Molecular Cell *23*, 49–59, July 7, 2006 ©2006 Elsevier Inc. DOI 10.1016/j.molcel.2006.05.025

Structural Basis for Polypyrimidine Tract Recognition by the Essential Pre-mRNA Splicing Factor U2AF65

E. Allen Sickmier,[1,3] Katherine E. Frato,[1] Haihong Shen,[2] Shanthi R. Paranawithana,[1,4] Michael R. Green,[2] and Clara L. Kielkopf[1,*]

[1] Department of Biochemistry and Molecular Biology
Johns Hopkins University Bloomberg School of Public Health
Baltimore, Maryland 21205

[2] Howard Hughes Medical Institute
Programs in Gene Function and Expression and Molecular Medicine
University of Massachusetts Medical School
Worcester, Massachusetts 01605

Summary

The essential pre-mRNA splicing factor, U2AF^{65}, guides the early stages of splice site choice by recognizing a polypyrimidine (Py) tract consensus sequence near the 3′ splice site. Since Py tracts are relatively poorly conserved in higher eukaryotes, U2AF^{65} is faced with the problem of specifying uridine-rich sequences, yet tolerating a variety of nucleotide substitutions found in natural Py tracts. To better understand these apparently contradictory RNA binding characteristics, the X-ray structure of the U2AF^{65} RNA binding domain bound to a Py tract composed of seven uridines has been determined at 2.5 Å resolution. Specific hydrogen bonds between U2AF^{65} and the uracil bases provide an explanation for polyuridine recognition. Flexible side chains and bound water molecules form the majority of the base contacts and potentially could rearrange when the U2AF^{65} structure adapts to different Py tract sequences. The energetic importance of conserved residues for Py tract binding is established by analysis of site-directed mutant U2AF^{65} proteins using surface plasmon resonance.

Introduction

Most transcripts of higher eukaryotes contain intervening sequences (introns) between the protein coding regions (exons) that must be excised by pre-mRNA splicing before nuclear export and translation of the mRNA product. The task of pre-mRNA splicing is accomplished through a series of ATP-dependent conformational rearrangements among constitutive splicing factors and small nuclear (sn)RNAs called the spliceosome (Jurica and Moore, 2003). In addition, alternative splicing factors generate transcript diversity for cell growth and differentiation by incorporating different exons into the final mRNA (Maniatis and Tasic, 2002). An essential splicing factor, U2 auxiliary factor (U2AF), recognizes consensus 3′ splice site sequences in the pre-mRNA and coordinates the initial states of spliceosome assembly. Since formation of the U2AF complex commits the pre-mRNA to be spliced (Michaud and Reed, 1991), U2AF/pre-mRNA interactions present a key target for regulation during alternative splicing, for example by Sex-lethal (SXL) (Valcarcel et al., 1993) or polypyrimidine tract binding protein (PTB) (Sharma et al., 2005). Accurate recognition of the 3′ splice site by U2AF is critical for pre-mRNA splicing, as demonstrated by the association of an estimated half of human genetic diseases with errors in splice site recognition (Garcia-Blanco et al., 2004).

U2AF is a heterodimer of two subunits. The large subunit (U2AF^{65}) recognizes an essential, polypyrimidine (Py) tract pre-mRNA consensus that is composed predominantly of uridines (Zamore et al., 1992). The small subunit (U2AF^{35}) associates tightly with a region near the U2AF^{65} N terminus, and contacts an adjacent "AG" consensus dinucleotide at the nearby intron-exon boundary (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999). Initially, the U2AF heterodimer binds the pre-mRNA as a ternary complex with a third protein, splicing factor 1 (SF1) (Abovich and Rosbash, 1997). SF1 recognizes the branchpoint consensus sequence (BPS) of the pre-mRNA where the first step of the splicing reaction ultimately takes place. In parallel with assembly of the U2AF/SF1/3′ splice site complex, the U1 small nuclear ribonucleoprotein (snRNP) associates with the 5′ splice site. Formation of this early splicing complex brings the BPS, 5′, and 3′ splice sites together in a structured conformation (Kent and MacMillan, 2002; Kent et al., 2003). Next, the U2 snRNP component of the spliceosome forms an ATP-dependent complex with the BPS and U2AF as SF1 dissociates. Stable association of the U2 snRNP requires an N-terminal, arginine-serine-rich (RS) domain of U2AF^{65} (Shen and Green, 2004; Valcarcel et al., 1996), and RNP-unwindases such as the U2AF-Associated-Protein-56KD (UAP56) (Fleckner et al., 1997). Ultimately, U2AF is released from the pre-mRNA before the splicing reaction is catalyzed by the active spliceosome.

