, Yang and Yu were able to show that pseudouridine at position 35 in yeast U2 snRNA is required for pre-mRNA splicing under certain conditions by semi-quantitative RT-PCR (Yang, McPheeters et al. 2005). Functional
information about pseudouridine residues in rRNA is not well known. In *E. coli*, deletion of pseudouridine synthase genes results some cell metabolic defects, which may indicate the role of pseudouridines involved in cooperative stabilization of a particular RNA conformation (Ofengand 2002). In yeast, single or even multiple deletions of a series box H/ACA RNAs had no effect on cell growth, and abolishing pseudouridylation at individual sites around the yeast PTC had no effect (Ofengand and Gournier 1998). However, when all pseudouridines were removed, the cells displayed a reduced protein synthetic rate and growth rate, thus suggesting that pseudouridines may influence both rRNA folding and ribosome assembly.

Pseudouridylation in tRNAs plays an important role in fine-tuning the local structure in which it resides, thereby influencing tRNA decoding activity, improving the fidelity of protein biosynthesis, and helping to maintain the proper reading frame (Harrington, Nazarenko et al. 1993). The pseudouridine residues on the local structure of the anticodon stem-loop (ASL) appear to be critical for the proper binding of tRNA to the ribosome (Dao, Guenther et al. 1994). Pseudouridines found in the anticodon properly play a role in alternative codon usage. In yeast, a minor tRNA\textsuperscript{Ile} variant has the anticodon $\psi\text{A}\psi$ (Senger, Auxilien et al. 1997). Moreover, eukaryotic cytoplasmic tRNA\textsuperscript{Yr} (anticodon G\text{\textPsi}A) is a suppressor of UAA and UAG stop codons (Tomita, Ueda et al. 1999).

1.3 Pseudouridylation in next two chapters

In the next two chapters of my thesis, I would like to talk about my studies on pseudouridylation through different angles. In Chapter 2, I utilized
Xenopus oocytes microinjection as a tool and pre-mRNAs as substrates to explore the roles of pseudouridylation in pre-mRNA splicing. Basicly, I performed a strategy named RNA-guided RNA modification to pseudouridylate pre-mRNA, assayed pre-mRNA splicing after targeting. I found single pseudouridine incorporated into pre-mRNA can negatively affect pre-mRNA splicing. Later, I further explored the mechanisms behind the inhibitory effect caused by pseudouridylation on pre-mRNA by both in vivo and in vitro assays and found that the inhibitory effect is correlated to U2AF$^{65}$ binding defect and the rigidity RNA backbone.

In Chapter 3, I studied pseudouridylation in another organism: S. cerevisiae. I was focusing on spliceosome U1 snRNA instead of pre-mRNA, and I was examining the mechanisms of pseudouridylation instead of exploring its functional roles. There are two conserved pseudouridines, Ψ5 and Ψ6, located at the 5’ arm of yeast U1 snRNA, but how Ψ5 and Ψ6 are introduced remain unclear. I utilized a pseudouridine synthase Cbf5p point mutant strain and Cbf5p conditional depletion strain, as well as another core protein of Box H/ACA RNP, Nhp2p, its conditional depletion strain, found Ψ5 and Ψ6 are Cbf5p and Nhp2p dependent by CMC-modification assay followed by primer extension. Furthermore, I found Ψ5 and Ψ6 are Box H/ACA RNP-dependent by in vitro pseudouridylation assay. The two pseudouridines at position 5 and 6 are mostly likely to be introduced by an RNA-dependent mechanism, but further studies are necessary to confirm this conclusion.
Chapter 2

A flexible RNA backbone within the polypyrimidine tract is required for U2AF\textsuperscript{65} binding and pre-mRNA splicing \textit{in vivo}
The polypyrimidine tract near the 3’ splice site is important for pre-mRNA splicing. Using pseudouridine incorporation and \textit{in vivo} RNA-guided RNA pseudouridylation, we have identified two important uridines in the polypyrimidine tract of adenovirus pre-mRNA. Conversion of either uridine into pseudouridine leads to a splicing defect in \textit{Xenopus} oocytes. Using a variety of molecular biology methodologies, we show that the splicing defect is due to the failure of U2AF\textsuperscript{65} to recognize the pseudouridylated polypyrimidine tract. This negative impact on splicing is pseudouridine-specific, as no effect is observed when the uridine is changed to other naturally-occurring nucleotides. Given that pseudouridine favors a C3’-endo structure, our results suggest that it is the backbone flexibility that is key to U2AF binding. Indeed, locking the key uridine in the C3’-endo configuration while maintaining its uridine identity blocks U2AF\textsuperscript{65} binding and splicing. This pseudouridine effect can also be applied to other pre-mRNA polypyrimidine tracts. Thus, our work demonstrates that \textit{in vivo} binding of U2AF\textsuperscript{65} to a polypyrimidine tract requires a flexible RNA backbone.
2.1 Introduction

In eukaryotic cells, the removal of intervening sequences (or introns) from messenger RNA precursors (pre-mRNA), a process known as splicing, is necessary for the production of mature mRNA (Staley and Guthrie 1998; Burge, Tuschl et al. 1999; Yu, Scharl et al. 1999; Jurica and Moore 2003). Pre-mRNA possesses several key consensus sequence elements that are required for accurate splicing. Of these, the 5’ splice site, the 3’ splice site and the branch site have been well documented (Burge, Tuschl et al. 1999). In higher eukaryotic pre-mRNA, there also exists a polypyrimidine tract (a sequence generally rich in uridines) adjacent to the 3’ splice site (Green 1991; Adams, Rudner et al. 1996; Burge, Tuschl et al. 1999).

Pre-mRNA splicing occurs in the spliceosome, a multi-component complex consisting of five small nuclear ribonucleoproteins (snRNPs)-U1, U2, U4, U5 and U6, and a large number of protein factors (Green 1991; Staley and Guthrie 1998; Burge, Tuschl et al. 1999; Yu, Scharl et al. 1999; Reed 2000; Jurica and Moore 2003). In vitro, the spliceosome is assembled in a step-wise fashion. Early in spliceosome assembly, U1 recognizes the 5’ splice site, and U2AF (U2 auxiliary factor), a heterodimer with a 65kDa subunit and a 35kDa subunit, recognizes the polypyrimidine tract/3’ splice site (Figure 2.1A). U2AF\(^{65}\) also interacts with other splicing factors, and helps recruit U2 to the branch site, thereby facilitating base-pairing interactions between U2 and the branch site. Finally, the U4/U5/U6 tri-snRNP joins the complex, converting it into a fully assembled spliceosome. Following a
series of conformational changes, the spliceosome becomes activated, and initiates splicing reactions.

The binding of U2AF\textsuperscript{65} to the polypyrimidine tract has been extensively studied (Wallace and Edmonds 1983; Singh, Valcarcel et al. 1995; Singh, Banerjee et al. 2000; Sickmier, Frato et al. 2006). Although polypyrimidine tracts vary widely in their nucleotide sequences, a pyrimidine (primarily uridine)-rich consensus sequence is sufficient for U2AF\textsuperscript{65} binding. Several years ago, the Kielkopf lab solved a crystal structure of the U2AF\textsuperscript{65} RNA binding domain bound to a polypyrimidine tract (Sickmier, Frato et al. 2006). The protein-RNA structure suggests that the side chains of U2AF\textsuperscript{65} are quite flexible in binding with RNA, thus offering a possible explanation as to how U2AF\textsuperscript{65} is able to bind a variety of polypyrimidine tract sequences (Sickmier, Frato et al. 2006). However, little is known about polypyrimidine tract requirements for binding.

Pseudouridine (Ψ) is found in stable eukaryotic RNAs, such as snRNAs, tRNAs and rRNAs (Massenent, Mougin et al. 1998; Ofengand and Gournier 1998; Stephenson, Karijolish et al. 2008). Ψ is derived from uridine via a post-transcriptional isomerization reaction known as pseudouridylation, where the nitrogen-carbon bond linking the base to the sugar ring is broken, and a new carbon-carbon bond forms, re-establishing the base-sugar linkage. The U-to-Ψ conversion frees the nitrogen on the base, thus creating an extra hydrogen bond donor in pseudouridine. In addition to a possible function in RNA-protein and RNA-RNA interactions, the extra hydrogen bond donor in Ψ can potentially add rigidity to the

Box H/ACA RNP-catalyzed (or RNA-guided) pseudouridylation is the primary mechanism responsible for most U-to-Ψ conversions in eukaryotes (Padgett, Konarska et al. 1984; Yu, Terns et al. 2005). Box H/ACA RNP consists of one box H/ACA RNA and four core proteins, including Cbf5/Nap57/Dyskerin (pseudouridylase), Nhp2, Gar1 and Nop10 (Wang and Meier 2004; Baker, Youssef et al. 2005; Charpentier, Muller et al. 2005; Yu, Terns et al. 2005). The RNA component of the RNP folds into a unique structure known as the helix-hinge-helix-tail (Figure 1B). The two internal loops each forms a pseudouridylation pocket that base-pairs with the RNA substrate, positioning the target uridine at the base of the upstem (Figure 2.1B) (Ganot, Bortolin et al. 1997; Ni, Tien et al. 1997). Once the target uridine is identified and positioned, the pseudouridylase Cbf5/Nap57/Dyskerin converts it into pseudouridine (Duan, Li et al. 2009; Liang, Zhou et al. 2009). This modification scheme has been extensively studied, and consequently, verified in various organisms (Ganot, Bortolin et al. 1997; Ni, Tien et al. 1997; Wang and Meier 2004; Baker, Youssef et al. 2005; Charpentier, Muller et al. 2005; Liang, Liu et al. 2007; Xiao, Yang et al. 2009).

In the current work, we use U-to-Ψ conversion to identify uridines in pre-mRNA that are important for splicing in vivo. We show that the conversion of uridine into pseudouridine (but not any other nucleotide) at either position 5 or position 7 in the polypyrimidine tract of adenovirus pre-mRNA (Figure 2.1) blocks pre-mRNA
splicing in *Xenopus* oocytes. Detailed analyses indicate that pseudouridylation at the critical uridine site brings rigidity to the backbone of the polypyrimidine tract. Consequently, the pseudouridylated polypyrimidine tract loses its ability to bind to U2AF<sup>65</sup>. Our results thus demonstrate that *in vivo* binding of U2AF<sup>65</sup> requires a flexible RNA backbone within the polypyrimidine tract.

### 2.2 Results

**Substitution of uridines with pseudouridines in the polypyrimidine tract blocks pre-mRNA splicing**

To identify uridines within pre-mRNA that are required for splicing, we substituted uridines of the adenovirus pre-mRNA splicing substrate with pseudouridines via an *in vitro* transcription (using \(\Psi\)TP instead of UTP). The pre-mRNA was then injected into *Xenopus* oocytes, and the *in vivo* splicing was assessed.

As shown in Figure 2.2A, while regular pre-mRNA with uridines was efficiently spliced, pre-mRNA containing pseudouridines completely failed to splice. In this experiment, because all uridines were changed to pseudouridines in the pre-mRNA, the minimal U-to-\(\Psi\) changes needed to block splicing remained unclear.

To narrow down the list of possible important uridine sites, we constructed three pairs of chimeric pre-mRNAs using a combination of site-specific RNase H cleavage directed by 2'-O-methylation RNA-DNA chimera and two-piece ligation (Figure 2.2B) (Yu, Shu et al. 1998). One RNA in each pair derived its 5' sequence [exon 1 plus the first two nucleotides of the intron (arrow 1), exon 1 plus the 5’ intron sequence prior to the polypyrimidine tract (arrow 2), or exon 1 plus all but the last...
Figure 2.1 Early stage of spliceosome assembly and box /ACA RNA in directing RNA pseudouridylation. (A) At the early stage in spliceosome assembly, U1 recognizes the 5’ splice site of pre-mRNA, and U2AF65 recognizes the polypyrimidine tract. The branch site (BS) and polypyrimidine tract (PolyPy) of adenovirus pre-mRNA are shown. U1 and U2AF (both the 65 and 35 subunits) are also indicated. The uridines in the polypyrimidine tract are sequentially numbered. (B) The hairpin-hinge-hairpin-tail structure of a box H/ACA RNA is shown. The internal loops of the hairpins serve as guides (or pseudouridylation pockets) that base-pair with their respective substrates. The thick lines represent substrate RNAs, and the thin lines stand for box H/ACA RNA. The target nucleotides to be pseudouridylated are indicated (Ψ). Box H and Box ACA are also indicated. The guide sequence of the 5’ pseudouridylation pockets is taken from PugU2-34/44, a Xenopus box H/ACA guide RNA targeting U2 at position 34 and 44.
three nucleotides of the intron (arrow 3)] from pseudouridine-substituted pre-mRNA transcript, and its 3’ sequence [all but the first two nucleotides of the intron plus exon 2 (arrow 1), the intron sequence starting from the polypyrimidine tract plus exon 2 (arrow 2), or the last three nucleotides of the intron plus exon 2 (arrow 3)] from regular pre-mRNA transcript (containing uridines); conversely, in the other RNA, the 5’ sequence was derived from regular pre-mRNA transcript, and its 3’ sequence was from pseudouridine-substituted transcript. These chimeric pre-mRNAs were separately injected into the nuclei of *Xenopus* oocytes for splicing assays. As shown in Figure 2.2C, pre-mRNAs containing pseudouridines in the 5’ portion of the molecule, including exon 1 and the first two nucleotides of intron (lane 3) or exon 1 plus the intron sequence until prior to the polypyrimidine tract (lane 5), spliced just as efficiently as did the control pre-mRNA containing no pseudouridines (lane 1). In contrast, the reciprocal pre-mRNAs containing pseudouridines in the 3’ portion of the molecule, starting with either the third nucleotide of the intron (lane 4) or the polypyrimidine tract (lane 6), completely failed to splice, just as the fully substituted pre-mRNA did (lane 2). However, when the 5’ portion of the molecule was extended to include the polypyrimidine tract, the outcome became completely different. Specifically, pre-mRNA containing pseudouridines in the 5’ portion of the molecule completely failed to splice (lane 7); however, the reciprocal chimeric pre-mRNA containing pseudouridines in the 3’ portion of the molecule (lane 8) spliced as efficiently as did the control pre-mRNA (containing no pseudouridines) (lane 1). Together, these results indicated that the uridine sites in the polypyrimidine tract
A

