Published in final edited form as:

J Mol Biol. 2011 February 4; 405(5): 1128–1138. doi:10.1016/j.jmb.2010.11.054.

RNA Induces Conformational Changes in the SF1/U2AF⁶⁵ Splicing Factor Complex

Ankit Gupta, Jermaine L. Jenkins, and Clara L. Kielkopf

Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA

Abstract

Spliceosomes assemble on pre-mRNA splice sites through a series of dynamic ribonucleoprotein complexes, yet the nature of the conformational changes remains unclear. Splicing Factor 1 (SF1) and U2 Auxiliary Factor (U2AF⁶⁵) cooperatively recognize the 3' splice site during the initial stages of pre-mRNA splicing. Here, we used small-angle X-ray scattering to compare the molecular dimensions and *ab initio* shape restorations of SF1 and U2AF⁶⁵ splicing factors, as well as the SF1/U2AF⁶⁵ complex in the absence and presence of AdML splice site RNAs. The molecular dimensions of the SF1/U2AF⁶⁵/RNA complex substantially contracted by 15 Å in the maximum dimension, relative to the SF1/U2AF⁶⁵ complex in the absence of RNA ligand. In contrast, no detectable changes were observed for the isolated SF1 and U2AF⁶⁵ splicing factors or their individual complexes with RNA, although slight differences in the shapes of their molecular envelopes were apparent. We propose that the conformational changes that are induced by assembly of the SF1/U2AF⁶⁵/RNA complex serve to position the pre-mRNA splice site optimally for subsequent stages of splicing.

Keywords

pre-mRNA splicing; protein-RNA interactions; U2 snRNP Auxiliary Factor; Splicing Factor-1; small-angle X-ray scattering

Introduction

Pre-mRNA splicing regulates expression of almost all human genes, by either including or excluding different protein coding exons and removing intervening noncoding introns.¹ The fidelity of pre-mRNA splicing is highly regulated, to avoid errors in pre-mRNA splicing that are frequently lethal or lead to human genetic diseases (reviewed in 2). For example, the

© 2010 Elsevier Ltd. All rights reserved.

Please address correspondence concerning the manuscript to: Clara L. Kielkopf: Tel: 585-273-4799, clara_kielkopf@urmc.rochester.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Accession numbers

SAXS data and models were deposited in the BIOISIS SAXS database (www.bioisis.net) with ID codes XX, YY, ZZ, XR, YR, and ZR for SF1¹⁻²⁵⁵, U2AF⁶⁵R123, SF1¹⁻²⁵⁵/U2AF⁶⁵R123, SF1¹⁻²⁵⁵/BPS, U2AF⁶⁵R123/Py tract, and SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS, respectively.

Supplementary Data

Supplementary data associated with this article are available online.

splicing factor U2AF⁶⁵ is essential for vertebrate development,³ and specific U2AF⁶⁵ deficiencies have been associated with cystic fibrosis,⁴ myotonic dystrophy,⁵ and cancers.^{6;} 7 As for U2AF⁶⁵, the splicing factor SF1 is required for mammalian cell viability⁸, and deficiencies of SF1 isoforms have also been shown to promote cancer proliferation.^{9;} 10; 11 In the early stages of splicing, a protein-protein complex between U2AF⁶⁵ and SF1 splicing factors recognizes consensus pre-mRNA sequences located near the 3' splice site (Fig. 1a). 12; 13; 14 These respective pre-mRNA sequences include a polypyrimidine (Py) tract directly adjacent the 3' splice site junction, and an upstream branch point sequence (BPS) that ultimately provides the nucleophile in the splicing reaction. Following association of the SF1/U2AF⁶⁵ complex, the core small nuclear ribonucleoprotein (snRNP) particles of the spliceosome are recruited to the pre-mRNA, and undergo a series of ATP-dependent rearrangements to achieve an active conformation for pre-mRNA splicing (reviewed in 15).

The U2AF⁶⁵ and SF1 splicing factors are composed of modular domains that primarily function to mediate protein/protein or protein/RNA interactions (Fig. 1b). A C-terminal, RRM-like domain of U2AF⁶⁵ with specialized features for protein rather than RNA binding (called a U2AF-Homology Motif, UHM16) is responsible for recognizing an N-terminal domain of SF1 (called a U2AF-Ligand Motif, ULM).^{17;} 18 Two central RNA recognition motifs (RRM) of U2AF⁶⁵ identify the Py tract sequence of the pre-mRNA.¹⁴ The N-terminal region of U2AF⁶⁵ contains a ULM that mediates heterodimerization with the U2AF small subunit (U2AF³⁵),¹⁹ which is required for splicing a subset of introns with short, divergent Py tracts.^{20;} 21; 22 An arginine-serine (RS) rich domain at the U2AF⁶⁵ N-terminus is essential for snRNP recruitment, and is thought to act by enhancing duplex formation between the U2 snRNA and the BPS of the pre-mRNA.²³ For SF1, a central K homology (KH) motif and adjoining Quaking homology 2 (QUA2) region recognize the BPS site.^{17;} 24 The SF1 domain between the KH-QUA2 and ULM is highly conserved from yeast to humans, and contains regulatory sites for phosphorylation.²⁵ The piecewise structures of many of these interacting domains have been determined, including SF1-ULM/U2AF⁶⁵-UHM, U2AF⁶⁵-RRM1-RRM2/Py tract, U2AF⁶⁵-ULM/U2AF³⁵-UHM, and SF1-KH-QUA2/BPS complexes (Fig. 1c).^{18;} 24; 26; 27; 28 Nevertheless, the overall structure of the SF1/U2AF⁶⁵ and its complex with the 3' splice site remain unknown.

