Dimerization and Protein Binding Specificity of the U2AF Homology Motif of the Splicing Factor Puf60

Lorenzo Corsini†, Michael Hothorn††, Gunter Stier†, Vladimir Rybin§, Klaus Scheffzek¶, Toby J. Gibson‡, and Michael Sattler†††‡‡

From the †Structural and Computational Biology Unit, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, the ‡Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädt Landstrasse 1, 85764 Neuherberg, Germany, and the ¶Munich Center for Integrated Protein Science and Chair Biomolecular NMR, Department Chemie, Technische Universität München, Lichtenbergstrasse 4, 85747 Garching, Germany

PUF60 is an essential splicing factor functionally related and homologous to U2AF65. Its C-terminal domain belongs to the family of U2AF (U2 auxiliary factor) homology motifs (UHM), a subgroup of RNA recognition motifs that bind to tryptophan-containing linear peptide motifs (UHM ligand motifs, ULMs) in several nuclear proteins. Here, we show that the Puf60 UHM is mainly monomeric in physiological buffer, whereas its dimerization is induced upon the addition of SDS. The crystal structure of PUF60-UHM at 2.2 Å resolution, NMR data, and mutational analysis reveal that the dimer interface is mediated by electrostatic interactions involving a flexible loop. Using glutathione S-transferase pulldown experiments, isothermal titration calorimetry, and NMR titrations, we find that Puf60-UHM binds to ULM sequences in the splicing factors SF1, U2AF65, and SF3b155. Compared with U2AF65-UHM, Puf60-UHM has distinct binding preferences to ULMs in the N terminus of SF3b155. Our data suggest that the functional cooperativity between U2AF65 and Puf60 may involve simultaneous interactions of the two proteins with SF3b155.

Pre-mRNA splicing is a stepwise process initiated by the recognition of sequence elements at the splice site by specific splicing factors (1). The branch point sequence is recognized by splicing factor SF1 (2, 3), whereas the polypyrimidine tract and the 3’ splice site AG-dinucleotide are bound by the heterodimer U2AF65-U2AF35 (4–7). Although SF1 alone interacts to SF1 and to the polypyrimidine tract (8). In the next step of splicing initiation U2 snRNP is brought to the 3’ splice site. This involves base pairing of the U2 RNA to the branch site RNA (9) and localization of the SF3b subunit p14 near the branch point adenosine by an interaction with the N terminus of the U2 snRNP component SF3b155 (10–12). Initial contacts between the U2 snRNP and the pre-mRNA are mediated by the N terminus of SF3b155 binding to U2AF65 and displacing SF1 from the branch point sequence (13).

The third RNA recognition motif (RRM) of U2AF65 and the RRM of U2AF35 share distinct sequence features that are not found in canonical RRMs and mediate binding to tryptophan-containing peptide sequences in cognate splicing factors (14–19). These noncanonical RRMs form a subgroup of RRMs called U2AF homology motif (UHM) (20). UHMs in other splicing factors have also been shown to bind to short tryptophan-containing linear motifs in U2AF65, SF1, and SF3b155 (21–24), which are thus called UHM ligand motifs (ULM) (23). High resolution structures of protein-peptide complexes involving U2AF35, UHM/U2AF65, ULM (14), U2AF65-UHM/SF1-ULM (15), and the UHM of SPF45 (splicing factor 45 kDa) bound to an ULM in SF3b155 (23) all share a very similar mode of molecular recognition, suggesting that UHMs in other proteins might bind similar linear motifs as well. Because the ULM consensus motif is rather soft (δ(R)₁₋₄=ε₀₋₁(W/D/E/N/Q)₁₋₂), its mere presence in a protein sequence cannot unambiguously identify a functional ULM. Candidate ULMs are found in many proteins, but a biological function has so far been assigned only to those in U2AF65 (17, 25), SF1 (16, 23), and SF3b155 (13, 23).