Uninjected     Injected

M               U               M               U               U               U
ΨΨ ΨΨ            ΨΨ ΨΨ            U               ΨΨ ΨΨ

622             527             424             307             217

Uninjected    Injected

201            238

190            160

147

Uninjected    Injected

B

Exon1


BS

PolyPy

Exon2

E1

(Ψ)

1

2

3

(Ψ)

E2

E1

(Ψ)

1

E2

(Ψ)

E1

(Ψ)

2

E2

(Ψ)

E1

(Ψ)

2

E2

(Ψ)

Site-specific cleavage

Two-piece Ligation

Pair 1

5'Ψ-3'U

5'U-3'Ψ
Figure 2.2 Mapping of the important uridines within adenovirus pre-mRNA. (A) Pre-
mRNA splicing in Xenopus oocytes. Lane 1 and 2 are un-injected pre-mRNAs,
regularly transcribed (U) and pseudouridine-substituted (Ψ), respectively. Lane 3 and
4 are injected pre-mRNAs, regularly transcribed (U) and pseudouridine-substituted
(Ψ), respectively. The RNA bands corresponding to the lariat intron, pre-mRNA, and
mature mRNA, are indicated. (B) Construction of chimeric pre-mRNAs using RNase
H site-specific cleavage directed by 2’-O-methyl RNA-DNA chimeras followed by
mix-and-match two-piece ligation. The three cleavage sites are indicated, and the
cleavage and ligation strategy, using cleavage site 1 as an example, is known.
Pseudouridine-substituted pre-mRNA is on the left, and the regularly transcribed
(containing U) pre-mRNA is on the right. Upon cleavage at the same site (directed by
2’-O-methyl RNA-DNA chimera), the 5’ and 3’ halves were mixed and matched.
Two-piece ligation was then followed, generating a pair of chimeric pre-mRNA. (C)
Three pairs of chimeric pre-mRNAs (lanes 3-8) as well as regularly transcribed (lane
1, all U) and pseudouridine-substituted (lane 2, all Ψ) pre-mRNAs were assayed for
splicing in Xenopus oocytes. 5’Ψ-3’U represents pre-mRNAs whose 5’ halves
contained Ψs and 3’ halves contained Us. 5’U-3’Ψ represents pre-mRNAs with
reciprocal combinations.
(a total of 9) were sensitive to pseudouridine substitution.

**Two of the nine uridines in the polypyrimidine tract are sensitive to U-to-Ψ change**

To further dissect the polypyrimidine tract, and to pinpoint the important uridine(s) in this region, we used a new strategy, namely, RNA-guided RNA pseudouridylation, to examine each of the nine uridine sites. Using PugU2-34/44, a known naturally-occurring Xenopus box H/ACA RNA, we constructed nine artificial box H/ACA guide RNAs, each targeting one of the nine uridines in the polypyrimidine tract. Specifically, for each artificial guide RNA, only the guide sequence in the 5’ pseudouridylation pocket of PugU2-34/44 was changed (to target the uridines in the polypyrimidine tract); all the other PugU2-34/44 sequences remained unchanged (see Figure 2.1B).

The artificial box H/ACA RNA guide RNAs and pre-mRNA were sequentially injected into *Xenopus* oocytes depleted of U2. The purpose of using U2-depleted oocytes was to slow down (or completely pause) splicing, thus providing enough time for the guide RNA to direct pre-mRNA pseudouridylation in the polypyrimidine tract. After a four-hour incubation, U2 snRNA was then injected, thereby resuming splicing competence in the oocytes. As shown in Figure 2.3A, while injection of seven of the nine guide RNAs has no effect on pre-mRNA splicing, injection of guide RNA targeting U₅ or U₇ (lane 13 and 15) resulted in greatly reduced splicing capability. To ensure that RNA-guided pseudouridylation occurred efficiently, we injected, under the same conditions, pre-mRNAs containing a single $^{32}$P at the target sites (rather than
the uniformly labeled pre-mRNA). Upon a four-hour incubation, RNA was recovered, and pseudouridylation efficiency was assessed using nuclease P1 digestion followed by TLC analysis. Our results indicated that RNA-guided pseudouridylation occurred almost equally efficiently (~70-80%) at the target sites (data not shown). Thus, our results suggested that the splicing function of only two uridine sites, U₅ and U₇, are sensitive to U-to-Ψ substitution.

To verify our observations, we constructed, using chemical synthesis and two-piece ligation, a pre-mRNA in which U₇ was completely substituted with Ψ₇. This pre-mRNA completely failed to splice, whereas the control pre-mRNA (containing U₇) underwent substantial splicing (Figure 2.3B). Thus, our experiments demonstrated that U₇ (and most likely U₅ as well) was a key uridine that could not tolerate U-to-Ψ substitution.

The inhibitory effect on splicing is pseudouridine-specific

Although it is an isomerization reaction, strictly speaking, the U-to-Ψ conversion could be considered a mutation. To gain a better understanding of the effect of this U-to-Ψ conversion, we next carried out a comparison analysis in which U-to-Ψ conversion was compared with a uridine-to-other-nucleotide mutation.

Specifically, we constructed pre-mRNAs containing either a U₇-to-G₇, U₇-to-C₇, or U₇-to-A₇ mutation in the polypyrimidine tract, and tested their ability to splice in Xenopus oocytes. Interestingly, these mutant pre-mRNAs spliced nearly as well as did the wild-type pre-mRNA (Figure 2.4A, lane 4 and 5, and data not shown), sharp contrast to the result shown in Figure 2.3B, where pre-mRNA containing Ψ₇ in
the polypyrimidine tract completely failed to splice. Consistently, in the presence of the artificial guide RNA targeting U₇, wild-type pre-mRNA splicing was greatly reduced (Figure 2.4A, lane6). Together, these results indicated that the inhibitory effect on pre-mRNA splicing was not merely due to a mutation. It appeared that the effect was pseudouridine-specific.

**The change of U to Ψ in the polypyrimidine tract affects binding of a splicing factor at an early stage in spliceosome assembly**

To identify the stage of spliceosome assembly or the function affected by the U-to-Ψ change, we conducted a native complex gel analysis (Yu, Shu et al. 1998). Wild-type pre-mRNAs containing U₇ in the polypyrimidine tract, mutant pre-mRNAs containing G₇ in the polypyrimidine tract, or altered pre-mRNAs containing Ψ₇ in the polypyrimidine tract, were injected into the oocyte nuclei under mineral oil. Ten minutes later, the nuclei were broken, mixed with native gel loading dye, and loaded on a native gel. As shown in Figure 2.4B, both wild-type pre-mRNA and mutant pre-mRNA containing G₇ in the polypyrimidine tract were able to assemble into splicing complexes A, B, and C (lane 2 and 3). In contrast, pre-mRNA containing Ψ₇ was unable to form any splicing-specific complexes (lane 4), suggesting that the effect manifested itself at an early stage during spliceosome assembly.

To gain some insights into the mechanisms by which the U-to-Ψ conversion in the polypyrimidine tract inhibited pre-mRNA splicing, we carried out a competition assay in *Xenopus* oocytes. Before injecting pre-mRNA, we injected a
Figure 2.3 Identification of two important pseudouridines in the polypyrimidine tract by RNA-guided RNA pseudouridylation. (A) Nine artificial box H/ACA guide RNAs targeting the nine uridines in the polypyrimidine tract (gRNA-U#1-gRNA-U#9) were separately injected into U2-depleted oocytes (lane 4, 5, 8, 9 13-17). Labeled pre-mRNA was then injected. Following pre-mRNA pseudouridylation, U2 snRNA was injected to rescue splicing activity. Lanes 1 and 10 are controls in which radiolabeled pre-mRNA was directly injected into oocytes. Lane 2, 6, and 11 are negatively controls in which radiolabeled pre-mRNA was injected into U2-depleted oocytes, and no rescuing U2 was injected later. Lanes 3, 7 and 12 are positive controls where radiolabeled pre-mRNA was injected into U2-depleted oocytes, and rescuing U2 was injected later. Bands corresponding to lariat intron, pre-mRNA, and rescuing U2 are indicated. Relative splicing efficiency was quantified for each reaction, and the ratio of lariat intron to total pre-mRNA (un-spliced pre-mRNA with lariat intron) is shown under each lane. (B) Pre-mRNA containing a wild-type $U_7$ (lane 1) or a $U_7$-to-$\Psi_7$ change (lane 2) in the polypyrimidine tract was synthesized (also see the top panel), and splicing was examined in Xenopus oocytes. Bands corresponding to the 2/3 lariat intermediate, lariat intron product and un-spliced pre-mRNA are indicated.
short RNA sequence, corresponding either to the wild-type polypyrimidine tract or an altered polypyrimidine tract where uridines were changed to pseudouridines (Figure 2.4C). While a dose-dependent inhibitory effect was observed when the wild-type polypyrimidine tract was injected (Figure 2.4D, compare lane 2 and 3 with lane 1), no effect was observed when the pseudouridine-substituted polypyrimidine tract was injected (compare lane 4 and 5 with lane 1). This result suggested that the polypyrimidine tract was competing in trans with pre-mRNA in the binding with a splicing factor, and that the substitution of uridines with pseudouridines abolished its ability to bind to this splicing factor.

**U-to-Ψ change in the polypyrimidine tract affects U2AF binding**

Given that the essential splicing factor U2AF$^{65}$ is known to bind to the polypyrimidine tract at an early stage of splicing, we reasoned that the U-to-Ψ change in the polypyrimidine tract may give rise to a U2AF$^{65}$-binding defect, thereby inhibiting splicing. To test this hypothesis, we carried out the immunoprecipitation analysis. After injection of radiolabeled pre-mRNA, either the wild-type pre-mRNA or the pre-mRNA containing Ψ$_7$ in the polypyrimidine tract, the oocyte nuclei were broken and immunoprecipitated with either anti-U2AF$^{65}$ antibody or anti-Sm antibody (specific for snRNPs) as a control. The co-precipitated RNA was recovered, and analyzed on a denaturing gel. As shown in Figure 2.5 A, while wild-type pre-mRNA was co-precipitated with U2AF$^{65}$ (lane 1), virtually no altered pre-mRNA, containing Ψ$_7$ in the polypyrimidine tract, was brought down by anti-U2AF$^{65}$ antibody (lane 2). In contrast, both pre-mRNAs, unmodified wild-type pre-mRNA and pre-mRNA
**Figure 2.4** Ψ-specific effect and impact on the assembly of early splicing complexes. (A) Wild-type pre-mRNA (lane 1, 2, 3, and 6) as well as a mutant pre-mRNA containing a U₇-to-G₇ change (lane 4 and 5) were assayed for splicing in Xenopus oocytes. Lane 1 and 4, splicing in intact oocytes; lane 3, splicing in U2-depleted oocytes; lane 2, splicing in oocytes where U2 was depleted and later reconstituted; lane 5 and 6, as in lane 2 with the exception that an artificial box H/ACA guide RNA, targeting position 7 of the polypyrimidine tract, was present. Splicing efficiency was quantified for each reaction, and the ratio of lariat intron to total mRNA (un-spliced pre-mRNA and lariat intron) is shown under each lane. (B) Wild-type pre-mRNA (lane 1 and 2) and mutant pre-mRNA containing a U₇-to-G₇ (lane 3) or a U₇-to-Ψ₇ (lane 4) change were assayed for spliceosome assembly in Xenopus oocytes. Lane 1 is a negative control where spliceosome assembly was assayed in U2 depleted oocytes. Splicing complexes A, B, and C, as well as the heterogeneous complex H are indicated. (C) Competitor adenovirus polypyrimidine tracts (with U, PolyPyU on the left, and with Ψ, PolyPyΨ on the right), sandwiched by two stem structures, are shown. (D) Wild-type pre-mRNA splicing was tested in the presence of competitor PolyPyU-15 fold (lane 2) and 45-fold (lane 3) molar excess relative to pre-mRNA, or in the presence of competitor PolyPyΨ-15 fold (lane 4) and 45-fold (lane 5) molar excess. To monitor the stability of the competitors, trace amount of radiolabeled PolyPyU or PolyPyΨ was mixed with the unlabeled competitor before injection. Lane 1, splicing without competitor. Bands corresponding to lariat intron, un-spliced pre-mRNA, spliced mRNA, as well as competitor PolyPyU or PolyPyΨ, are indicated.
containing Ψ7 in the polypyrimidine tract, were precipitated by anti-Sm antibody (lane 3 and 4). These results are consistent with the fact that both U1 snRNP and U2AF independently bind to pre-mRNA at early times during spliceosome assembly, and indicated that U7-to-Ψ7 change blocked binding with U2AF65, but not with U1. Furthermore, in the case of the wild-type pre-mRNA, which was spliced, no lariat intermediate and product were co-precipitated with U2AF65 (lane 1); the lariat intermediate and product were only in the supernatant (lane 5). However, the lariat intermediate and product were precipitated by anti-Sm antibody. These results are in agreement with the factor that U2AF65 is released from pre-mRNA prior to functional spliceosome formation (Bennett, Michaud et al. 1992; Chiara, Palandjian et al. 1997; Tisserant and Konig 2008), whereas U2, U5, both of which are targets of anti-Sm, remain bound to the lariat intermediate and product after first and second steps of splicing (Moore, Query et al. 1993; Sharp 1994).