Several lines of evidence indicate that the 5' and 3' splice sites communicate during the earliest stages of splicing, despite intervening pre-mRNA sequences that are up to thousands of nucleotides in length. For U2AF⁶⁵ and SF1 to efficiently associate with the 3' splice site during the initial stages of splicing, the U1 snRNP of the spliceosome must be present at the 5' splice site.^{29;} 30 RNA substrates labeled with directed hydroxyl-radical probes confirm that the 5' and 3' splice sites are within close proximity (10–20 Å) even in this early splicing complex.³¹ The inter-splice site interaction depends on the presence of a Py tract in the pre-mRNA substrate for recognition by U2AF⁶⁵. Further, use of a U2AF⁶⁵-tethered hydroxyl-radical probe demonstrates that the U2AF⁶⁵ N-terminus interacts with the BPS³², despite the simultaneous engagement of the C-terminal U2AF⁶⁵-UHM by the SF1/U2AF⁶⁵/BPS complex (Fig. 1a, b).¹⁸ A bent pre-mRNA configuration that places the 5' splice site near the 3' splice site is consistent with the functions of the N-terminal U2AF⁶⁵ RS domain to contact the BPS and promote annealing of the BPS/U2 snRNA duplex (Fig. 1a).^{23;} 33; 34

Despite the plethora of biochemical evidence in support of these conformational changes and the availability of structures for the individual splicing factor domains, the field currently lacks structural information for the early splice site assemblies. To investigate the overall conformations and influence of RNA binding on SF1, U2AF⁶⁵, SF1/U2AF⁶⁵ complexes, we used small-angle X-ray scattering (SAXS) to characterize the low resolution shapes of the splicing factors in the presence and absence of prototypical AdML splice site sequences. This method is suitable for complexes of intermediate size and has the advantage

of determining macromolecular shapes in solution.^{35; 36} To the best of our knowledge, these results represent the first overall views of these splicing factor and RNA complexes.

Design and quality of small-angle X-ray scattering experiments

The boundaries of the protein constructs and sequences of the RNA sites used for SAXS are shown in Fig. 1d,e. Both the U2AF⁶⁵ and SF1 constructs (U2AF⁶⁵R123 and SF1¹⁻²⁵⁵) contain the essential domains of the SF1/U2AF⁶⁵ complex and bind the 3' splice site consensus sequences. The U2AF⁶⁵R123 construct is composed of the two RRM for Py tract recognition and a UHM for SF1 interaction. The RS domain of U2AF⁶⁵ was omitted, since it nonspecifically dominates the RNA affinity and causes protein aggregation that interferes with SAXS.^{37; 38} The SF1¹⁻²⁵⁵ construct is composed of a KH-QUA2 domain for BPS recognition and a ULM for interaction with the U2AF⁶⁵ domain. The C-terminal proline-rich domain and zinc knuckle of SF1 were omitted, since these regions lack known function in formation of the splice site complex and like the RS domain are prone to aggregation that interferes with SAXS. Sequences from the prototypical 3' splice site of the adenovirus major late first intron (AdML) were chosen as the RNA binding sites, since this substrate is independent of the U2AF³⁵ small subunit²¹ and the activities of U2AF⁶⁵ and SF1 in AdML splicing have been studied extensively (for example, 14; 39). Serendipitously, the RNA sites contribute less than 10% of the total scattering mass of the splicing factor complexes. Consequently, the SAXS data primarily reflects the protein conformations.

To verify that the samples used for SAXS were sufficiently concentrated to ensure homogeneous protein/protein and protein/RNA complexes in solution, we measured the apparent equilibrium dissociation constants (K_D) of the complexes. Previously, we had determined a high affinity for the association of SF1¹⁻²⁵⁵ with the U2AF⁶⁵-UHM interaction domain (K_D 12 nM, near the detection limits of the calorimetry method)⁴⁰, indicating that the SF1¹⁻²⁵⁵/U2AF⁶⁵R123 complex would be stable under the conditions of the SAXS experiments. To address the affinities of the protein/RNA complexes, we used fluorescence anisotropy to measure the K_D values of the splicing factors for fluorescein-labeled RNAs (Fig. 1f-h). Consistent with previous reports,^{24; 26} the K_D s for SF1¹⁻²⁵⁵ binding the 9-nucleotide BPS and U2AF⁶⁵R123 binding the 14-nucleotide Py tract of the AdML splice site were in the micromolar range ($0.7 \pm 0.1 \mu\text{M}$ and $0.3 \pm 0.1 \mu\text{M}$, respectively). The SF1¹⁻²⁵⁵/U2AF⁶⁵R123 complex displayed a nanomolar affinity for the 25-nucleotide AdML 3' splice site (3'SS; $2.6 \pm 0.8 \text{ nM}$), which was ~100–200-fold higher than the affinities of the individual proteins for the isolated BPS and Py tract RNAs. Although the large positive cooperativity observed here is enhanced by differences in the lengths of the RNA sites, previous reports likewise document that the presence of SF1 or U2AF⁶⁵ pre-bound to the AdML splice site increases the apparent RNA affinity of its protein partner.^{13; 41} Altogether, these affinity measurements ensure that the sample concentrations used for SAXS data collection were significantly greater than the K_D values of the protein-protein and protein/RNA complexes, by at least 3800-fold for the lowest concentration of SF1¹⁻²⁵⁵/U2AF⁶⁵R123, 150-fold for the lowest concentrations of U2AF⁶⁵R123/Py tract or SF1¹⁻²⁵⁵/BPS, and 14600-fold for the lowest concentration of SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS.