PUF60 (poly-U-binding factor 60 kDa, also called FIR, Hfp, Ro-bp1) is a splicing factor homologous to and complementary in function to U2AF65. Similarly to U2AF65, its domain structure consists of a predicted intrinsically unstructured N terminus, two central RRM domains, and a C-terminal UHM. The UHM domain is special in that it has been reported to mediate homodimerization of Puf60 in SDS-PAGE (21). Full-length Puf60 was found to interact with itself in yeast two-hybrid analysis revealing that the dimer interface is mediated by electrostatic interactions involving a flexible loop. Using glutathione S-transferase pulldown experiments, isothermal titration calorimetry, and NMR titrations, we find that Puf60-UHM binds to ULM sequences in the splicing factors SF1, U2AF65, and SF3b155. Compared with U2AF65-UHM, Puf60-UHM has distinct binding preferences to ULMs in the N terminus of SF3b155. Our data suggest that the functional cooperativity between U2AF65 and Puf60 may involve simultaneous interactions of the two proteins with SF3b155.

Pre-mRNA splicing is a stepwise process initiated by the recognition of sequence elements at the splice site by specific splicing factors (1). The branch point sequence is recognized by splicing factor SF1 (2, 3), whereas the polypyrimidine tract and the 3’ splice site AG-dinucleotide are bound by the heterodimer U2AF65-U2AF35 (4–7). Although SF1 alone interacts only weakly with the branch point sequence, this interaction is stabilized significantly by U2AF65, which binds simultaneously to SF1 and to the polypyrimidine tract (8). In the next step of
Puf60 was discovered as a poly-U RNA-binding protein required to reconstitute splicing in depleted nuclear extracts. Its function is enhanced by the presence of U2AF65, but not by the small U2AF subunit, U2AF35 (21). Puf60 and U2AF65 can interact in vitro and in yeast cells (21, 26, 27). It was recently demonstrated that Puf60 and U2AF65 mutually enhance their affinity for binding polyuridylic tract RNA in a cooperative fashion. Moreover, the ratio of U2AF65 to Puf60 can directly influence selective inclusion or skipping of alternatively spliced exons in several genes (28). The function of Puf60 in splicing is thus closely linked to the function of U2AF65.

In addition to its role in alternative splicing, Puf60 also controls human c-myc gene expression. Under the synonym FIR (FBP-interacting repressor), Puf60 was reported to interact with and inhibit the transcription factor FBP (EUSE (far upstream sequence element)-binding-protein), an activator of c-myc promoters (29). Probably because of a similar mechanism, mutations in the Drosophila homolog of Puf60, Hp (Half Pint), lead to increased expression of d-myc genes, thus negatively regulating cell cycle progression (30). Hp mutations also lead to aberrant splicing of specific mRNAs in Drosophila ovari- (31). Similar to its mammalian ortholog Puf60, Hfp is thus a regulator both of transcription and of alternative splicing.

Here, we report that Puf60 UHM is mainly monomeric under physiological conditions, whereas it dimerizes upon the addition of SDS. The crystal structure of PUF60-UHM and mutational analysis reveal that the dimerization is entirely mediated by electrostatic interactions. NMR relaxation data show that the dimer interface involves a loop that is highly flexible in solution. Furthermore, we show that PUF60-UHM binds to ULM sequences in U2AF65, SF1, and SF3b155. The UHMs in PUF60 and U2AF65 show preferences for binding to different ULMs in the N terminus of SF3b155. We propose that PUF60 and U2AF65 may cooperatively recruit U2 snRNP by simultaneously binding to SF3b155.