To confirm our immunoprecipitation results, we carried out the in vitro binding assay (Moore, Query et al. 1993; Sharp 1994) using U2AF65 and the polypyrimidine tract of adenovirus pre-mRNA containing either U7 or Ψ7. As shown in Figure 2.5B, while U2AF65 bound to the U7-containing polypyrimidine tract with high affinity (Kd = 1.8 × 10^-7 M) (lanes 1-4), binding of U2AF65 to the Ψ7-containing polypyrimidine tract was greatly reduced (Kd = 8.1 × 10^-6 M) (lanes 5-8). Interestingly, based on native gel analysis, the U7-containing polypyrimidine tract appeared to adopt two conformations, with a vast majority of the molecules in a low-mobility conformation and a small fraction in a high-mobility conformation (lanes 1-
4). It appeared that only the low-mobility fraction shifted upon addition of U2AF^{65} (lanes 1-4), suggesting that this conformation was active in binding with U2AF^{65}. In contrast, the Ψ_{7}-containing polypyrimidine tract adopted only the high-mobility conformation (lanes 5-8), which was barely shifted by the addition of U2AF^{65} (lanes 5-8). Thus, these results have reinforced the notion that a single pseudouridine in the polypyrimidine tract can alter the folding (or conformation) of polypyrimidine tract, impacting negatively on U2AF^{65} binding.

**U2AF^{65} binding requires a flexible polypyrimidine tract backbone**

It is well established that through a bridging water molecule, pseudouridine can form a hydrogen bond with its own phosphate backbone, thus favoring the C3'-endo sugar ring configuration and bringing rigidity to the backbone and the base (Arnez and Steitz 1994; Charette and Gray 2000; Newby and Greenbaum 2002; Kolev and Steitz 2006). Our observed pseudouridine-specific effect on U2AF binding to the polypyrimidine tract suggested that it was perhaps the inflexibility of the backbone that negatively impacted U2AF binding. To test this hypothesis, we synthesized a pre-mRNA in which the important uridine (U_{7}) within the polypyrimidine tract was locked into the C3'-endo sugar ring configuration (the nucleotide itself was kept as uridine) (Kierzek, Ciesielska et al. 2005). This pre-mRNA was injected into *Xenopus* oocytes for splicing assay.

As shown in Figure 2.6, while the pre-mRNA containing a regular uridine (U_{7}) was spliced efficiently (lane1), splicing was almost completely blocked when U_{7} was locked up in the C3'-endo configuration (locked nucleic acid, LNA, in which its 2'-O
was linked, through a carbon, to the 4’-C position (lane 3). As a control, when an unimportant uridine (U₃) was locked in the C3’-endo configuration, splicing was not affected (lane 2).

To further confirm our observations and to rule out the possibility that the 2’-OH of U₇ played a role in splicing (2’-OH was altered in LNA), we synthesized additional pre-mRNAs, in which the 2’-OH group of U₇ was changed to 2’-F or 2’-H. It was reported that, in relative terms, 2’-F favors the C3’-endo configuration, whereas 2’-H brings flexibility to the backbone (Kolev and Steitz 2006). As shown in Figure 2.6, splicing occurred efficiently when pre-mRNA with a 2’-H at U₇ was used (lane 4). In contrast, when pre-mRNA with a 2’-F at U₇ was used, a reduction in splicing, albeit small, was evident (lane 5). These results, coupled with the results of conventional mutagenesis analysis (Figure 2.4), suggest that it is the sugar backbone configuration (rather than the nucleotide identity or the 2’-OH moiety) at position 7 that is key to U2AF₆₅ binding and splicing.

**U-to-Ψ change also affects the splicing of pre-mRNA bearing a β-globin polypyrimidine tract**

It is known that the polypyrimidine tract sequences in various mammalian pre-mRNAs, although pyrimidine-rich, are quite different. To test whether the effect of U-to-Ψ change in the adenovirus polypyrimidine tract can be applied to other pre-mRNAs, we replaced the polypyrimidine tract of the adenovirus pre-mRNA
<table>
<thead>
<tr>
<th></th>
<th>Pellet</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>aU2AF&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>aS</td>
<td>aU2AF&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>aU2AF&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
<tr>
<td>aS</td>
<td>aU2AF&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>U#7ψ&lt;sup&gt;-&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Pellet | Sup. | Correlation
---|---|---
1 | 2 | 3 | 4 | 5 | 6 | 7 | 8
**Figure 2.5** Effect of U$_7$-to-$\Psi_7$ change on U2AF$^{65}$ binding. (A) Wild-type pre-mRNA (odd numbered lanes) or pre-mRNA containing a $\Psi_7$ in the polypyrimidine tract (even numbered lanes) was injected into the nuclei of Xenopus oocytes. Nuclear extracts were subsequently prepared, and immunoprecipitated with anti-U2AF$^{65}$ (lane 1, 2, 5, and 6), or anti-Sm (lane 3, 4, 7, and 8) antibodies. Lanes 1-4 are precipitated fraction (Pellet), and lanes 5-8 are unbound fraction (Supernatant). (B) In vitro U2AF$^{65}$-polypyrimidine binding assay. 5’-end radiolabeled U$_7$-containing polypyrimidine tract (U$_7$) (lanes 1-4) or $\Psi_7$-containing polypyrimidine tract ($\Psi_7$) (lanes 5-8; this panel was over-exposed in order to visualize the shifted complex) was incubated with indicated amount of pure U2AF$^{65}$ (lane 1-8). After reaction, the RNA-protein complex (indicated as U2AF-RNA) and unbound RNA (indicated as Low and High Mobil) were resolved on the native gel. The sequences of U$_7$ and $\Psi_7$ are shown on the top.
Figure 2.6 U2AF65 binding requires RNA backbone flexibility within the polypyrimidine tract. Wild-type pre-mRNA (lane 1), as well as pre-mRNA in which U3 (lane 2) or U7 (lane 3) in the polypyrimidine tract was locked in the C3'-endo configuration (LNA), or the 2'-OH group of U7 was changed to 2'-H (lane 4) or 2'-F (lane 5), were assayed for splicing in Xenopus oocytes. Relative splicing efficiency [mature mRNA/(mRNA+lariat intron+un-spliced pre-mRNA)] was quantified for each reaction, and the quantification is shown under each lane.
**Figure 2.7** Ψ-effect can be applied to different polypyrimidine tracts. Pre-mRNA containing an adenovirus (lane 1 and 2) or β-globin (lane 3 and 4) polypyrimidine tract was assayed for splicing in Xenopus oocytes. In lanes 1 and 3, uridines were incorporated into the polypyrimidine tracts, whereas in lane 2 and 4, pseudouridines were incorporated into the polypyrimidine tracts. Bands, corresponding to lariat intron, un-spliced pre-mRNA, and spliced mRNA, are indicated. The intron sequence of pre-mRNA containing β-globin polypyrimidine tract is shown on the bottom.
(UCCCCUUUUUUCUC) with that of the β-globin pre-mRNA
(UUUUCCCACCCUU). The new pre-mRNA was synthesized in vitro, injected into
Xenopus oocytes, assayed for splicing. As shown in Figure 2.7, while pre-mRNA
(with an adenovirus polypyrimidine tract) did (lane 2), pre-mRNA containing
psuedouridines completely failed to splice (lane 4), just as its pseudouridine-
substituted parental pre-mRNA did (lane 2). Further analysis indicated that the second
5′-most uridine (out of six uridines) in the β-globin polypyrimidine tract could not
tolerate U-to-Ψ change (data not shown). These results suggested that the effects of
the U-to-Ψ change could be applied to different polypyrimidine tracts, despite the fact
that their sequences are quite different (see Discussion).

2.3 Discussion

Using the Xenopus oocyte microinjection system, we have identified two
uridines U₅ and U₇ in the pre-mRNA polypyrimidine tract of adenevirous pre-mRNA
that cannot tolerate U-to-Ψ change without compromising splicing. Indeed, U-to-Ψ
conversion at site 7 (likely site 5 as well) inhibited the binding of U2AF, an essential
splicing factor in higher eukaryotes, thus suggesting that the two uridines are critical
for U2AF binding. Further dissection indicated that a flexible backbone, which was
rigidified upon the U-to-Ψ conversion, was required for U2AF binding. Our current
work has also raised some interesting questions with regard to the observed effect of
U-to-Ψ change.

U-to-Ψ conversion in the pre-mRNA polypyrimidine tract inhibits splicing: a Ψ-
specific backbone effect
According to our experimental data (Figure 2.4), while a U-to-ψ change at position 7 of the polypyrimidine tract of adenovirus pre-mRNA resulted in a defect in U2AF binding and pre-mRNA splicing, U-to-other-nucleotide (G, C or A) changes led to no defect in splicing. These results indicate that the observed effect is not simply due to conventional mutations, but is due to U-to-ψ isomerization and is therefore a ψ-specific effect.

What is unique about ψ? Isomerization of U to ψ involves the breakage of a N-C bond that used to link the base with the sugar ring, and the concurrent formation of a new C-C bond which re-establishes the linkage between the base and sugar. Consequently, this reaction frees a nitrogen, resulting in an extra hydrogen bond donor in the newly formed ψ. It is possible that this hydrogen bond donor specifically forms a hydrogen bond with some other factor(s), thereby competing with the binding with U2AF₆⁵. Alternatively, the new hydrogen bond donor could form a hydrogen bond with a water molecule, which in turn would form another two hydrogen bonds with the phosphodiester-backbone of the RNA (Arnez and Steitz 1994; Charette and Gray 2000; Newby and Greenbaum 2002; Kolev and Steitz 2006). Consequently, both the ψ base and its sugar ring backbone become rigid. Such a rigid configuration may disfavor binding with U2AF, giving rise to a splicing defect phenotype. Our experimental results strongly support the hypothesis that a flexible sugar backbone at position 7 (and most likely at position 5 as well), or a free rotation between C3’-endo (N form) and C2’-endo (S form), is required for U2AF₆⁵ binding in vivo. Specifically, we have shown that, when the U₇ is locked in the C3’-endo configuration (LNA via
2\textsuperscript{\prime}′-O, 4\textsuperscript{\prime}′-C linkage), the polypyrimidine tract becomes inactive in binding with U2AF\textsuperscript{65} (Figure 2.6). However, when the uracil base at position 7 is changed to other bases (Figure 2.4) or the 2\textsuperscript{\prime}′-OH group at position 7 is change to 2\textsuperscript{\prime}′-H (Figure 2.6), splicing is not affected. Thus, our results indicate that it is the sugar ring backbone flexibility, rather that than the base identity or the 2\textsuperscript{\prime}′-OH moiety, that is critical for U2AF\textsuperscript{65} binding. Along this line, it has been reported that the rigid nature of the ψ-containing Sm binding site is accountable for the defect in the binding of U7 snRNA with its Sm proteins (Arnez and Steitz 1994; Charette and Gray 2000; Newby and Greenbaum 2002; Kolev and Steitz 2006).

In regard to the structural basis of polypyrimidine tract-U2AF\textsuperscript{65} binding, the crystal structure of a U2AF\textsuperscript{65} RNA binding domain bound to a seven-uridine tract (U\textsubscript{1}U\textsubscript{2}U\textsubscript{3}U\textsubscript{4}U\textsubscript{5}U\textsubscript{6}U\textsubscript{7}) RNA has been solved (Sickmier, Frato et al. 2006). The structural data indicates that there is only one 2\textsuperscript{\prime}′-OH (the 2\textsuperscript{\prime}′-OH of U\textsubscript{5}) that interacts with U2AF\textsuperscript{65}; there are no contacts between U2AF\textsuperscript{65} and the 2\textsuperscript{\prime}′-OH groups of the other six uridines. This is consistent with our experimental results, which indicate that at least the 2\textsuperscript{\prime}′-OH group of U\textsubscript{7} in the adenovirus polypyrimidine tract is not required for U2AF\textsuperscript{65} binding (Figure 2.6). Although most interactions are through contacts between uridine residues and U2AF\textsuperscript{65} (Sickmier, Frato et al. 2006), the C5 and C6 atoms of the uridine residues show no contact with U2AF\textsuperscript{65}. As pseudouridine is the C5-glycoside isomer of pseudouridine, this would indicate that the N1 atom of the pseudouridine base does not interact with U2AF\textsuperscript{65} and alter its binding (Figure 2.5). Importantly, the crystal structure also indicates that the sugar puckers of the first three
uridines adopt the C2’-endo conformation (C. Kielkopf, unpublished data), a structure often adopted by DNA rather than RNA. Notably, while binding with U2AF$^{65}$, the seven-uridne tract forms a turn, with the third nucleotide being at the turn of the chain. Given the above structural information, it is tempting to align the seven-uridine tract RNA with the adenovirus polypyrimeidine tract. Assuming the alignment begins with U$_5$, U$_6$, and U$_7$ we would predict that these residues will adopt the C2’-endo conformation. Thus, replacing them with pseudouridines or with LNA would favor the C3’-endo conformation resulting in a profound defect in U2AF$^{65}$ binding and pre-mRNA splicing (Figure 2.5 and Figure 2.6). However, our results do not offer an explanation as to why U$_6$, the second-most 5’ uridine of the seven-uridine polypyrimeidine tract, was not identified as have a requirement for backbone flexibility. Perhaps the C2’-endo conformation of U$_6$ is not as crucial for U2AF$^{65}$ binding.

Is the U-to-ψ effect universal to all polypyrimeidine tracts?

We have shown that U-to-ψ change in the polypyrimeidine tract of adenovirus pre-mRNA blocks U2AF binding and pre-mRNA splicing. It appears that this effect of U-to-ψ change is not limited to the adenovirus polypyrimeidine tract, as when the polypyrimeidine tract of β-globin pre-mRNA was used, a similar effect was observed, suggesting that this effect is universally applicable.

Given that the sequences of polypyrimeidine tracts are known to vary between various pre-mRNAs, the commonality of determinants in different sequences is somewhat puzzling. Which uridine(s) (appearingly with no commonality in
sequence context) is critical, or in other words, which uridine(s) cannot be changed to pseudouridine(s) without compromising U2AF binding and pre-mRNA splicing?