SF1¹⁻²⁵⁵ and U2AF⁶⁵R123 were expressed separately as GST fusion proteins in *E. coli* as described^{40; 42}, and the GST-tags were removed following purification by glutathione-affinity chromatography. All of the proteins and protein-protein complexes used for SAXS experiments were further purified by ion-exchange chromatography and size exclusion chromatography. For preparation of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complex, the purified proteins and synthetic RNA (Dharmacon Technologies) were mixed in a 1:1:1.2 ratio, and the complex was isolated following size exclusion chromatography. The relatively large size of the AdML 3'SS RNA (7.7 kDa) allowed this high affinity complex to be concentrated

using a 2 kDa MWCO filter without detectable loss of the RNA component (Supp. Table S1, Supp. Fig. S1). The shorter BPS RNA (2.7 kDa) approached the MWCO of the filter and bound with a lower affinity, raising selective RNA loss during the early stages of SF1¹⁻²⁵⁵/BPS sample concentration as a concern. To ensure a 1:1 stoichiometry, the SF1¹⁻²⁵⁵/BPS sample was reconstituted following size exclusion chromatography of the SF1¹⁻²⁵⁵ protein. The U2AF⁶⁵R123 complex with Py tract (4.2 kDa) used for detailed analysis here was prepared in a similar manner to the SF1¹⁻²⁵⁵/BPS complex. Control SAXS experiments comparing U2AF⁶⁵R123/Py tract complexes with RNA added either before or after size exclusion chromatography lacked detectable differences in the molecular dimensions (Supp. Fig. S2). All samples (SF1¹⁻²⁵⁵, U2AF⁶⁵R123, SF1¹⁻²⁵⁵/U2AF⁶⁵R123, SF1¹⁻²⁵⁵/BPS, U2AF⁶⁵R123/Py tract, SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS) were monodisperse by dynamic light scattering ($\geq 98\%$ of the scattering contributed by appropriately sized macromolecules, data not shown).

Short and long exposures for three different concentrations of each sample were collected at SIBYLS beamline 12.3.1 of the Advanced Light Source, Lawrence Berkeley National Laboratory. Following subtraction of matching buffer profiles, the final dataset was obtained as previously described²⁶ by merging the datasets collected at different exposures and concentrations. Superposition of scaled scattering curves, linearity of Guinier plots⁴³, and consistent radii of gyration (R_G) over the range of concentrations confirmed the absence of sample aggregation (Fig. 2). In general, the Kratky plots for each of the splicing factors and RNA complexes are broad parabolas followed by slightly increasing $I(q)*q^2$ with increasing q (Fig. 3), consistent with well-folded domains connected by flexible linkers⁴⁴.

Comparison of SF1¹⁻²⁵⁵ in the absence and presence of BPS RNA

Overall, the Kratky plots and pairwise distance distribution ($P(r)$) functions of SF1¹⁻²⁵⁵ compared with SF1¹⁻²⁵⁵/BPS indicate that RNA induces relatively minor changes in the SF1 shape (Fig. 3a, d). The SF1¹⁻²⁵⁵/BPS Kratky plot is similar to that of SF1¹⁻²⁵⁵ with the exception of a slightly more prominent parabolic region characteristic of globular proteins⁴⁴. This may result from RNA-induced structuring of KH-QUA2 loop regions and/or nonspecific RNA interactions by the basic ULM, which is unstructured^{18; 40} and exhibits weak RNA affinity in the absence of U2AF⁶⁵ (data not shown). The $P(r)$ functions of both apo-SF1¹⁻²⁵⁵ and the SF1¹⁻²⁵⁵/BPS complex share similar values of R_G (respectively ~ 30 and 29 Å) and the maximum dimension (D_{max}) (~ 105 Å) (Table 1). Both $P(r)$ curves exhibit a maximum at ~ 25 Å that is likely to correspond to intraparticle distances of the KH-QUA2 domain, and extended tails characteristic of elongated shapes⁴⁵ (Fig. 3a). A minor shoulder at ~ 45 Å in the $P(r)$ function for the SF1¹⁻²⁵⁵/BPS complex may arise from interdomain distances, again suggesting a somewhat more defined interdomain structure for the complex of SF1 with RNA than for the apo-protein.

To obtain three-dimensional information from the scattering profiles, *ab initio* shapes were restored using the program DAMMIN.⁴⁶ This program was chosen since its dummy atom model makes no assumptions concerning the nature of the macromolecule, whereas the alternative program GASBOR⁴⁷ constrains its dummy residues to the Ca positions of a polypeptide chain. Although the structure of the KH-QUA2 domain is available,²⁴ docking of the high resolution structure with the SF1¹⁻²⁵⁵ molecular envelope is expected to be unreliable since this domain comprises only $\sim 50\%$ of the scattering mass of the construct employed here. Ten iterations of DAMMIN⁴⁶ demonstrate acceptable uniformity (normalized spatial discrepancies, NSD of 0.80 for apo-SF1¹⁻²⁵⁵ and 0.68 for SF1¹⁻²⁵⁵/BPS, Table 1) and excellent matches with the experimental scattering data ($\chi \sim 1$) (Table 1, Fig. 2a,b). The averaged, filtered molecular envelopes of apo-SF1¹⁻²⁵⁵ and SF1¹⁻²⁵⁵/BPS are shown in Fig. 4b. The similar *ab initio* shape reconstructions of apo-SF1¹⁻²⁵⁵ and

SF1¹⁻²⁵⁵/BPS complex were consistent with their shared $P(r)$ characteristics (Fig. 4a). Both samples adopt three-lobed ellipsoids, where the distinct lobes may correspond to the C-terminal KH-QUA2 domain, the central phosphorylated domain of unknown structure, and the N-terminal ULM of SF1. The averaged shape of the SF1¹⁻²⁵⁵/BPS sample appears smaller and thinner in one lobe than the corresponding lobe of the apo-SF1¹⁻²⁵⁵ complex. Since the final models represent a frequency map of the dummy atom positions,⁴⁸ constriction of this lobe on RNA binding may reflect a slightly reduced flexibility of the globular SF1 KH-QUA2 domain – an interpretation supported by the enhanced parabolic region of the SF1¹⁻²⁵⁵/BPS Kratky plot.