U2AF^{65} preferentially binds uridine-rich RNA sequences, as shown by in vitro genetic selection experiments with U2AF^{65} that enrich polyuridine sequences (Singh et al., 1995), and chemical modification of the uridine-N3 or O4 atoms inhibits U2AF^{65} binding by ~100-fold (Singh et al., 2000). Accordingly, Py tracts composed of long uridine stretches promote use of adjacent 3′ splice sites (Coolidge et al., 1997; Reed, 1989). However, natural mammalian Py tracts vary in length and sequence composition (Senapathy et al., 1990). U2AF^{65} universally recognizes these diverse natural Py tracts, which are frequently interrupted with cytosines or purines, albeit with a broad (200-fold) range of affinities (Zamore et al., 1992). In contrast, the alternative splicing factors SXL and PTB bind specific Py tract sequences of regulated splice sites (guanosine-containing uridine tracts or alternating [CU] tracts, respectively) (Perez et al., 1997; Singh et al., 1995; Sosnowski et al., 1989; Valcarcel et al., 1993).

Despite their distinct Py tract specificities, the RNA binding domains of U2AF^{65} (Ito et al., 1999), SXL (Handa

*Correspondence: ckielkop@jhsph.edu
[3] Present address: Department of Molecular Structure, Amgen Inc., One Amgen Center Drive, Thousand Oaks, California 91320.
[4] Present address: Department of Chemistry, Hampton University, Hampton, Virginia 23668.

et al., 1999), and PTB (Conte et al., 2000; Oberstrass et al., 2005; Simpson et al., 2004) are composed of a similar structural scaffold of consecutive RNA recognition motifs (RRM). The RRM, one of the most common types of eukaryotic RNA binding domains, is characterized by two ribonucleoprotein consensus motifs (RNP1 and RNP2) with aromatic and basic residues that interact with single-stranded RNAs (Maris et al., 2005). Specifically, the two central RRMs (RRM1 and RRM2) of U2AF^{65} comprise the minimal Py tract binding domain (U2AF^{65}1,2) (Banerjee et al., 2003, 2004; Zamore et al., 1992). To investigate how these two U2AF^{65} RRMs accomplish versatile Py tract recognition, we present the X-ray structure and a complementary mutational analysis of the U2AF^{65} RNA binding domain in complex with polyuridine RNA. The structure reveals that U2AF^{65} recognizes uridines through a network of hydrogen bond interactions with the base edges rather than shape selection of the smaller pyrimidine compared with purine bases. A significant number of side chain and water-mediated hydrogen bonds may explain the ability of U2AF^{65} to bind a variety of natural Py tract sequences.

Results

X-Ray Structure Determination of a U2AF^{65} Variant with Polyuridine RNA

Initial attempts to cocrystallize U2AF^{65} or U2AF^{65}1,2 with polyuridine RNAs of 12–20 nucleotides were unsuccessful. One strategy to engineer protein crystallization without interfering with functional activity is to shorten poorly conserved loop regions (Mazza et al., 2002; Nolen et al., 2001). The length and sequence composition of the 30 residue linker between RRM1 and RRM2 of human U2AF^{65} are phylogenetically variable (Figure 1), and this linker region is predicted to lack a well-defined structure (http://bioinf.cs.ucl.ac.uk/disopred/ and Shamoo et al. [1995]). Accordingly, several U2AF^{65} variants with shortened linkers were screened for cocrystallization with single-stranded polyuridine oligonucleotides of various lengths (Sickmier et al., 2006). The best crystals were obtained from a human U2AF^{65} fragment containing RRM1 and RRM2 (dU2AF^{65}1,2; residues 148–336 lacking residues 238–257), in complex with a polyuridine RNA dodecamer (rU$_{12}$).