According to the data obtained from our experiments, it is U₅ and U₇ of the adenovirus polypyrimidine tract (UCCCCUUUU₅UU₇UUCC), and U₂ of the β-globin polypyrimidine tract (UU₂UUCCACCCUU). Interestingly, there is a discrepancy between the two polypyrimidine tracts. What is the molecular basis for the inhibitory effect we observed? Again, in regards to the seven-uridine tract bound to the U2AF₆⁵ RNA binding domain, if the polypyrimidine tract begins with the 5’-most C, U₂ of the β-globin polypyrimidine tract (CUU₂UUCC) becomes the third pyrimidine and would be equivalent to U₃ of the seven-uridine tract, thus offering a possible explanation of why substitution of this uridine with a pseudouridine resulted in a splicing defect.

Thus, we predict that the sugar pucker of the 5’-most C and the third pyrimidine (U₂) prefer to adopt a C2’-endo conformation when bound by U2AF₆⁵. In our experiments, the only sequence difference between the two pre-mRNAs is the polypyrimidine tract sequence. The distances between the two tracts and branch site or the 3’ splice site are the same. However, in nature, the locations of polyprimidine tracts in introns are quite versatile. It will be interesting to identify the important uridine residues in different polypyrimidine tracts within their natural pre-mRNA context, and determine whether the distance to either the branch point sequence or the 3’ splice site influence which uridine residues are functionally restricted to a C2’-endo conformation.

The polypyrimidine tract provides a possible target for gene silencing in vivo
We have shown that a U-to-ψ change at a specific site(s) in the polypyrimidine tract effectively blocks pre-mRNA splicing. Using artificial box H/ACA guide RNAs to target the critical uridines, we have also recapitulated this inhibitory effect in vivo. Thus, such an approach provides a way to regulate gene expression at the level of pre-mRNA splicing. However, it has long been noted that box H/ACA RNAs are localized to the nucleoli and Cajal bodies, where their natural target RNAs, rRNAs and snRNAs, are located. This raises an interesting questions as to whether mRNA or pre-mRNA, which is known to be absent in the nucleoli and Cajal bodies, will be modified by the box H/ACA-guided mechanism.

However, recent work from several labs suggest hat the guide RNAs may be dispersed throughout the nucleoplasm. For instance, the Gall lab has demonstrated that Drosophila cells lacking Cajal bodies are capable of modifying snRNAs (Deryusheva and Gall 2009). Work from our lab and others suggests that some specific guide RNAs may indeed reside within the nucleoplasm (Liang, Xu et al. 2002; Zhao, Li et al. 2002). More directly, it has been shown that an artificial guide RNA can direct nucleoplasmic pre-mRNA modification in yeast and mammalian cells (Semenov, Vratskih et al. 2008; Zhao and Yu 2008). Our current work also indicates that pseudouridylation can be introduced into pre-mRNA. All these lines of evidence suggest that the approach may in principle be used for gene silencing and regulation. Given the well known mechanisms of RNA-guided RNA modification and pre-mRNA splicing, it would be relatively simple and straightforward to design artificial
guide RNAs to target important uridines in pre-mRNA, thereby regulating pre-mRNA splicing *in vivo*.

### 2.4 Materials and methods

**Construction of various adenovirus pre-mRNAs, U2 snRNA and artificial box H/ACA guide RNAs**

T7 *in vitro* transcription was used to generate uniformly radiolabeled wild-type adenovirus pre-mRNA and mutant adenovirus pre-mRNA containing a U-to-G substitution at position 5 or 7 (U₅ or U₇) in the polypyrimidine tract, as previously described (Yu, Shu et al. 1998; Yu 2000). The full length of in vitro transcribed adenovirus pre-mRNA is 415 nucleotides (1-66, exon 1; 67-309, intron; 310-415, exon 2). T7 *in vitro* transcription was also used for generation of trace radiolabeled U2 snRNA for splicing reconstitution experiments, and trace radiolabeled artificial box H/ACA guide RNAs for targeted pre-mRNA pseudouridylation. Based on PugU2-34/44, a naturally-occurring Xenopus box H/ACA guide RNA (Zhao, Li et al. 2002), we constructed, via PCR, the T7-DNA template for transcription of artificial box H/ACA guide RNA. The 5’ pseudouridylation pocket of PugU2-34/44 (Figure 1B) was changed to fit the new target site, whereas the remaining nucleotide sequences remained unchanged.

Chimeric adenovirus pre-mRNAs, whose 5’ halves were derived from regularly transcribed pre-mRNA and 3’ halves were from pseudouridine-substituted pre-mRNA (or in reciprocal combination), were constructed according to previously published protocol (Figure 2.2B) (Yu, Shu et al. 1998). First, regular pre-mRNA and
pseudouridine-substituted pre-mRNA were synthesized via T7 in vitro transcription. Specifically, while regular pre-mRNA was transcribed in the presence of regular nucleotides (NTPs), in vitro synthesis of pseudouridine-substituted pre-mRNA was carried out under the similar conditions except the UTP was substituted with ψTP (Sierra Bioresearch., Tuscon, AZ) in the transcription reaction. Second, regular pre-mRNA and pseudouridine-substituted pre-mRNA were subjected to site-specific RNase H cleavage directed by 2'-O-methyl RNA-DNA chimeras. Finally, by mixing and matching, the 5’ half of regular pre-mRNA and the 3’ half of pseudouridine-substituted pre-mRNA (or reciprocally, the 5’ half of pseudouridine-substituted pre-mRNA and the 3’ half regular pre-mRNA) were aligned using an antisense bridging oligonucleotide, and ligated by T4 DNA ligase (Fermentas). Three 2’-O-methyl RNA-DNA chimeras were used to direct the cleavage at three positions: phosphodiester bond between nucleotides 2 and 3 with respect to the 5’ splice site (cleavage 1), phosphodiester bond between nucleotides-19 and -20 with respect to the 3’ splice site (cleavage 2), and phosphodiester bond between nucleotides -3 and -4 with respect to the 3’ splice site (cleavage 3). Upon ligation, three pairs of chimeric adenovirus pre-mRNAs were created.

To construct adenovirus pre-mRNA containing a 2’-F-uridine (IDT), 2’-H-uridine (IDT) or 2’-4’ locked uridine (C3’-endo configuration) (Kierzek, Ciesielska et al. 2005) at position 7 in the polypyrimidine tract, an in vitro transcribed 5’ fragment (nucleotides 1-295), an in vitro transcribed 3’ fragment (nucleotides 309-415), and a chemically synthesized middle-fragment oligoribonucleotide converting nucleotides
296-308 (5’-[^{32}P]UUUUU[dU]UUCACA-3’, where dU is 2’-deoxy-uridine; 5’-[^{32}P]UUUUU[FU]UUCACA-3’, where FU is 2’-F-uridine; 5’-[^{32}P]U[2’-4’U]UUUUUCCACA-3’, where 2’-4’U is a uridine locked in C3’-endo configuration, or 5’-[^{32}P]UUUUU[2’-4’U]UUCACA-3’, where 2’-4’U is a uridine locked in C3’-endo configuration), were aligned using a bridging DNA oligonucleotide complementary to nucleotides 278-329 of adenovirus pre-mRNA (5’ fragment: 3’fragment:middle fragment:bridge oligo= 1:1:4:1), and ligated together by T4 DNA ligase (Fermentas), essentially as described (Yu 2000).

Adenovirus pre-mRNA containing a pseudouridine at position 7 (Ψ7) within the polypyrimidine tract was constructed by two-pieces ligation. The 5’ piece (nucleotides 1-295) was in vitro transcribed, and the 3’ piece (nucleotides 296-319, where nucleotide 301 is equivalent to position 7 of the polypyrimidine tract, and nucleotides 310-319 correspond to the first 10 nucleotides of exon 2) was chemically synthesized (Dharmacon) and 5’ phosphorylated with [γ^{32}P]ATP. Upon hybridization with the bridging oligonucleotide (complementary to nucleotides 278-329), the two pieces of adenovirus pre-mRNA were the ligated by T4 ligase.

To produce a stable adenovirus polypyrimidine tract for competition assay, a T7-polypyrimidine tract template
(TAATACGACTCTATAGGGATGCAGTAACTGCATGGAAACCGCAGACCG AACGTCCTTTTTTTTCACAGCACCACCTCTCGACGAGTCTCGAGAGAATT) was generated by overlapping PCR. T7 transcription with UTP or ΨTP generated an
RNA containing, respectively, a regular polypyrimidine tract or a pseudouridine-substituted polypyrimidine tract, flanked by a 5’ and a 3’ stem-loop structure.

**Xenopus oocyte microinjection and reconstitution of pre-mRNA splicing**

Microinjection and reconstitution were performed essentially as previously described (Yu, Shu et al. 1998). For checking the splicing activity of a pre-mRNA, a single injection was used. Specifically, 9nL of $^{32}$P-uniformly radiolabeled or $^{32}$P-singly radiolabeled pre-mRNA was directly injected into the nuclei of *Xenopus* oocytes. After a one-hour incubation, the nuclei were isolated, RNA was recovered, and splicing was assayed on a 7% polyacrylamide-8M urea gel (acrylamide: bis = 19:1; EM Science).

For pseudouridylation-splicing reconstitution assay, multiple injections were used. Briefly, 46 nL of a mixture of 2 mg/mL antisense U2 DNA oligonucleotide (complementary to nucleotides 28-42) and 1 mg/mL in vitro transcribed artificial box H/ACA guide RNA, targeting uridines of the adenovirus pre-mRNA polypyrimidine tract, were injected into the cytoplasm of *Xenopus* oocytes. After an overnight (or ~ 4h) incubation, 9 nL of radiolabeled adenovirus splicing substrate (500,000 cpm/µL) was injected into the nucleus of oocytes. Following another overnight (~ 4h) incubation, 40 nL of 50 ng/µL in vitro transcribed U2 snRNA was injected into the cytoplasm of the oocytes to trigger splicing. After an overnight (or ~ 1h) reconstitution, nuclei were isolated. Upon proteinase K treatment, total nuclear RNA was extracted with PCA and then precipitated with ethanol. The recovered RNA was loaded on a 7% polyacrylamide-8M urea gel (acrylamide: bis =
Radiolabeled pre-mRNA and splicing products were visualized after autoradiography.

For competition assay, the procedure was essentially the same, with one slight modification. Specifically, at the step of injecting radiolabeled adenovirus pre-mRNA, and excess amount (~15-45-fold excess over labeled RNA) of trace-radiolabeled polypyrimidine tract (either regular or pseudouridine-substituted) was also injected.

**Analysis of splicing complexes by native gel electrophoresis**

Spliceosome assembly was analyzed according to previous reports (Konarska 1989; Yu, Shu et al. 1998). Briefly, 10 min after nuclear injection of 9 nL of radiolabeled adenovirus splicing substrate (500,000 cpm/µL) nuclei were isolated and broken by pipetting up and down several times in 10 µL of loading dye containing 9mM HEPES (pH 7.9), 22.5 mM KCl, 0.09mM EDTA, 0.22 mM DTT, and 1 mg/mL heparin. The samples were then loaded onto a 4% polyacrylamide native gel (acrylamide:bis = 80:1). The splicing complexes were visualized by autoradiography.

**Immunoprecipitation**

Anti-Sm (Y12) and Anti-U2AF65 immunoprecipitation were carried out essentially as described (Lerner, Lerner et al. 1981; Yu, Shu et al. 1998). Briefly, 9 nL of radiolabeled adenovirus splicing substrate (500,000 cpm/µL) were injected into the nuclei of Xenopus oocytes. After a 10 min-incubation, the oocyte nuclei were isolated and broken by pipetting up and down 20 times using P10 tips, followed by
vigorous mixing in the Net-2 buffer (50 mM Tris-HCl at pH 7.5, 150 mM NaCl, 0.05% NP-40). The sample was then clarified by centrifugation at 13,000g for 5 min, and the supernatant was mixed with anti-Sm antibodies (Y12) or Anti-U2AF<sup>65</sup> antibodies pre-bound to protein A-Sepharose. The mixture was then nutated for 2 h at 4°C. After a brief centrifugation, the beads were washed four times with Net-2 buffer and digested with proteinase K at 42°C in G50 buffer (20mM Tris-HCl at pH 7.5, 300 mM sodium acetate, 2mM EDTA, and 0.03% SDS). Co-precipitated RNA was recovered by PCA extraction and ethanol precipitation and analyzed on a 7% denaturing gel.

**U2AF<sup>65</sup>-polypyrimidine tract binding assay**

U2AF<sup>65</sup> was expressed in E. coli, and purified by ion-exchange chromatography, as described (Zamore and Green 1989; Valcarcel, Singh et al. 1993). The polypyrimidine tract of adenovirus pre-mRNA was synthesized in vitro (Dharmacon), and 5’-radiolabeled with [γ-32P]ATP and plynucleotide Kinase (Promega). Binding assay was carried out as described (Moore, Query et al. 1993; Sharp 1994). Briefly, a 10 µl-reaction mixture, containing 10 mM Tris-HCl, pH 7.5, 5% glycerol, 1 mM DTT, 50 mM KCl, 0.5 units/µl RNasin, 0.09 µg/µl BSA, 0.15 µg/µl carrier tRNA, protein dilutions (0, 5, 10, 20 ng/µl) and 1 fmol of 5’ radiolabeled RNA probe (polypyrimidine tract), was incubated at 30°C for 30 min. The reaction was then immediately loaded onto a 5% native polyacrylamide gel (bis:acrylamide = 1: 39). After electrophoresis, the polypyrimidine tract and polypyrimidine tract-U2AF<sup>65</sup> complex were visualized by autoradiography, and Kd
was calculated. The 5’ radiolabeled polypyrimidine tracts used in the binding assay were: U₇ probe, 5’-GUC CCU UUU UU₇U UCC ACA GCT CGC GGT TG-3’; Ψ₇ probe, 5’-GUC CCU UUU UΨ₇U UCC ACA GCT CGC GGT TG-3’.