Comparison of U2AF⁶⁵R123 in the absence and presence of Py tract RNA

The $P(r)$ functions of the apo-U2AF⁶⁵R123 protein and its complex with the AdML Py tract RNA are superimposed in Fig. 3b. The unliganded protein and its RNA complex share R_G values of ~33 Å and extend ~115 Å in the D_{max} (Table 1). In both cases, the skewed, asymmetric $P(r)$ profiles are characteristic of elongated particles with a close to linear distribution of domains⁴⁵, and both display similar first maxima at ~22 Å corresponding to the interatomic distances within an RRM or UHM. A second maximum at ~45 Å in the $P(r)$ function of apo-U2AF⁶⁵R123 corresponds to a shoulder previously observed for a U2AF⁶⁵ construct composed solely of the two RRM domains (U2AF⁶⁵R12) (also shown in Fig. 3b).²⁶ These data suggest a beads-on-a-string shape for U2AF⁶⁵R12 or U2AF⁶⁵R123, in which the position of this second maximum corresponds to the distance between RRM2 and the overall centers of the other domains. The addition of RNA to form the U2AF⁶⁵R123/Py tract complex slightly compresses the parabolic region of the Kratky plot relative to U2AF⁶⁵R123, but has little effect on the increase in $I(q)*q^2$ at higher q (Fig. 3e). This observation suggests that the poorly-conserved linker regions connecting the well-folded RRM and UHM domains remain flexible following RNA binding. Accordingly, deletion or substitution of the inter-RRM linker sequences of U2AF⁶⁵ has no detectable effect on *in vitro* RNA splicing or RNA binding, little effect on the structural characteristics of the U2AF⁶⁵ RNA binding domain determined by SAXS, and the crystal structure of a U2AF⁶⁵ variant exhibits few inter-RRM contacts.^{27; 42}

Ten independent iterations of DAMMIN⁴⁶ gave consistent *ab initio* models from the scattering profiles (NSD of 0.59 for apo-U2AF⁶⁵R123 and 0.55 for U2AF⁶⁵R123/Py tract, Table 1) and acceptable fits with the scattering data ($\chi \sim 1$) (Table 1; Fig. 2a,b). The average, filtered molecular envelopes of the apo-U2AF⁶⁵R123 and its Py tract complex are shown in Fig. 4b. The availability of high resolution structures for the U2AF⁶⁵ RRM and UHM domains, which contribute ~80% of the scattering mass of the construct, enabled rigid body models of these structures connected by *ab initio* linker regions to be fit against the apo-U2AF⁶⁵R123 SAXS data using the program BUNCH⁵⁰. Ten independent iterations of BUNCH starting from random configurations of the domains resulted in reasonable spatial agreement among the final models (NSD 1.22). In all cases, the two RRM domains were located slightly closer to one another than to the UHM, consistent with a 30-residue linker between RRM1 and RRM2 compared with a 40-residue linker between RRM2 and the UHM. The ‘most typical’ BUNCH model (NSD 1.18) is shown superimposed on the DAMMIN dummy atom model in the left panel of Fig. 4b, and has reasonable agreement with the SAXS data (χ 1.26). Consistent with the two maxima observed in the $P(r)$ function, superposition with the BUNCH models of apo-U2AF⁶⁵R123 indicate that the three distinct lobes of the DAMMIN shape are likely to correspond to the RRM1, RRM2, and UHM domains.

In the presence of the Py tract RNA, the U2AF⁶⁵R123/Py tract complex changes to an oblong ellipsoid (Fig. 4b). Based on biochemical expectations^{14; 17; 24; 49} and superposition

with the apo- U2AF⁶⁵R123 models, it appears likely that the enlarged lobe at the base of the orientation shown in Fig. 4b corresponds to U2AF⁶⁵ RRM1 and RRM2 bound to the Py tract RNA, whereas the head may represent the RNA-free UHM domain.

Comparison of SF1¹⁻²⁵⁵/U2AF⁶⁵R123 in the absence and presence of 3' splice site RNA

The $P(r)$ function of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123 complex substantially contracts in the presence of the RNA site, decreasing D_{max} values from ~150 to 135 Å for the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS RNA complex (Fig. 3c, Table 1). Likewise, addition of RNA slightly compresses and alters the lower resolution region of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS Kratky plot from a single to bimodal parabola (Fig. 3f). The higher resolution region of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS Kratky plot remains similar to that of SF1¹⁻²⁵⁵/U2AF⁶⁵R123. Since the parabolic shapes of the Kratky plot correspond to globular folded domains whereas the increasing values at higher resolution reflect flexibility⁴⁴, the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complex appears to become more compact and globular without undergoing substantial changes in flexibility compared with the protein complex in the absence of RNA.

The *ab initio* shapes of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123 and SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complexes show reasonable consistency among ten iterations (NSD 0.76 and 0.69, respectively) and fit the experimental data ($\chi \sim 1$) (Table 1, Fig. 2a,b). The molecular envelope of SF1¹⁻²⁵⁵/U2AF⁶⁵R123 displays two protrusions at the top and right of the viewpoint shown in Fig. 4c, and a larger, extended lobe at the base. In the RNA complex, the lower lobe contracts towards the center of the scattering mass, whereas the top and right protrusions remain distinct. Although the locations of the component proteins within the molecular envelope cannot be verified without access to a neutron source for scattering experiments with deuterium-labeled samples, one possible explanation is that the upper protrusions correspond to the RRM1 and RRM2 domains of U2AF⁶⁵ and the larger base to the SF1/U2AF⁶⁵UHM complex. This interpretation would account for the increased mass in the central region of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complex. Since each RNA binding domain (KH-QUA2 of SF1; RRM1 and RRM2 of U2AF⁶⁵) is respectively tethered by binding the BPS and the adjacent Py tract of the AdML splice site RNA, the KH-QUA2 domain of SF1 may be constrained closer to the U2AF⁶⁵ RRMs in the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS compared with the SF1¹⁻²⁵⁵/U2AF⁶⁵R123 complexes (Fig. 4d).