The structure of the dU2AF^{65}1,2 complex with polyuridine RNA was determined at 2.5 Å resolution by molecular replacement using the X-ray structure of the isolated RRM1 domain (PDB code 2FZR) as the search model. Electron density calculated from the initial molecular replacement solution unambiguously revealed three nucleotides bound to RRM1. Following iterations of model building, refinement, and density modification, a complete structure was built for RRM1 and RRM2 of one polypeptide and 7 of the 12 uridines in the asymmetric unit (rU$_7$). No electron density was observed for the remaining five uridines of the oligonucleotide, indicating that these were either disordered or degraded during crystallization. Crystallographic data and refinement statistics are given in Table 1.

Characterization of U2AF^{65} Variant

To confirm that residues 238–257 of the U2AF^{65} RRM1-RRM2 linker were dispensable for pre-mRNA splicing, the activity of a full-length U2AF^{65} variant containing the linker deletion (dU2AF^{65}) was tested in vitro (Figure 2). A titration of the wild-type U2AF^{65} or dU2AF^{65} linker variant with U2AF^{65}-depleted nuclear extract showed that splicing of the adenovirus major late pre-mRNA (*Ad ML*) was restored following addition of comparable levels of both proteins (between 0.25 and 0.50 μM), regardless of the linker composition. Thus, the deleted linker residues were dispensable for U2AF^{65} to function during pre-mRNA splicing in vitro.

The RNA binding properties of the wild-type U2AF^{65}1,2 and the variant dU2AF^{65}1,2 fragments were measured using nitrocellulose filter binding assays with single-stranded RNAs. Both U2AF^{65}1,2 and dU2AF^{65}1,2 proteins bound an RNA oligonucleotide composed of 20 uridines (rU$_{20}$) with comparable apparent equilibrium dissociation constants (K_D 0.37 ± 0.06 μM or 0.41 ± 0.04 μM respectively, data not shown). These binding constants were on the order of those previously measured for a U2AF^{65} fragment binding to natural Py tracts using electrophoretic mobility shift assays (EMSAs) (Zamore et al., 1992). Neither U2AF^{65}1,2 nor dU2AF^{65}1,2 detectably bound a polyguanosine RNA (rG$_{20}$), indicating that the sequence preference for polyuridine over polyguanosine is >200-fold assuming a ~100 μM K_D limit for nitrocellulose filter binding (Hall and Kranz, 1999). In summary, our modification of the RRM1-RRM2 linker region had no apparent effect on the in vitro splicing activity or RNA binding characteristics of U2AF^{65}, establishing that the dU2AF^{65}1,2/rU$_7$ complex is a reliable model system in which to study Py tract recognition by U2AF^{65}.

The RNA affinity of U2AF^{65}1,2 is relatively low compared with the subnanomolar affinity of the corresponding SXL RRMs for its regulated *tra* Py tract (K_D ~5 × 10^{-11} M for an SXL fragment containing RRM1 and RRM2) (Kanaar et al., 1995), although in practice the relative affinities of the full-length U2AF^{65} and SXL proteins for this sequence are closer (K_D 10^{-8} M and 10^{-9} M, respectively) (Valcarcel et al., 1993). Association of the U2AF^{65}/pre-mRNA complex with additional splicing factors such as SF1 (Berglund et al., 1998) and U2AF^{35} (Merendino et al., 1999; Wu et al., 1999; Zorio and Blumenthal, 1999) is functionally important to ensure the affinity and specificity of U2AF^{65} for the entire 3′ splice site consensus (BPS-Py-tract-AG). The binding constants of the isolated U2AF^{65} RRMs for rU$_{20}$ (K_D 4.0 ± 2.0 μM and >100 μM for RRM2 and RRM1, respectively, data not shown), were similar to the RNA affinities of other isolated RRMs (Shamoo et al., 1995). Using these values, a binding constant can be predicted for U2AF^{65}1,2/rU$_{20}$, assuming a 30 residue unstructured linker separating two RRMs. The predicted K_D (0.8 μM, calculated using equations derived by Shamoo and coworkers [Shamoo et al., 1995]) is consistent with the observed K_D of the U2AF^{65}1,2 fragment, which provides further evidence that the U2AF^{65}1,2 interdomain linker does not significantly contribute to RNA binding.

Overall Structure

The repeating unit of the crystal is a 1:1 stoichiometric dU2AF^{65}1,2/rU$_7$ complex, although the dU2AF^{65} RRM1 and RRM2 that interact with the same rU$_7$ strand are contributed by distinct polypeptide chains (Figure 3A