**Pseudouridylation assay**

To assess the efficiency of pseudouridylation, pre-mRNA containing a single $^{32}$P in the polypyrimidine tract (5’ of a test uridine), rather than a uniformly radiolabeled pre-mRNA, was injected into the U2-depleted *Xenopus* oocytes containing an artificial box H/ACA guide RNA. Following an overnight (or ~4h) incubation, nuclei were isolated and total nuclear RNA was recovered by PCA extraction and ethanol precipitation. The recovered RNA was then digested with nuclease P1 (final 200 µg/mL in 3 µL of sodium acetate at pH 5.2) for 1h at 37°C. The digested sample was then dotted on TLC PEI plates (EM Science) and chromatographed in HCl-H₂O-isopropanol (15:15:70, v/v/v) buffer for ~ 6-7h. $^{32}$pU and $^{32}$pΨ were separated and visualized by autoradiography. The ratio of $^{32}$pU/$^{32}$pΨ was determined using a PhosphorImager (Molecular Dynamics).
Chapter 3

Pseudouridylation of yeast U1 snRNA may be catalyzed by an RNA-dependent mechanism
Abstract

Pre-mRNA splicing is an essential RNA processing reaction by which intervening sequences (introns) are removed and protein-coding sequences (exons) are ligated together. Pre-mRNA splicing is carried out by the spliceosome, a large RNA-protein complex consisting of 5 small uridine-rich RNAs (U1, U2, U4, U5 and U6 snRNA) and many proteins. U1 snRNA is involved in recognition of the 5' splice site (5'SS) through complementary base-pairing interactions. Interestingly, the 5' region of U1 snRNA responsible for the recognition of the 5'SS contains two post-transcriptionally modified uridines (Ψ5 and Ψ6). Furthermore, these two modifications are evolutionarily conserved from yeast to humans. Preliminary studies in S. cerevisiae indicate that Cbf5p, a pseudouridine synthase, and Nhp2p, one of the core proteins from box H/ACA RNPs, are involved in the formation of Ψ5 and Ψ6 in yeast U1 snRNA. An in vitro pseudouridylation assay also suggests that Ψ6 formation is RNA-dependent, further supporting the notion that this particular modification is introduced by a box H/ACA RNP. However, the set of predicted RNA candidates responsible for the two sites modifications are experimentally tested to be incorrect.
3.1 Introduction

A critical processing step in the maturation of pre-mRNA is pre-mRNA splicing, which is catalyzed via the spliceosome (Moore, Query et al. 1993). The spliceosome, a multiple-component complex, contains five phylogenetically conserved small nuclear RNAs (U1, U2, U4, U5 and U6 snRNA) and a large number of protein factors (Will and Luhrmann 2001). During spliceosome assembly, extensive interactions between snRNA and pre-mRNA occur, which facilitates two specific transesterification reactions that remove the intron and ligate the two exons together. The well characterized RNA-RNA interactions contain base pairing between the U1 snRNA 5’ invariant sequence with the 5’ splice site of the pre-mRNA (Zhuang and Weiner 1986); the base pairing between the highly conserved U2 snRNA branch site recognition region and the branch site of the pre-mRNA (Zhuang and Weiner 1989) and the base pairing between U2 and U6 snRNAs (U2-U6 helices, I, II, and III) (Hausner, Giglio et al. 1990). In addition to these Watson-Crick base-pairing interactions, the conserved loop of U5 snRNA interacts with both exon sequences at the 5’SS and 3’SS (Sontheimer and Steitz 1993). Additionally, 5’SS interacts with 3’SS themselves (Collins and Guthrie 2001). During the ordered and dynamic spliceosome assembly process, U1 and U4 snRNAs are displaced before the first step of splicing, additional conformational changes occur after the first step of splicing, leading to the second splicing step that generates mature mRNA (Yu, Scharl et al. 1999).

The initial step of pre-mRNA splicing is the recognition of the 5’SS by the U1
snRNP particle. In yeast, the 5'SS sequence is first recognized by base pairing with the 5' arm of U1 snRNA, and the U1 snRNP-associated proteins, including U1Cp, Prp40p, Nam8p, as well as the Sm complex (Kao and Siliciano 1996; Puig, Gottschalk et al. 1999; Zhang, Abovich et al. 2001). Base pairing between U1 snRNA and the conserved 5'SS of pre-mRNA is important for the commitment complex formation, however, Rosbash's lab has reported that some protein components of yeast U1 snRNP can interact with the 5'SS to some extent without the RNA-RNA base-pairing interactions (Du and Rosbash 2001). In the late spliceosome assembly process, U1 snRNA is replaced by U6 at the 5'SS through Prp28p DEAD box helices in an ATP-dependent manner (Staley and Guthrie 1999). Recently, for the first time natural occurring U1-independent pre-mRNA splicing in humans has been reported (Fukumura, Taniguchi et al. 2009).

All the five spliceosomal snRNAs (U1, U2, U4, U5, and U6 snRNA) contain extensive post-transcriptional modifications, including pseudouridylation and 2'-O methylation (Massenet, Mougin et al. 1998). Remarkably, these modifications are highly conserved across species and often cluster near functionally important regions of cellular RNAs involved in spliceosome RNA-RNA interactions, suggesting a possible functional role of these modifications in pre-mRNA splicing. U1 snRNA contains two pseudouridines and two 2'-O-methylated nucleotides normally presented at its 5' arm. The two pseudouridines at positions 5 and 6 are conserved from yeast to humans.
Experimental data accumulated thus far suggests that in higher eukaryotes pseudouridylation of spliceosomal snRNAs are catalyzed by box H/ACA RNPs on RNA-dependent mechanism (Yu, Terns et al. 2005). Each RNP contains four core proteins and a single RNA that serves as a guide for site-specific modification (Ganot, Bortolin et al. 1997; Ni, Tien et al. 1997). In *S. cerevisiae*, pseudouridylation of U2 snRNA can be catalyzed by either on RNA-dependent or on RNA-independent mechanism (Ma, Zhao et al. 2003; Ma, Yang et al. 2005). But the mechanism by which the other pseudouridylation sites in yeast spliceosomal snRNAs are catalyzed remains unknown.

In 1990’s, the functional roles of spliceosomal snRNA pseudouridylation have started to be discovered. Among the 5 U snRNAs, the molecular function of U2 snRNA pseudouridylation has been documented most extensively in vertebrate and *S. cerevisiae*. The 6 pseudouridines in the branch site recognition region and 3 pseudouridines within the 5’ arm of U2 are all required for snRNP assembly and pre-mRNA splicing in higher eukaryotes as determined by using a *Xenopus* oocyte reconstitution system (Yu, Shu et al. 1998; Zhao and Yu 2004). Under certain conditions, pseudouridine 35 in the branch site recognition region of yeast U2 snRNA is necessary for pre-mRNA splicing and cell growth (Yang, McPheeters et al. 2005). However, the functional significance of pseudouridylation of the other U snRNAs has yet to be investigated.

In yeast, a base-paring interaction between the 5’ end of U1 snRNA and the 5’SS of pre-mRNA is essential for pre-mRNA splicing. The necessity for this
interaction has been demonstrated in various ways. First, mutations in either the pre-mRNA 5’SS region or the 5’ end of U1 snRNA causes a dramatic decrease in commitment complex formation, inhibits subsequent steps of spliceosome assembly (Seraphin, Kretzner et al. 1988). Second, no U1 snRNP-independent commitment to splicing has been reported in yeast. Interestingly, there are two evolutionarily conserved pseudouridines in U1 snRNA, and both are located within the 5’SS recognition region. Pseudouridine 5 in U1 snRNA non-Watson-Crick interacts with U₄ of the 5’SS of pre-mRNA in yeast, but pairs with an A₄ in higher eukaryotes. Marion’s group has shown that the presence of this mismatch in the yeast U1 snRNP-pre-mRNA complex stabilized the interaction between the U1 snRNP and the 5’SS (Libri, Duconge et al. 2002). Furthermore, data from Krainer’s lab suggests that U1 snRNA pseudouridylation functions in 5’ splice site selection (Roca, Sachidanandam et al. 2005).

Our present work seeks to define the mechanism by which pseudouridine at positions 5 and 6 are introduced. Our preliminary results indicate that the pseudouridine synthase Cbf5p and one of the box H/ACA RNP core proteins Nhp2p are involved in the formation of both pseudouridines. Additionally, our data suggests that an RNA component is required for Ψ⁶ formation.

3.2 Results

Yeast Cbf5 point mutation cbf5D95A abolishes pseudouridylation in U1 snRNA (Ψ⁵, Ψ⁶) in vivo
The essential protein Cbf5p is the yeast pseudouridine synthase. Carbon’s group has shown that yeast strains expressing mutated cbf5 genes in a cbf5Δ null background are viable at 25°C, but display pronounced cold- and heat-sensitive growth phenotypes. Substitution of alanine for an aspartic acid residue in the conserved XLD motif of Cbf5p (cbf5D95A) abolishes pseudouridylation of rRNA in vivo (Zebarjadian, King et al. 1999). In order to investigate the mechanisms for pseudouridylation at position 5 and 6 in U1 snRNA, we took advantage of the cbf5D95A mutant yeast strain to determine the Ψ content in U1 snRNA. To map the pseudouridines, total RNA was extracted from the wild-type strain BY4741 and Cbf5 point mutant strain cbf5D95A, and pseudouridylation of endogenous U1 snRNA was analyzed by CMC modification followed by primer extension (Bakin and Ofengand 1993). In the wild-type strain, primer extension clearly detected both pseudouridines in U1 snRNA (Figure 3.1A, lane 2). However, in the cbf5D95A strain, both Ψ5 and Ψ6 formations are greatly diminished (Figure 3.1A, lane 4). We also assessed yeast U2 snRNA pseudouridylation as a positive control. U2 snRNA pseudouridylation at position 35 and 44 are modified via an RNA-independent mechanism, while Ψ42 in U2 snRNA is introduced by snR81 box H/ACA RNP (Ma, Zhao et al. 2003; Ma, Yang et al. 2005). As shown in Figure 3.1B, pseudouridylation at position 42 was dramatically reduced in cbf5D95A stain, while Ψ35 and Ψ44 formation remained unchanged compared to the wild-type strain. These results indicate that Cbf5p is involved in the pseudouridylation at both sites in yeast U1 snRNA.
Figure 3.1 The *cbf5*D95A mutant abolishes the formation of Ψ5 and Ψ6 in yeast U1 snRNA. (A) The yeast U1 snRNA pseudouridylation assay (CMC modification followed by primer extension) was performed by using total RNA isolated either from the wild-type strain (lane 1 and 2) or *cbf5*D95A mutant strain (lane 3 and 4). The lane 1 and 3, CMC modification was omitted. The lanes labeled A, G, C and U represents a yeast primer-extension sequencing ladder. The two primer-extension stops/pauses representing the two pseudouridines (Ψ5 and Ψ6) in U1 snRNA are indicated. (B) The yeast U2 snRNA pseudouridylation assay was conducted exactly the same as in panel A except that the primer-extension primer was complementary to yeast U2 snRNA. The three pseudouridines (Ψ35, Ψ42 and Ψ44) in U2 snRNA are indicated.
Depletion of either Nhp2p or Cbf5p precludes Ψ5 and Ψ6 formation in vivo

To further investigate whether pseudouridyllation of yeast U1 snRNA is RNA dependent, we analyzed U1 pseudouridyllation in vivo under conditions in which one of the box H/ACA RNP core proteins was depleted. Because all four core proteins of the box H/ACA RNP are essential, we utilized a conditional depletion approach. There are two strains in which the transcription of either NHP2 or CBF5 is under the control of the P_GAL promoter. The two mutant strains grew as well as the wild-type strain in galactose-containing medium. However, when we switched to glucose-containing medium, the Gal promoter was shut off, and Nhp2 or Cbf5 was gradually depleted in the mutant cells. The mutant cells grew significantly slower than the wild-type cells after about four generations (~10 h after medium switch). At three different time points (6, 16 and 48h) after the medium switch, both mutant cells were collected. Total RNA was extracted and pseudouridyllation of U1 snRNA was analyzed by CMC modification followed by primer extension.

As shown in figure 3.2A and 3.2C, when RNA was collected from cells that had been incubated for only 6h after the medium switch, primer extension clearly detected both two U1 pseudouridine stops/pauses, regardless of which strain was used. However, when RNA was isolated from nhp2 or cbf5 depletion strains that had grown for 16h or 48h after the medium switch, U1 pseudouridyllation was greatly diminished compared to the RNA isolated before the medium switch.
Figure 3.2 Depletion of either Nhp2p or Cbf5p specifically abolishes Ψ5 and Ψ6 formation in yeast U1 snRNA in vivo. (A) The U1 pseudouridylation assay (CMC modification followed by primer extension) was performed by using total RNA isolated from the cbf5-depletion strain that had been incubated in YPD for 0h (lane 1 and 2), 6h (lane 3), or 16h (lane 4), or 48h (lane 5). In lane 1, CMC modification was omitted. The two primer extension stops/pauses representing the two pseudouridines (Ψ5 and Ψ6) in U1 snRNA are indicated. (B) Pseudouridylation assay for yeast U2 snRNA was performed exactly the same as in panel A except that the primer used for primer extension assay was complementary to yeast U2 snRNA. The natural occurring pseudouridylation sites are indicated. (C) Pseudouridylation assay for U1 snRNA was performed exactly the same as in panel A, except that the total RNA was isolated from the nhp2-depletion strain. In lane 1, 3, 5 and 7, CMC modification was omitted. (D) Pseudouridylation assay for U2 snRNA was performed exactly as in panel C, except that the primer used for primer extension assay was complementary to yeast U2 snRNA.
We also assessed yeast U2 snRNA pseudouridylation at position 42, which is catalyzed by snR81 box H/ACA RNP (Ma, Yang et al. 2005). The signal of Ψ42 was dramatically reduced in *nhp2* or *cbf5* depletion strains that had been incubated for 16h and 48h after the medium switch (Figure 2B and 2D). The formation of Ψ35 and Ψ44, which are catalyzed by Pus7p (Ma, Zhao et al. 2003) and Pus1p (Massenet, Motorin et al. 1999) respectively, were essentially unchanged.