Summary and implications for pre-mRNA Splicing

The SAXS solution analysis presented here represents the major functional domains of SF1 and U2AF⁶⁵ in the absence and presence of splice site RNAs. Since the bound RNAs contribute only ~10% of the total scattering mass, these data primarily reflect and provide evidence for changes in the protein conformations following RNA binding. The slight changes observed in the overall shapes of the U2AF⁶⁵R123/Py tract and SF1¹⁻²⁵⁵/BPS complexes, compared with the unliganded proteins, can be attributed to localized changes in the RNA binding domains of these splicing factors. For example, BPS binding is expected to be limited to the KH-QUA2 domain of SF1, and may account for the similar but more defined *ab initio* shape of the SF1¹⁻²⁵⁵/BPS complex compared with apo-SF1¹⁻²⁵⁵ (Fig. 4a). Conversely, the RRM1 and RRM2 domains of U2AF⁶⁵ are expected to bind in a linear arrangement to the continuous Py tract sequences of the pre-mRNA, consistent with an ellipsoidal rather than lobed shape of the U2AF⁶⁵R123/Py tract complex (Fig. 4b). Notably, the molecular dimensions of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complex substantially contract when bound to an RNA oligonucleotide containing the 3' splice site consensus sequences (Fig. 3c, Fig. 4c). These comparisons demonstrate that formation of the overall SF1¹⁻²⁵⁵/

U2AF⁶⁵R123/3'SS complex, rather than RNA binding by the individual splicing factors, leads to a major structural change. It is possible that this transition serves as a structural checkpoint to protect the pre-mRNA from entering the splicing pathway until an appropriate SF1/U2AF⁶⁵ complex has assembled at the 3' splice site.

Substantial evidence has accumulated in support of a bent conformation for the 3' splice site when bound in the early splicing complex with U2AF⁶⁵ and SF1. Functionally, the 5' and 3' splice sites communicate either directly or indirectly in this earliest splicing factor complex. The U1 snRNP must be present at the 5' splice site for U2AF to efficiently assemble with the 3' splice site²⁹, as well as for the U2 snRNP to stably associate with the BPS in the subsequent step of splicing.³⁰ Moreover, the *S. cerevisiae* homolog of SF1 (Msl5p) interacts with the Prp40p subunit of the U1 snRNP at the 5' splice site, yet binds the BPS near the 3' splice site.⁵¹ Unambiguous chemical evidence is offered by directed hydroxyl radical probing experiments, which indicate that the 5' and 3' splice sites are within close proximity even in the early splicing factor complex.³¹ This proximity is dependent on the presence of an intact Py tract consensus sequence in the pre-mRNA³¹, and either SF1 or U2AF⁶⁵.³⁶ Accordingly, a bent conformation of the bound Py tract is required to account for the specific interactions mediated by the U2AF⁶⁵ domains, including the U2AF⁶⁵UHM/SF1 complex at the BPS,^{17; 24} crosslinking of the RRM1 and RRM2 domains respectively with the 3' and 5' ends as well as the central nucleotides of the Py tract,⁴⁹ crosslinking of the N-terminal RS domain of U2AF⁶⁵ with the BPS,²³ and directed cleavage of RNA sequences both upstream of the BPS and downstream of the 3' splice site by a hydroxyl radical probe attached to the U2AF⁶⁵ N-terminus.³²

We propose a model based on these known interactions among the U2AF⁶⁵, SF1 and splice site sequences in which the component RNA binding domains are constrained near the center of mass of the complex in the RNA-bound conformation (Fig. 4d). The RNA-induced compression of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS complex presented here is consistent with and supports the model of a bent conformation for the bound splice site. Similarities between the high *q* regions of the Kratky plots for SF1¹⁻²⁵⁵/U2AF⁶⁵R123 and SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS (Fig. 3f) rule out extensive folding of unstructured domains as the chief basis for the RNA-induced changes in the molecular dimensions of the complexes.⁵² Instead, the exact locations of the subunits and domains need to be defined in the longer term by neutron scattering or high resolution structures of the early splice site complexes. This low resolution structural analysis provides a glimpse of conformational changes among essential splicing factors during the early stages of pre-mRNA splicing, and represent an important step towards the future high resolution studies needed to fully elucidate a complete structural model of the mechanism for pre-mRNA splice site choice.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by a grant from the National Institutes of Health (R01 GM070503) to C.L.K. We are grateful to Dr. G.L. Hura and Dr. R. Gillilan for indispensable guidance with SAXS data collection and analysis, and to Dr. J.E. Wedekind and Dr. M.R. Green for insightful discussions. SAXS data were collected at the SIBYLS beamline at the Advanced Light Source, Lawrence Berkeley National Laboratory, which is supported in part by the DOE program Integrated Diffraction Analysis Technologies (IDAT) and the DOE program Molecular Assemblies Genes and Genomics Integrated Efficiently (MAGGIE) under Contract Number DE-AC02-05CH11231 with the U.S. Department of Energy.