**Tandem affinity purification targeting Cbf5p co-purifies the pseudouridylase activity for formation of Ψ6 in yeast U1 snRNA**

With the evidence that the Cbf5 mutant strain abolishes pseudouridine formation at positions 5 and 6, and depletion of either Nhp2p or Cbf5p precludes both pseudouridylation *in vivo*, it is likely that a box H/ACA RNP complex is responsible for Ψ5 and Ψ6 formation in yeast U1 snRNA. To test this possibility, we took advantage of the TAP tagged Cbf5p prepared preparation in our lab to investigate its pseudouridylase activity for yeast U1 snRNA by performing a tandem affinity purification (TAP) procedure (Ma, Yang et al. 2005).

A synthetic U1 snRNA substrate with a single $^{32}$P label at the 5’ side of the uridine at position 6 was used. This single labeled U1 snRNA was incubated with the Cbf5-TAP preparation, and assayed for the pseudouridylation by thin layer chromatography (TLC). The pseudouridylation assay demonstrated that the Cbf5-TAP-tagged preparation could carry out pseudouridylation of uridine at position 6.
Figure 3.3 The formation of Ψ6 in yeast U1 snRNA requires an RNA component. (A) Yeast U1 snRNA containing a single $^{32}$P label at the 5’ side of uridine at position 6 was used to assess the pseudouridylase activity of the TAP preparation targeting Cbf5 (lane 2). Pseudouridylase activity of Cbf5-TAP preparation disappeared after treatment with micrococcal nuclease (MN) (lane 3). As a negative control, water was used in place of the TAP preparation (lane 1). (B) Yeast U2 snRNA containing a single $^{32}$P label at the 5’ side of uridine at position 42 was used. The pseudouridylase activity of Cbf5-TAP preparation was assessed exactly the same as in panel A.
(Figure 3.3A, lane 2). Given that a complete box H/ACA RNP contains an RNA guide as well as protein components, we hypothesize that yeast U1 snRNA pseudouridylation at position 6 may be catalyzed by an RNA-guided mechanism. To test this hypothesis experimentally, we treated the Cbf5-TAP preparation with micrococcal nuclease in the presence of Ca$^{2+}$. After inactivation of the nuclease by chelating Ca$^{2+}$ with EGTA, we performed the pseudouridylation assay using the U1 snRNA substrate containing a single $^{32}$P label at the 5’ side of the uridine at position 6. Unlike the untreated Cbf5-TAP preparation which was active in pseudouridylase activity at position 6 (Figure 3.3A, lane 2), the Cbf5-TAP preparation treated with micrococcal nuclease completely lost its pseudouridylase activity (lane 3). As a control, a reconstitution reaction was carried out using yeast U2 snRNA containing a single $^{32}$P label at the 5’ side of uridine at position 42, a naturally occurring pseudouridylation site known to be modified by snR81 box H/ACA RNP. As expected, the Cbf5-TAP preparation effectively converted uridine to pseudouridine at position 42 (Figure 3.3B, lane 1), whereas the micrococcal nuclease-treated Cbf5-TAP preparation did not catalyze this pseudouridylation reaction (Figure 3.3B, lane 2). Taken together, these results indicate that a complete box H/ACA RNP, including both RNA and protein components, is likely to constitute the pseudouridylase activity that catalyzes the formation of $\Psi_6$ in yeast U1 snRNA. According to these results we propose that pseudouridylation at position 6 in yeast U1 snRNA is carried out by an RNA-dependent mechanism.
None-essential proteins have no synergistic effect on pseudouridylation of U1 snRNA

Data from the *in vitro* experiment indicates that the uridine residue at position 6 is pseudouridylated when incubated with a Cbf5-TAP preparation. However, the extent of the conversion is not comparable to that of the pseudouridylation at position 42 in U2 snRNA. Thus the question arises whether proteins that loosely interact with the box H/ACA RNP are required to enhance the modification at this particular position in yeast U1 snRNA. To address this question we took advantage of the fact that all ORFs of the yeast genome have been systematically deleted. We screened 54 non-essential proteins that each is able to interact with one or more box H/ACA RNP core proteins to test their effects on pseudouridylation of yeast U1 snRNA (Table 3.1). Total RNA was isolated from both deletion strains and wild-type strain, and pseudouridylation of U1 snRNA was analyzed by CMC modification followed by primer extension. Compared to the wild-type strain, both Ψ5 and Ψ6 formation was unchanged (Figure 3.4 and data not shown). Collectively, these results suggest that the individual non-essential proteins screened have no synergerstic effect on pseudouridylation in U1. However, we can not exclude the possibility that a redundant mechanism may be required for the enhancement of the modification.

Computational predicted RNA candidates are not the guides for Ψ5 and Ψ6 in yeast U1 snRNA
Table 3.1 non-essential proteins that interact with the core proteins of box H/ACA RNP

<table>
<thead>
<tr>
<th>standard name</th>
<th>systematic name</th>
<th>Functional description</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPS24B</td>
<td>YIL069C</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>RPL4A</td>
<td>YBR031W</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>RPL13B</td>
<td>YMR142C</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>NPL3</td>
<td>YDR432W</td>
<td>mRNA binding; poly(A) RNA binding</td>
<td>Gar1p</td>
</tr>
<tr>
<td>KAP104</td>
<td>YBR107C</td>
<td>nuclear localization sequence binding</td>
<td>Gar1p</td>
</tr>
<tr>
<td>IMD2</td>
<td>YHR216W</td>
<td>chromatin binding; IMP dehydrogenase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>N/A</td>
<td>YCR016W</td>
<td>Unknown</td>
<td>Cbf5p;Nhp2p</td>
</tr>
<tr>
<td>COQ10</td>
<td>YOL008W</td>
<td>Ubiquinone binding</td>
<td>Nhp2p</td>
</tr>
<tr>
<td>RPL7B</td>
<td>YPL198W</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>MUS81</td>
<td>YDR386W</td>
<td>crossover junction endodeoxyribonuclease activity; endonuclease activity</td>
<td>Nhp2p</td>
</tr>
<tr>
<td>NSR1</td>
<td>YGR159C</td>
<td>RNA binding; single-stranded DNA binding</td>
<td>Cbf5p;Gar1p</td>
</tr>
<tr>
<td>SBP1</td>
<td>YHL034C</td>
<td>RNA binding; single-stranded DNA binding</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>SRO9</td>
<td>YCL037C</td>
<td>RNA binding; single-stranded DNA binding</td>
<td>Cbf5p;Nhp2p</td>
</tr>
<tr>
<td>HSP82</td>
<td>YPL240C</td>
<td>ATPase activity; unfolded protein binding</td>
<td>Cbf5p;Nhp2p</td>
</tr>
<tr>
<td>PDC5</td>
<td>YLR134W</td>
<td>pyruvate decarboxylase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>VIP1</td>
<td>YLR410W</td>
<td>inositol heptakisphosphate kinase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>RTF1</td>
<td>YGL244W</td>
<td>RNA polymerase II transcription elongation factor activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>HEK2</td>
<td>YBL032W</td>
<td>mRNA binding; poly(A) RNA binding</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>TIF4632</td>
<td>YGL049C</td>
<td>translation initiation factor activity</td>
<td>Cbf5p;Gar1p</td>
</tr>
<tr>
<td>FUN12</td>
<td>YAL035W</td>
<td>GTPase activity; translation initiation factor activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>PRL8B</td>
<td>YLL045C</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>GAL80</td>
<td>YML051W</td>
<td>specific transcriptional repressor activity; kinase inhibitor activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>RPS17A</td>
<td>YML024W</td>
<td>Structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>N/A</td>
<td>YMR031C</td>
<td>Unknown</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>standard name</td>
<td>systematic name</td>
<td>Functional description</td>
<td>Interactions</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>DBP3</td>
<td>YGL078C</td>
<td>ATP-dependent RNA helicase activity; RNA-dependent ATPase activity</td>
<td>Cbf5p; Gar1p; Nhp2p; Nop10p</td>
</tr>
<tr>
<td>N/A</td>
<td>YGR283C</td>
<td>Unknown</td>
<td>Cbf5p; Gar1p; Nhp2p</td>
</tr>
<tr>
<td>HTA2</td>
<td>YBL003C</td>
<td>DNA binding</td>
<td>Cbf5p; Gar1p; Nhp2p</td>
</tr>
<tr>
<td>N/A</td>
<td>YMR310C</td>
<td>Unknown</td>
<td>Cbf5p; Gar1p; Nhp2p</td>
</tr>
<tr>
<td>TIF4631</td>
<td>YGR162W</td>
<td>translation initiation factor activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>ACF2</td>
<td>YLR144C</td>
<td>glucan endo-1,3-beta-D-glucosidase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>KEM1</td>
<td>YGL173C</td>
<td>5'-3' exoribonuclease activity; deoxyribonuclease activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>UBP10</td>
<td>YNL186W</td>
<td>ubiquitin-specific protease activity</td>
<td>Cbf5p; Gar1p</td>
</tr>
<tr>
<td>RPS4A</td>
<td>YJR145C</td>
<td>structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>SAT4</td>
<td>YCR008W</td>
<td>protein kinase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>ROM2</td>
<td>YLR371W</td>
<td>phosphatidylinositol-4,5-biophosphate binding</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>MRP7</td>
<td>YNL005C</td>
<td>peptidyltransferase activity; structural constituent of ribosome</td>
<td>Gar1p</td>
</tr>
<tr>
<td>RPL16A</td>
<td>YIL133C</td>
<td>structural constituent of ribosome; RNA binding</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>HMO1</td>
<td>YDR174W</td>
<td>DNA binding activity</td>
<td>Nhp2p</td>
</tr>
<tr>
<td>JJJ1</td>
<td>YNL227C</td>
<td>ATPase activator activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>EST1</td>
<td>YLR233C</td>
<td>RNA binding; single-stranded DNA binding; telomeric DNA binding</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>GIN4</td>
<td>YDR507C</td>
<td>protein kinase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>RPL27A</td>
<td>YHR010W</td>
<td>structural constituent of ribosome</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>CHD1</td>
<td>YER164W</td>
<td>ATPase activity; RNA polymerase II transcription elongation factor activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>MIS1</td>
<td>YBR084W</td>
<td>formate-tetrahydrofolate ligase activity; methylenetetrahydrofolate cyclohydrolase activity</td>
<td>Gar1p</td>
</tr>
<tr>
<td>NAT1</td>
<td>YDL040C</td>
<td>Contributes to peptide alpha-N-acetyltransferase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>LHP1</td>
<td>YDL051W</td>
<td>RNA binding; tRNA binding</td>
<td>Gar1p</td>
</tr>
<tr>
<td>PUF6</td>
<td>YDR496C</td>
<td>mRNA binding; specific transcriptional repressor activity</td>
<td>Cbf5p; Gar1p</td>
</tr>
<tr>
<td>HMT1</td>
<td>YBR034C</td>
<td>protein-arginine N-methyltransferase activity</td>
<td>Gar1p</td>
</tr>
<tr>
<td>standard name</td>
<td>systematic name</td>
<td>Functional description</td>
<td>Interactions</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------</td>
<td>------------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>N/A</td>
<td>YPL009C</td>
<td>Unknown</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>SNU66</td>
<td>YOR308C</td>
<td>Unknown</td>
<td>Nhp2p</td>
</tr>
<tr>
<td>IMD3</td>
<td>YLR432W</td>
<td>IMP dehydrogenase activity</td>
<td>Cbf5p</td>
</tr>
<tr>
<td>DSE4</td>
<td>YNR067C</td>
<td>glucan endo-1,3-beta-D-glucosidase activity</td>
<td>Nhp2p</td>
</tr>
<tr>
<td>NOP6</td>
<td>YDL213C</td>
<td>RNA binding</td>
<td>Cbf5p;Gar1p;Nhp2p</td>
</tr>
<tr>
<td>STM1</td>
<td>YLR150W</td>
<td>DNA binding;telomeric DNA binding</td>
<td>Nhp2p</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>rps24bΔ</th>
<th>rpl4aΔ</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMC</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

**Figure 3.4** Depletion of non-essential protein RPS24B or RPL4A which interacts with Cbf5p does not abolish pseudouridylation in yeast U1 snRNA *in vivo*. The U1 pseudouridylation assay (CMC modification assay followed by primer extension) was performed using total RNA isolated either from the wild-type strain, *rps24b*-depletion strain or *rpl4a*-depletion strain that had been incubated in YPD until $A_{600}$ reached 2.0-3.0. When appropriate, CMC treatment is indicated by – (untreated) and + (treated) for RNA from all strains. The two primer extension stops/pauses representing the two pseudouridines in U1 snRNA are indicated.
Table 3.2 Predicted RNA candidates for pseudouridylation in yeast U1 snRNA by snoGPS

<table>
<thead>
<tr>
<th>standard name</th>
<th>Ch</th>
<th>predicted length (nt)</th>
<th>expression verified</th>
<th>target site</th>
<th>snoGPS score</th>
</tr>
</thead>
<tbody>
<tr>
<td>snR35</td>
<td>XV</td>
<td>204</td>
<td>Yes</td>
<td>Ψ6</td>
<td>18.67</td>
</tr>
<tr>
<td>NR</td>
<td>VIII</td>
<td>167</td>
<td>N/A</td>
<td>Ψ6</td>
<td>32.98</td>
</tr>
<tr>
<td>NR</td>
<td>V</td>
<td>233</td>
<td>N/A</td>
<td>Ψ5</td>
<td>27.10</td>
</tr>
<tr>
<td>NR</td>
<td>V</td>
<td>142</td>
<td>N/A</td>
<td>Ψ5</td>
<td>25.27</td>
</tr>
<tr>
<td>NR</td>
<td>VIII</td>
<td>157</td>
<td>N/A</td>
<td>Ψ5</td>
<td>25.66</td>
</tr>
</tbody>
</table>

‘Ch’ indicates chromosome; NR is for not reported

Figure 3.5 Examining the U1 snRNA pseudouridylation after disrupting the genes of predicted RNA candidates. The presence of Ψ5 and Ψ6 was monitored by CMC modification followed by primer extension. Results are shown from 6 strains, each depleted of a different predicted RNA candidate or the wild-type strain, BY4742. When appropriate, CMC treatment is indicated by – (untreated) and + (treated) for RNA from all strains. Positions of pseudouridine residues are indicated as Ψ5 and Ψ6. (A) Predicted RNA candidate snR35 for Ψ6. (B) Predicted RNA candidates for Ψ5. (C) Predicted RNA candidates for either Ψ5 or Ψ6.
As indicated in both *in vivo* and *in vitro* studies, formation of $\Psi_5$ and $\Psi_6$ may be RNA-dependent. In order to find the guides for modifying these two pseudouridines, Dr. Peter Schattner used a computational approach (snoGPS), first screened all known *S cerevisae* box H/ACA RNAs against uridines at position 5 and 6 in U1 snRNA, and found that the target uridine at position 6 matches with snR35, which is known to be responsible for converting U1191 to $\Psi_{1191}$ in 18S rRNA. No target-guide match was found between uridine at position 5 and known H/ACA RNAs. However, there is a known example that snR82 is capable of modifying two nearby uridine residues at position 2349 and 2351 in 25s rRNA, it is possible that snR35 can modify both sites. To test whether the formation of $\Psi_5$ and $\Psi_6$ in U1 snRNA is dependent on snR35 *in vivo*, a snR35-deletion strain (with no growth defect phenotype) was created. Total RNA was isolated from the deletion strain and the wild-type strain, and pseudouridylation of U1 snRNA was analyzed by CMC modification followed by primer extension. As shown in Figure 3.5A, the formation of both $\Psi_5$ and $\Psi_6$ in U1 snRNA was not affected in the snR35 deletion strain.

Next, Dr. Schattner screened the whole *S cerevisae* genome against uridines at position 5 and 6 in U1 snRNA by using the same computational approach, and found several candidates (Table 3.2). To test whether these candidates are correct, we constructed different deletion strains: a particular deletion strain with a particular predicted candidate RNA gene deleted. Again, pseudouridylation was analyzed by CMC modification followed by primer extension. As shown in Figure 3.5B and 3.5C,
both \( \Psi_5 \) and \( \Psi_6 \) in U1 snRNA were detected in all the deletion strains as well as the wild-type strain. These results indicate that the predictions for pseudouridylation at position 5 and 6 in U1 snRNA are not correct.

### 3.3 Discussion

We have demonstrated that pseudouridylation of yeast U1 snRNA at position 5 and 6 is catalyzed by an RNA-dependent mechanism. Using a box H/ACA RNP pseudouridylase synthase Cbf5p point mutation strain, we demonstrate that the pseudouridylase activity is Cbf5p-dependent (Figure 3.1). Using the \( nhp2 \) or \( cbf5 \) conditional depletion strain, we further demonstrate that the two core protein components of the box H/ACA RNP are required for both modification sites (Figure 3.2). The \textit{in vitro} pseudouridylation TAP experiment showed that the Cbf5-TAP preparation copurified with the pseudouridylase activity at position 6, but the activity was lost after micrococcal nuclease digestion (Figure 3.3). These data further suggests that pseudouridylation at position 6 is RNA-dependent. Finally, the pseudouridylation assay showed that all candidate RNAs as predicted through snoGPS for pseudouridylation at position 5 and 6 through snoGPS computational program are not correct (Figure 3.5). Our results indicate that the conversion of U to \( \Psi \) in yeast U1 snRNA is most likely catalyzed by a box H/ACA RNP via an RNA-guided mechanism. If it is true, the mechanism will likely be similar to the previously identified RNA-dependent mechanism catalyzing yeast U2 pseudouridylation at
position $\Psi_{42}$. However, the responsible guides for the two modifications remain unknown.

**Spliceosomal snRNA Pseudouridylation**

The spliceosomal U snRNAs, which are necessary for pre-mRNA splicing, have long been known to contain a large number of post-transcriptional modifications, especially pseudouridylation. Many of the pseudouridines are located in similar regions across species. There are 24 known sites of pseudouridylation in human major spliceosomal snRNAs. Computational and experimental data indicates that pseudouridylation of spliceosomal snRNAs is RNA-dependent (by Box H/ACA RNPs) in higher eukaryotes. Indeed, through size fractionation of RNAs, co-immunoprecipitation with antibodies against box H/ACA core proteins, and bioinformatics-based approaches, 16 of the 24 known sites of pseudouridylation within the 5 major spliceosomal snRNAs have been proven or predicted to be catalyzed by RNA-dependent mechanism in various vertebrate organisms (Huttenhofer, Kiefmann et al. 2001; Vitali, Royo et al. 2003; Kiss, Jady et al. 2004; Schattner, Decatur et al. 2004; Yu, Terns et al. 2005; Schattner, Barberan-Soler et al. 2006). Verified by both *in vivo* and *in vitro* modification experiments, box H/ACA RNA U85 in human and fruit fly is responsible for converting U46 to $\Psi_{46}$ in U5 snRNA (Jady and Kiss 2001). Another box H/ACA RNA, U89, was predicted to be able to direct the same site pseudouridylation as U85 in human U5 snRNA (Darzacq, Jady et al. 2002). In *S. cerevisiae*, however, only 6 pseudouridine residues were
found. Unlike that of higher eukaryotic spliceosomal snRNAs, pseudouridylation of yeast spliceosomal snRNAs can be catalyzed by either an RNA-dependent mechanism or an RNA-independent mechanism, as identified for yeast U2 pseudouridylation. $\Psi_{35}$ and $\Psi_{44}$ of U2 snRNA are the two sites catalyzed via protein-only mechanism by pseudouridylation synthase Pus7p (Ma, Zhao et al. 2003) and Pus1p (Massenet, Motorin et al. 1999) respectively. However, pseudouridylation at position 42 is catalyzed by snR81 box H/ACA RNP in an RNA-dependent manner (Ma, Yang et al. 2005). There are two conserved pseudouridines located at the 5’ arm of U1 snRNA. Our current data shows the first indication that pseudouridylation in yeast U1 snRNA is RNA-dependent.

**Pseudouridylation at position 5 in yeast U1 snRNA**

In our current work, we observed that pseudouridylation of U₅ in yeast U1 snRNA is Cbf5p and Nhp2p-dependent. We hypothesize that $\Psi_5$ is introduced by an RNA-dependent mechanism, because both Cbf5p and Nhp2p are the core proteins of a box H/ACA RNP. However, we cannot rule out the possibility that this modification may be catalyzed without the use of a guide RNA. There are couple reasons. First, all the predicted RNA candidates responsible for formation of $\Psi_5$ are tested to be incorrect. Second, it has been reported that in archaea purified Cbf5 by itself is capable of modification of tRNA U55 *in vitro*, but the pseudouridylase activity of Cbf5 is enhanced in the presence of the box H/ACA RNP complex (Roovers, Hale et al. 2006).
Pseudouridylation at position 6 in yeast U1 snRNA

Our Tap experiment suggests that pseudouridylation at position 6 in yeast U1 snRNA requires an RNA component (Figure 3.3). Interestingly, the modification of $\Psi_6$ observed is much lower compared to that of $\Psi_{42}$ in yeast U2 snRNA. Besides four core proteins, there are some proteins loosely bound with box H/ACA RNP by interacting with one or more of the core proteins. It is possible that our expressed Cbf5-TAP preparation lost some loosely bound proteins which may be required for this specific pseudouridylase activity. However, our single deletion experiment showed that none of the non-essential proteins screened has any effect. One explanation is that we only deleted one protein gene in one yeast strain. It is possible that more than one specific protein needs to incorporate into the complex to enhance the pseudouridylase activity. Another alternative explanation for the low efficient pseudouridylase activity could be that the required protein(s) is essential. The essential protein Naf1p has been suggested to function as a yeast H/ACA snoRNP assembly factor and shows a weak association with the mature H/ACA snoRNAs in vivo (Yang, Rotondo et al. 2002). Further study is necessary to clarify these possibilities.

Why predicted RNA candidates are not correct?

Both our in vivo and in vitro data suggests that pseudouridylation of yeast U1 snRNA may require a box H/ACA RNP. However, computational identified RNA candidates for $\Psi_5$ and $\Psi_6$ are experimentally tested to be incorrect. There are several
explanations for why disrupting the coding sequences of these RNAs individually did not block modification. One is that the modifications may be mediated by redundant mechanisms. Such redundancies might involve more than one RNA candidate acting with one or more box H/ACA RNA-dependent pseudouridine synthases. This issue could be addressed by disrupting all candidate RNAs in the same yeast strain. Another possibility could be that two mechanisms may co-exist as was discovered for the 2′-O-methylation of LSU-U2918 in *S. cerevisiae* (Bonnerot, Pintard et al. 2003). Besides, it is possible that the guides for pseudouridylation of U1 snRNA were not yielded by snoGPS approach. The guide RNA may have some unusual structural feature(s) that causes it to elude the program. Combination with other computational methods may enhance the possibilities to identify the potential guides.

Besides U1 and U2 snRNA in *S. cerevisiae*, U5 snRNA contains one post-transcriptionally pseudouridylated modification at position 99. It will be interesting to characterize the mechanism by which this pseudouridine in catalyzed. All the pseudouridines in yeast U2 snRNA are functionally important (Yu, Shu et al. 1998; Zhao, Li et al. 2002; Zhao and Yu 2004; Yang, McPheeters et al. 2005). The two pseudouridines in U1 snRNA are located in the 5′ arm and base pair with 5′SS of premRNA during the early stage of spliceosome assembly. It will be very interesting to address the functional significance of these two pseudouridines.

### 3.4 Materials and Methods

**Conditional Depletion (Promoter Shutoff) Assay**
Yeast strain transformed with a plasmid containing either Cbf5 gene or Nhp2 gene under the control of the P\textsubscript{Gal} promoter was grown in YPGAL (1% yeast extract, 2% galactose, 2% peptone) at 30°C until A\textsubscript{600} reached 2.0-3.0. At this point, the yeast cells were pelleted and resuspended in YPD. At different time points after switching the medium, cell growth was monitored by A\textsubscript{600}, and a fraction of the cells was collected. Total RNA was extracted from these cells and analyzed for U1 pseudouridylation.

**Pseudouridylation assays**

The U1 pseudouridylation assay was performed exactly as described previously (Ma, Zhao et al. 2003). In brief, total RNA was isolated using Trizol (Invitrogen). Extracted total yeast RNA was modified with CMC (N-cyclohexyl-N’-(2-morpholinoethyl)-carbodiimide metho-p-toluolsufonate) at 37°C for 20 min in a 30 µl reaction containing 10 µg of total RNA, 0.17 M CMC, 50 mM Bicine (pH 8.3), 4 mM EDTA, and 7 M urea. The modified RNA was then treated with sodium bicarbonate buffer (50 mM, pH 10.4) at 37°C for 2h. RNA was recovered by PCA extraction and ethanol precipitation and was subsequently subjected to primer extension analysis using a 5’ \textsuperscript{32}P-radiolabeled oligodeoxynucleotide complementary to nucleotides 55-77 of \textit{S. cerevisae} U1 snRNA. CMC modification-specific stops/pauses were visualized by autoradiography of 6% polyacrylamide/8M urea gels.

**Construction of single labeled substrates**
The U1 substrate used for the \textit{in vitro} TAP pseudouridylation study was constructed via three-piece ligation. Specifically, there are three pieces of RNAs: a middle 15-mer chemical synthesized RNA oligonucleotide corresponding to the target uridine and its immediate upstream sequence, a 5’ 10-mer chemical synthesized RNA oligonucleotide corresponding to the sequence immediately downstream of the target uridine and a 3’ piece of \textit{in vitro} transcribed RNA corresponding to the sequence immediately upstream of the middle synthesized RNA. The middle piece of 15-mer RNA was phosphorylated with $\gamma^{32}$P ATP and plynucleotide kinase. The 5’ end labeled 15-mer RNA was then aligned with the 5’ 10-mer and 3’ transcribed RNA through a complementary DNA bridging oligonucleotide, and subsequently ligated together by DNA ligase. The RNA substrate thus synthesized contained a single $^{32}$P 5’ of the target uridine at site 6 of U1 snRNA. The U2 substrate was prepared exactly as described previously (Ma, Yang et al. 2005).

\textit{In vitro} TAP pseudouridylation experiment

Micrococcal nuclease-treated or untreated Cbf5-TAP-tagging preparations were used to reconstitute pseudouridylation \textit{in vitro} with various RNAs. About 1000 cpm (<0.001 fmol) of single $^{32}$P-radiolabeled yeast U1 snRNA at the 5’ side of U6 or U2 snRNA at the 5’ side of U42 was mixed with 2 $\mu$l of each micrococcal nuclease-treated TAP-tagging preparation or original TAP-tagging preparation that had not been treated with micrococcal nuclease. As a control, we used the same volume of water in place of the TAP preparation. The mixture was then brought to a final
volume of 20 µl containing 100 mM Tris-HCl, pH 8.0, 100 mM ammonium acetate, 5 mM MgCl₂, 2 mM DTT and 0.1 mM EDTA. The pseudouridylation reaction was performed at 30°C for 1h. The single radiolabeled U1 or U2 snRNA was then recovered via PCA extraction and ethanol precipitation, and was subsequently subjected to P1 nuclease digestion and TLC analysis, exactly as described (Ma, Zhao et al. 2003).