References

1. Wang ET, Sandberg R, Luo S, Khrebukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. Alternative isoform regulation in human tissue transcriptomes. *Nature* 2008;456:470–476. [PubMed: 18978772]
2. Garcia-Blanco MA, Baraniak AP, Lasda EL. Alternative splicing in disease and therapy. *Nat Biotechnol* 2004;22:535–546. [PubMed: 15122293]
3. Golling G, Amsterdam A, Sun Z, Antonelli M, Maldonado E, Chen W, Burgess S, Haldi M, Artzt K, Farrington S, Lin SY, Nissen RM, Hopkins N. Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nat Genet* 2002;31:135–140. [PubMed: 12006978]
4. Zuccato E, Buratti E, Stuaní C, Baralle FE, Pagani F. An intronic polypyrimidine-rich element downstream of the donor site modulates cystic fibrosis transmembrane conductance regulator exon 9 alternative splicing. *J Biol Chem* 2004;279:16980–16988. [PubMed: 14966131]
5. Tiscornia G, Mahadevan MS. Myotonic dystrophy: the role of the CUG triplet repeats in splicing of a novel DMPK exon and altered cytoplasmic DMPK mRNA isoform ratios. *Mol Cell* 2000;5:959–967. [PubMed: 10911990]
6. Davies RC, Calvio C, Bratt E, Larsson SH, Lamond AI, Hastie ND. WT1 interacts with the splicing factor U2AF⁶⁵ in an isoform-dependent manner and can be incorporated into spliceosomes. *Genes Dev* 1998;12:3217–3225. [PubMed: 9784496]
7. Maeda T, Furukawa S. Transformation-associated changes in gene expression of alternative splicing regulatory factors in mouse fibroblast cells. *Oncol Rep* 2001;8:563–566. [PubMed: 11295081]
8. Tanackovic G, Kramer A. Human splicing factor SF3a, but not SF1, is essential for pre-mRNA splicing *in vivo*. *Mol Biol Cell* 2005;16:1366–1377. [PubMed: 15647371]
9. Shitashige M, Satow R, Honda K, Ono M, Hirohashi S, Yamada T. Increased susceptibility of SF1(+/-) mice to azoxymethane-induced colon tumorigenesis. *Cancer Sci* 2007;98:1862–1867. [PubMed: 17900258]
10. Shitashige M, Naishiro Y, Idogawa M, Honda K, Ono M, Hirohashi S, Yamada T. Involvement of splicing factor-1 in beta-catenin/T-cell factor-4-mediated gene transactivation and pre-mRNA splicing. *Gastroenterology* 2007;132:1039–1054. [PubMed: 17383426]
11. Cattaruzza M, Schafer K, Hecker M. Cytokine-induced down-regulation of zfm1/splicing factor-1 promotes smooth muscle cell proliferation. *J Biol Chem* 2002;277:6582–6589. [PubMed: 11748220]
12. Berglund JA, Chua K, Abovich N, Reed R, Rosbash M. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell* 1997;89:781–787. [PubMed: 9182766]
13. Berglund JA, Abovich N, Rosbash M. A cooperative interaction between U2AF⁶⁵ and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev* 1998;12:858–867. [PubMed: 9512519]
14. Zamore PD, Patton JG, Green MR. Cloning and domain structure of the mammalian splicing factor U2AF. *Nature* 1992;355:609–614. [PubMed: 1538748]
15. Wahl MC, Will CL, Luhrmann R. The spliceosome: design principles of a dynamic RNP machine. *Cell* 2009;136:701–718. [PubMed: 19239890]
16. Kielkopf CL, Lucke S, Green MR. U2AF homology motifs: protein recognition in the RRM world. *Genes Dev* 2004;18:1513–1526. [PubMed: 15231733]
17. Rain JC, Rafi Z, Rhani Z, Legrain P, Kramer A. Conservation of functional domains involved in RNA binding and protein-protein interactions in human and *Saccharomyces cerevisiae* pre-mRNA splicing factor SF1. *RNA* 1998;4:551–565. [PubMed: 9582097]
18. Selenko P, Gregorovic G, Sprangers R, Stier G, Rhani Z, Kramer A, Sattler M. Structural basis for the molecular recognition between human splicing factors U2AF⁶⁵ and SF1/mBBP. *Mol Cell* 2003;11:965–976. [PubMed: 12718882]
19. Rudner DZ, Kanaar R, Breger KS, Rio DC. Interaction between subunits of heterodimeric splicing factor U2AF is essential *in vivo*. *Mol Cell Biol* 1998;18:1765–1773. [PubMed: 9528748]
20. Zorio DA, Blumenthal T. Both subunits of U2AF recognize the 3' splice site in *Caenorhabditis elegans*. *Nature* 1999;402:835–838. [PubMed: 10617207]

21. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3' splice site AG by the splicing factor U2AF³⁵. *Nature* 1999;402:832–835. [PubMed: 10617206]
22. Merendino L, Guth S, Bilbao D, Martinez C, Valcarcel J. Inhibition of msl-2 splicing by Sex-lethal reveals interaction between U2AF³⁵ and the 3' splice site AG. *Nature* 1999;402:838–841. [PubMed: 10617208]
23. Valcarcel J, Gaur RK, Singh R, Green MR. Interaction of U2AF⁶⁵ RS region with pre-mRNA branch point and promotion of base pairing with U2 snRNA. *Science* 1996;273:1706–1709. [PubMed: 8781232]
24. Liu Z, Luyten I, Bottomley MJ, Messias AC, Hounnigou-Molango S, Sprangers R, Zanier K, Kramer A, Sattler M. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science* 2001;294:1098–1102. [PubMed: 11691992]
25. Manceau V, Swenson MC, Le Caer JP, Sobel A, Kielkopf CL, Maucuer A. Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF⁶⁵. *FEBS J* 2006;273:577–587. [PubMed: 16420481]
26. Jenkins JL, Shen H, Green MR, Kielkopf CL. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF⁶⁵. *J Biol Chem*. 2008
27. Sickmier EA, Frato KE, Shen H, Paranawithana SR, Green MR, Kielkopf CL. Structural basis of polypyrimidine tract recognition by the essential pre-mRNA splicing factor, U2AF⁶⁵. *Mol Cell* 2006;23:49–59. [PubMed: 16818232]
28. Kielkopf CL, Rodionova NA, Green MR, Burley SK. A novel peptide recognition mode revealed by the X-ray structure of a core U2AF³⁵/U2AF⁶⁵ heterodimer. *Cell* 2001;106:595–605. [PubMed: 11551507]
29. Michaud S, Reed R. A functional association between the 5' and 3' splice site is established in the earliest prespliceosome complex (E) in mammals. *Genes Dev* 1993;7:1008–1020. [PubMed: 8504926]
30. Barabino SM, Blencowe BJ, Ryder U, Sproat BS, Lamond AI. Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. *Cell* 1990;63:293–302. [PubMed: 2170025]
31. Kent OA, MacMillan AM. Early organization of pre-mRNA during spliceosome assembly. *Nat Struct Biol* 2002;9:576–581. [PubMed: 12091875]
32. Kent OA, Reayi A, Foong L, Chilibeck KA, MacMillan AM. Structuring of the 3' splice site by U2AF⁶⁵. *J Biol Chem* 2003;278:50572–50577. [PubMed: 14506271]
33. Lee CG, Zamore PD, Green MR, Hurwitz J. RNA annealing activity is intrinsically associated with U2AF. *J Biol Chem* 1993;268:13472–13478. [PubMed: 7685763]
34. Shen H, Green MR. A pathway of sequential arginine-serine-rich domain-splicing signal interactions during mammalian spliceosome assembly. *Mol Cell* 2004;16:363–373. [PubMed: 15525510]
35. Putnam CD, Hammel M, Hura GL, Tainer JA. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 2007;40:191–285. [PubMed: 18078545]
36. Vachette P, Koch MH, Svergun DI. Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. *Methods Enzymol* 2003;374:584–615. [PubMed: 14696389]
37. Yue BG, Ajuh P, Akusjarvi G, Lamond AI, Kreivi JP. Functional coexpression of serine protein kinase SRPK1 and its substrate ASF/SF2 in *Escherichia coli*. *Nucleic Acids Res* 2000;28:E14. [PubMed: 10666475]
38. Caceres JF, Krainer AR. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *EMBO J* 1993;12:4715–4726. [PubMed: 8223480]
39. Arning S, Gruter P, Bilbe G, Kramer A. Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 1996;2:794–810. [PubMed: 8752089]
40. Thickman KR, Swenson MC, Kabogo JM, Gryczynski Z, Kielkopf CL. Multiple U2AF⁶⁵ binding sites within SF3b155: thermodynamic and spectroscopic characterization of protein-protein interactions among pre-mRNA splicing factors. *J Mol Biol* 2006;356:664–683. [PubMed: 16376933]

41. Manceau V, Swenson M, Le Caer JP, Sobel A, Kielkopf CL, Maucuer A. Major phosphorylation of SF1 on adjacent Ser-Pro motifs enhances interaction with U2AF⁶⁵. *FEBS J* 2006;273:577–587. [PubMed: 16420481]
42. Jenkins JL, Shen H, Green MR, Kielkopf CL. Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF⁶⁵. *J Biol Chem* 2008;283:33641–33649. [PubMed: 18842594]
43. Guinier, A.; Fournet, G. Small-angle scattering of X-rays. New York: Wiley Interscience; 1955.
44. Mertens HD, Svergun DI. Structural characterization of proteins and complexes using small-angle X-ray solution scattering. *J Struct Biol* 2010;172:128–141. [PubMed: 20558299]
45. Feigin, LA.; Svergun, DI. Structure Analysis by Small-Angle X-Ray and Neutron Scattering. New York: Springer; 1987.
46. Svergun DI. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999;76:2879–2886. [PubMed: 10354416]
47. Svergun DI, Petoukhov MV, Koch MH. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 2001;80:2946–2953. [PubMed: 11371467]
48. Volkov V, Svergun DI. Uniqueness of *ab initio* shape determination in small angle scattering. *J Appl Cryst* 2003;36:860–864.
49. Banerjee H, Rahn A, Davis W, Singh R. Sex lethal and U2 small nuclear ribonucleoprotein auxiliary factor (U2AF⁶⁵) recognize polypyrimidine tracts using multiple modes of binding. *RNA* 2003;9:88–99. [PubMed: 12554879]
50. Petoukhov MV, Svergun DI. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 2005;89:1237–1250. [PubMed: 15923225]
51. Abovich N, Rosbash M. Cross-intron bridging interactions in the yeast commitment complex are conserved in mammals. *Cell* 1997;89:403–412. [PubMed: 9150140]
52. Bernado P, Perez Y, Blobel J, Fernandez-Recio J, Svergun DI, Pons M. Structural characterization of unphosphorylated STAT5a oligomerization equilibrium in solution by small-angle X-ray scattering. *Protein Sci* 2009;18:716–726. [PubMed: 19309697]
53. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, MR.; Appel, RD.; Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. In: Walker, JM., editor. *The Proteomics Protocols Handbook*. Humana Press; 2005. p. 571-607.
54. Svergun DI. Determination of regularization parameters in indirect-transform methods using perceptual criteria. *J Appl Cryst* 1992;25:495–503.
55. Kozin MB, Svergun DI. Automated matching of high and low resolution structural models. *J Appl Cryst* 2001;34:33–41.

Fig. 1.