Construction of yeast predicted RNA deletion strain

All predicted RNAs responsible for pseudouridylation at site 5 and 6 in yeast U1 were provided by Dr. Peter Schattner. To delete each predicted RNA gene, PCR was used to generate a DNA fragment containing the predicted RNA-Δ:Kan′ cassette and predicted RNA-flanking sequences (-45 bp to +45 bp, relative to the 5′ and 3′ ends of the predicted RNA gene, respectively). Specifically, the template for the PCR reaction was the genomic DNA isolated from strain 24124. The 5′ primer was a 60-mer oligodeoxynucleotide corresponding to the 45 nucleotides immediately upstream of the guide RNA gene and the first 15 nucleotides of a KanMX4 cassette, and the 3′ primer was a 60-mer oligodeoxynucleotide complementary to the 45 nucleotide sequence immediately downstream of the guide RNA gene and the last 15 nucleotides of KanMX4 gene. The PCR product was transformed into the wild-type strain BY4741, and the predicted RNA-deletion strain was selected on a YPD medium containing 0.2 mg/ml geneticin. The deletion of the guide RNA was verified by PCR analysis using the 5′ primer corresponding to a sequence upstream of the
predicted RNA gene and either of the two 3’ primes complementary to a KanMX4 (KanB, 5’ GCCGCCTGCAGTACTCTTTATCATGCAAGAAG3’) sequence or a predicted RNA sequence, respectively. A PCR product with predicted size was generated when the 5’ predicted RNA verify primer and the 3’ KanB primer were used; no product was observed when the 3’ KanB was replaced by predicted RNA extension primer (data not shown).

All the PCR deletion primers for the predicted RNAs are listed here:

U1.U5_142ntΔF1: 5’ TGGGTGCCTAAAAACTTTATTTTTCGTGAGGCGGTGGGAG
AAATAAGGCTTTTCAATTCAT 3’

U1.U5_142ntΔR1: 5’ GAACAGTTTACGTAAAGTTACCTGTTTTGTTGAGGCTCAT
TCTACAGCGATGATAAGCTTG3’

UPS U1.U5_142ntΔF1: 5’ GAGAACTTTATTTTCTGC3’

UPS U1.U5_142ntΔR1: 5’ GAAAATGGCGAGCTGA3’

U1.U5_187ntΔF1: 5’ GAGATTATAGTAGTACCAATTTTCAATTACTGCAATTACT
TATTCAAGGCTTTTCAATTCAT3’

U1.U5_187ntΔR1: 5’ TTTAAAGTGTGGCCAGAACGCTTCTTTGAACCTTGATTG
AATTTAAAGGCTTTGATGCTG3’

UPS U1.U5_187ntΔF1: 5’ CAACATCCAGGAGACA3’
UPS U1.U5_187nt\(\Delta\)R1: 5’CTTCCTAATTG TAGTCGC3’

U1.U5_233nt\(\Delta\)F1: 5’ATAGATTTACATATATG TAGTATTTTATCAATT TATTATTGAT CAAAATGCTTT TCAATTCAT3’

U1.U5_233nt\(\Delta\)R1: 5’TGAGTTTTAATGAGT CCAGATCACTTT TAGTACG GTT TAAATTGC GATGATAAGCTGTC3’

UPS U1.U5_233nt\(\Delta\)F1: 5’CAAACC CTTTT TCTTTAC3’

UPS U1.U5_233nt\(\Delta\)R1: 5’CTGGATGT TGATCACAAATG3’

U1.U5_157nt\(\Delta\)F1: 5’AGGAGATGAAGC CTTTAC TCAAGTCACATCAA AATAATCGA AGCCAACAGC TTTTCAATT CAT3’

U1.U5_157nt\(\Delta\)R1: 5’GACTTCCTACTAGACC CAGATATAATG GATACCACG GTATACACGATG GATAAGCTGTC3’

UPS U1.U5_157nt\(\Delta\)F1: 5’CTTGGAGGCTTAATC TA3’

UPS U1.U5_157nt\(\Delta\)R1: 5’CTTATGCTGCTTT AAAGT3’

U1.U6_167nt\(\Delta\)F1: 5’CCAAGGAAGAGTAAATGC CCTCGAGCC ACGCATTGTGGT GCCTTTT TCAATTCAAT3’

U1.U6_167nt\(\Delta\)R1: 5’TTGGCCGACGACTGGT TCTGGGCCGGGCTCGAGCCGGAA GTGGTAACGATGATAAGCTGTC3’
UPS U1.U6_167ntΔF1: 5' TCACAGTAAGCACATCT3'  

UPS U1.U6_167ntΔR1: 5' GGTAACGCTGTTCCAAG3'
Chapter 4

Conclusion
4.1 A flexible RNA backbone within the polypyrimidine tract is required for U2AF<sup>65</sup> binding and pre-mRNA splicing *in vivo*

Our study identified two uridine residues, U<sub>5</sub> and U<sub>7</sub>, in the polypyrimidine tract (a total of 9) of adenovirus pre-mRNA that were sensitive to pseudouridine substitution. This inhibitory effect is pseudouridine-specific and happens at an early stage of spliceosome assembly. Furthermore, our data suggests that the U-to-Ψ change in the polypyrimidine tract causes a change in sugar pucker configuration at position 7, resulting in a defect in U2AF<sup>65</sup> binding, thus inhibiting pre-mRNA splicing. Moreover, we demonstrate that the inhibitory effect of U-to-Ψ isomerization on pre-mRNA splicing can be applied to different polypyrimidine tracts, despite the fact that their sequences are quite different.

The crystal structure of a U2AF<sup>65</sup> RNA binding domain bound to a seven-uridine tract (U<sub>1</sub>U<sub>2</sub>U<sub>3</sub>U<sub>4</sub>U<sub>5</sub>U<sub>6</sub>U<sub>7</sub>) RNA indicates that the sugar pucker of the first three uridines adopt a C2’-endo sugar configuration. As our data indicates that a C2’-endo configuration at position 7 is important for pre-mRNA splicing, it would be interesting to determine the crystal structure of a U2AF<sup>65</sup> RNA binding domain bound to the polypyrimidine tract of adenovirus pre-mRNA in the presence of uridine at position 7. The crystallographic analyses will likely provide insight to the mechanism behind the inhibitory effects of Ψ on U2AF<sup>65</sup> binding to the polypyrimidine tract.

Pseudouridine differs from its parental uridine in that it has an additional hydrogen bond donor. This additional hydrogen bond donor allows pseudouridine to
coordinate a water molecule between its N1-H and the 3’ phosphate of its preceding residue (Auffinger 1998). The water molecule strongly favors a C3’-endo sugar configuration, thus it is reasonable to ask whether the phosphate oxygen of the Ψ at position 7 is functional during pre-mRNA splicing. In order to address this question, we could use a phosphorothioate substitution interference strategy to substitute the phosphate oxygen at position 7. The sulfur substitution has minimal effects on RNA secondary structure, however, it can affect hydrogen bonding. If the interaction between the water molecule and the phosphate of Ψ7 is critical for splicing, the phosphorothioate substitution will alter the bond, thus reversing the inhibitory effect of U-to-Ψ substitution on pre-mRNA splicing.

While our results indicate that a C3’-endo sugar configuration of pseudouridine causes a U2AF65-binding defect, thus inhibiting pre-mRNA splicing, these analyses were all in the adenovirus pre-mRNA. In order to exclude the possibility that the intronic and/or exonic sequences of adenovirus pre-mRNA are necessary for the inhibitory effects observed, we could first construct chimeric RNAs in which the adenovirus polypyrimidine tract is placed in the context of a different pre-mRNA. Following the expression of the chimeric RNA we can pseudouridylate U7 and determine whether U2AF65 binding is still inhibited. If each substrate shows a defect of U2AF65-binding and splicing in the presence of Ψ7, this will indicate that the inhibitory effect from U7-to-Ψ7 change is not dependent on intronic or exonic sequences, but the polypyrimidine tract.
We found that pseudouridylation at position U\(_2\) in the \(\beta\)-globin polypymidine tract has an inhibitory effect on pre-mRNA splicing, however, the important uridines for splicing in the adenovirus pre-mRNA polypyrinididine tract are at sites U\(_5\) and U\(_7\). As the experiments were performed with a pre-mRNA substrate having the \(\beta\)-globin polypymidine tract within the adenovirus pre-mRNA context, it is yet to be determined if with a substrate having the \(\beta\)-globin polypymidine tract within its own pre-mRNA will show the same effect from a U\(_2\)-to-\(\Psi\)\(_2\) change. Thus, it will be necessary to dissect the contribution of \(\Psi\) in the \(\beta\)-globin polypymidine tract within its native context, and see whether the major effect is still at position 2. In addition, it will be important to have the crystal structures of U2AF\(^{65}\) bound to the \(\beta\)-globin polypymidine tract within its native pre-mRNA in the presence of U\(_2\) to determine whether a C2’-endo sugar configuration is similarly required for U2AF\(^{65}\) binding.

In our study, we used a new strategy named RNA-guided-RNA modification to target uridines in the polypymidine tract. This approach can introduce pseudouridylation in a very efficient and site-specific way. RNA pseudouridylation can alter the RNA backbone structure, thus leading to an inhibitory effect on pre-mRNA splicing. Many RNAs function properly if they maintain their natural structures. If we introduce pseudouridylation into these RNAs, we may impair their normal functions. Thus our RNA-guided-RNA modification will be a very useful tool to regulate RNA functions \textit{in vivo}. 

4.2 Pseudouridylation of U1 small nuclear RNA may be catalyzed by an RNA-dependent mechanism in *Saccharomyces cerevisiae*

In *S. cerevisiae*, there are two pseudouridines (Ψ5 and Ψ6) located at the 5’ end of U1 snRNA. Here, we demonstrate that both Cbf5p and Nhp2p are required for the pseudouridylation of yeast U1 snRNA. In addition, using an *in vitro* pseudouridylation assay, pseudouridylation of position U6 of U1 snRNA appears to be dependent on an RNA component(s). Thus, we believe that the Ψ6 (and perhaps Ψ5 as well) is likely introduced by the RNA-dependent mechanism.

However, in our *in vitro* pseudouridylation reconstitution assay, modification at position U6 was not nearly comparable to that of U42 in yeast U2 snRNA. It is possible that when we prepared the TAP preparation, important factors required for modification of U6 were lost during the purification. To test this possibility, we could mix crude yeast extract or total protein isolated from yeast cells with the Cbf5-TAP preparation to add back those lost factors. If the modification activity is restored, this would indicate that another protein component is necessary for U6 modification. If the crude yeast extract can enhance the efficiency of the modification, we will be more confident that Ψ6 formation is box H/ACA RNP-dependent.

Regarding whether an RNA component(s) is involved in the modifications of U5 and U6, Dr. Peter Schattner (University of California Santa Cruz) has predicted several H/ACA RNA candidates involved in formation of Ψ5 and Ψ6.
However, deletions of all individual candidates still result in modification of $U_5$ and $U_6$. It is possible that U1 snRNA pseudouridylation is catalyzed by multiple mechanisms, i.e. more than one guide RNA. In order to clarify this possibility, we could construct deletion strains with two or more candidates deleted and analyze U1 snRNA pseudouridylation. In addition, it is also possible that the correct H/ACA RNA was not predicted by Dr. Schattner’s approach. Thus, we can take advantage of a previously built Box HACA RNA library and test the candidates from the library for their ability to reconstitute U1 pseudouridylation at position 5 and 6 in the micrococcal nuclease-treated Cbf5-TAP preparation.

Lastly, it has been reported that 2’-O-methylation of LSU-U2918 is introduced by two co-existing mechanisms (both RNA-independent and RNA-dependent mechanisms) (Bonnerot, Pintard et al. 2003). Although we found that pseudouridylation at position $U_5$ and $U_6$ of U1 snRNA are Cbf5p and Nhp2p-dependent, we cannot rule out the possibility that other pseudouridine synthases also modify these positions. In order to determine whether other pseudouridine synthases are involved in U1 snRNA, it is worth the effort to construct deletion strains of all known pseudouridine synthases in *S. cerevisiae* and determine whether U1 snRNA is still modified in these strains.

### 4.3 Concluding remarks

The results of this thesis provide insight into the various fields of RNA biology, in particular pre-mRNA splicing and RNA modification. In this regard, we expanded the targets of H/ACA RNAs, that is, we demonstrated that pre-messenger
RNA is a suitable target for H/ACA RNA-mediated pseudouridylation in *Xenopus* oocytes. Using H/ACA RNA-mediated pseudouridylation as a tool we elucidated an *in vivo* requirement for backbone flexibility within the polypurimidine tract during 3’-splice site selection of pre-mRNA splicing. While it is not fully understood as to why a rigid backbone inhibits U2AF^{65} binding, our study demonstrates the need to carry out further detailed analyses of U2AF^{65} binding to various polypurimidine tracts. In addition, this thesis work started to unravel the mechanism behind U1 snRNA pseudouridylation.

While significant progress is being made at elucidating the mechanism of pre-mRNA splicing as well as determining the enzymes required for snRNA pseudouridylation, several questions exist. For instance, is backbone flexibility required at other locations of the pre-mRNA or snRNAs during intron removal? Does pseudouridylation naturally exist in messenger RNA? In addition, the mechanism of U1 and U5 snRNA pseudouridylation has not been fully addressed. Finally, what are the functional roles of U1 and U5 snRNA pseudouridylation during pre-mRNA splicing? We believe that the experimental approaches and assays used here will be useful in addressing these questions.
References