(a) Model of SF1/U2AF⁶⁵ interactions with the 3' splice site. BPS, branch point sequence; Py tract, polypyrimidine tract. (b) Schematic diagram of SF1 and U2AF⁶⁵ domains. The proteins are oriented from C- to N-terminus to show the interactions discussed in the text relative to the 5' to 3' orientation of the pre-mRNA. Pro-rich, proline-rich; Zn, zinc knuckle; KH-QUA2, K-homology and Quaking homology 2 motifs; (P), phosphorylated domain; UHM, U2AF homology motif; RRM, RNA recognition motif, ULM, U2AF ligand motif; RS, arginine-serine-rich. (c) Available high resolution structures of SF1 (green) and U2AF⁶⁵ (blue) domains present in the constructs used in this study. From left to right, these include SF1-KH-QUA2/BPS (PDB ID 1K1G), SF1-ULM/U2AF⁶⁵-UHM (PDB ID 1O0P), and U2AF⁶⁵-RRM1-RRM2/Py tract (PDB ID 2G4B). (d) SF1¹⁻²⁵⁵ and U2AF⁶⁵RRM123 protein constructs used in this study. (e) Sequences of RNA oligonucleotides derived from the 3' splice site of the adenovirus major late first intron (AdML) used in this study. The BPS is colored green and the Py tract is blue. (f–h) Representative fluorescence anisotropy curves for (f) SF1¹⁻²⁵⁵, (g) U2AF⁶⁵RRM123, or (h) U2AF⁶⁵RRM123/SF1¹⁻²⁵⁵ binding the fluorescein-labeled RNA sites shown in d. The average apparent equilibrium dissociation constants (K_D) and standard deviations of three experiments are given. Experimental conditions and data analysis were as described.²⁶ a.u., arbitrary units.

Fig. 2.

(a–b) X-ray scattering profiles of splicing factors in the (a) absence or (b) presence of RNA sites. Experimental data (colored triangles) from a range of concentrations and exposures were merged as described²⁶, and are shown compared with data calculated from the most typical DAMMIN model (solid lines). (c–d) Guinier plots of $\ln(I(q))$ versus q^2 in the range $qR_G < 1.3$ for splicing factors in the (c) absence or (d) presence of RNA sites.

SF1¹⁻²⁵⁵ or SF1¹⁻²⁵⁵/BPS, green; U2AF⁶⁵R123 or U2AF⁶⁵R123/Py tract, blue; SF1¹⁻²⁵⁵/U2AF⁶⁵R123 or SF1¹⁻²⁵⁵/U2AF⁶⁵R123/3'SS, cyan. The concentrations of the protein/RNA complexes were calculated as described in Supp. Table S1, using the estimated extinction coefficients of the RNA components at 280 nm and the calculated extinction coefficients of

the protein components⁵³. For $a-d$, the relative scattering intensities are displaced along the logarithmic y-axis for clarity.

Fig. 3.

(a–c) Comparison of $P(r)$ functions calculated from the experimental scattering profiles using the program GNOM⁵⁴ highlight a significant change in the dimensions of the SF1¹⁻²⁵⁵/U2AF⁶⁵R123 complex when bound to RNA. The functions in the presence and absence of RNA are scaled relative to one another according to the integrated areas of the curves. The radii of gyration (R_G) and maximum intraparticle sizes (D_{max}) are given in Table 1. (d–f) Comparison of Kratky plots rule out large disorder-order transitions of the splicing factors following formation of their RNA complexes. (a), (d) SF1¹⁻²⁵⁵ (green); (b), (e) U2AF⁶⁵R123 (blue) and the minimal RNA binding domain, U2AF⁶⁵R12 (grey, shown for comparison from 42); (c), (f) SF1¹⁻²⁵⁵/U2AF⁶⁵R123 (cyan). RNA complexes are colored magenta.

Fig. 4.

Ab initio models of (a) SF1^{1–255} (green), (b) U2AF⁶⁵R123 (blue) or (c) SF1^{1–255}/U2AF⁶⁵R123 (cyan) and their RNA complexes (magenta). Molecular envelopes calculated in ten iterations of DAMMIN⁴⁶ with different seed values were superimposed, averaged and filtered using the programs SUPCOMB⁵⁵ and DAMAVER.⁴⁸ A typical rigid body model generated using the program BUNCH⁵⁰ of the U2AF⁶⁵ RRM1, RRM2, and UHM structures connected by *ab initio* interdomain linkers is shown superimposed using SUPCOMB⁵⁵ on the DAMMIN *ab initio* model of U2AF⁶⁵R123. Viewpoints are consistent between the separated and superimposed molecular envelopes of each sample. Surfaces of the protein-RNA complexes are semitransparent in the superimposed representations. (d) A suggested model for the changes observed following binding of the splice site RNA in the SF1^{1–255}/U2AF⁶⁵R123/3'SS complex.

Table 1

Overall parameters and discrepancy values derived from scattering data

Sample	R_G^S (Å)	R_G^P (Å)	D_{max} (Å)	$\langle\chi_{ab}\rangle$	$\langle NSD_{ab}\rangle$
SF11-255	30.2 ± 0.2	31.4 ± 0.1	105 ± 10	1.04 ± 0.01	0.80 ± 0.03
SF11-255/BPS	29.4 ± 0.2	30.3 ± 0.1	105 ± 10	0.99 ± 0.02	0.68 ± 0.02
U2AF ⁶⁵ R123	31.4 ± 0.3	33.8 ± 0.1	115 ± 10	0.99 ± 0.01	0.59 ± 0.03
U2AF ⁶⁵ R123/Py tract	32.1 ± 0.2	33.8 ± 0.1	115 ± 10	0.97 ± 0.01	0.55 ± 0.01
SF11-255/U2AF ⁶⁵ R123	39.4 ± 0.2	41.9 ± 0.1	150 ± 10	1.02 ± 0.02	0.76 ± 0.02
SF11-255/U2AF ⁶⁵ R123/3'SS	38.0 ± 0.2	38.1 ± 0.1	135 ± 10	1.08 ± 0.02	0.69 ± 0.03

R_G^S , radius of gyration value from Guinier analysis of the final merged data; R_G^P , radius of gyration value from $P(r)$ analysis; D_{max} , maximum size; $\langle\chi_{ab}\rangle$ values are the average discrepancies between the experimental data and the computed scattering calculated from *ab initio* models resulting from ten independent DAMMIN trials. $\langle NSD_{ab}\rangle$ is the average normalized spatial discrepancy among ten *ab initio* envelopes.