**EXPERIMENTAL PROCEDURES**

**Protein Preparation**—Recombinant Puf60-UHM (residues 460–559, wild type, and mutants), thioredoxin-Puf60-UHM, SF45-UHM (301–401), U2AF35-UHM (residues 38–152), and U2AF65-UHM (residues 369–475) were expressed from modified pET9d vectors with tobacco etch virus protease cleavable, N-terminal His6, and glutathione S-transferase (GST) tags. Unlabeled proteins were expressed in Escherichia coli S-transferase (GST) tags. Unlabeled proteins were expressed in Escherichia coli BL21(DE3)pLyss in LB medium. All of the proteins were purified with nickel-nitrilotriacetic acid-agarose (Qiagen) under standard conditions and buffer exchanged to phosphate-buffered saline. For the preparation of ULM peptides for NMR titrations see the supplemental data. Isotopically 13C- and/or 15N-labeled proteins were expressed in minimal (M9) medium supplemented with 13C-6-glucose and/or 15NH4Cl. NMR samples were concentrated to 0.3–1.0 mM in 20 mM Na2PO4 buffer (pH 6.8), 150 mM NaCl, and 5 mM β-mercaptoethanol. The samples used for crystallization were additionally purified by size exclusion chromatography on a Superdex™ 75 16/60 prep grade column.

**NMR**—All of the NMR spectra were recorded at 300 K on a Bruker DRX500 spectrometer, processed with NMRpipe (32), and analyzed with NMRView (33). Backbone 1H, 15N, and 13C resonances were assigned with standard triple resonance experiments (34). 15N relaxation data were recorded as described (35). Dissociation constants were derived from chemical shift displacements in HSQC spectra upon the addition of ligands as described (36) (see supplemental data).

**Crystallization and Data Collection**—For crystallization, the chimeric thioredoxin-Puf60(460–559) fusion protein was concentrated to about 70 mg/ml in 20 mM Tris (pH 7.0), 150 mM NaCl, 5 mM β-mercaptoethanol. The crystals were grown by vapor diffusion from hanging drops composed of 1 μl of protein solution and 1 μl of crystallization buffer (1.4 mM (NH4)2SO4, 50 mM potassium formate) suspended over 1 ml of the latter as reservoir solution. The crystals grew to sizes of about 100 × 100 × 500 μm and were cryoprotected by serial transfer into a solution containing 20% (v/v) ethylene glycol, 1.5 mM (NH4)2SO4, 50 mM potassium formate). Diffraction data were recorded at beam-line PX01 of the Swiss Light Source (Villigen, Switzerland). Data processing and scaling was carried out with XDS (37).

**Structure Determination and Refinement**—The structure of the thiorodoxin-Puf60 fusion protein was solved by molecular replacement as implemented in PHASER (38). The structure of *E. coli* thiorodoxin (Protein Data Bank code 2TRX) and a homology model of Puf60-UHM generated with MODELLER (39) based on the structure of free SF45-UHM (Protein Data Bank code 2PE8) as a template were used as search models. The solution comprises eight Trx-Puf60-UHM monomers that were refined in alternating cycles of model correction in COOT (40), and restrained refinement as implemented in REFMAC (41) and PHENIX.REFINE (Ref. 42; see Table 1 for structural statistics). Structures were visualized with PYMOL (DeLano Scientific LLC, San Carlos, CA). The eight UHM domains in the unit cell of the crystal structure can be superimposed onto a reported solution structure of Puf60-UHM (Protein Data Bank code 2DNY) with root mean square deviations of 0.9–1.1 Å over 90 of 100 Ca atoms. The solution structure, however, does not indicate dimerization of the Puf60-UHM.

**GST Pulldown Experiments**—GST-tagged ULMs (1 nmol) were mixed with 3 nmol of Hisα-tagged UHMs in 150 μl of phosphate-buffered saline supplemented with 2 mM β-mercaptoethanol and 0.1% (w/v) Igepal CA-630 at 22 °C and mixed vigorously for 1 h. For GST precipitation, 8 μl of glutathione-Sepharose 4B (Amersham Biosciences) pre-equilibrated in phosphate-buffered saline were added and mixed vigorously for 30 min. The beads were washed three times for 1 min in the buffer described above and analyzed by SDS-PAGE. Western blotting was carried out with α-Puf60 antibody (Abcam 22819).

**RESULTS**

**Puf60-UHM Is Mainly Monomeric in Physiological Buffer**—Puf60 interacts with itself in yeast two-hybrid analyses (26, 27), and its C-terminal UHM domain has been shown to form dimers resistant to denaturing SDS-PAGE (21). We used NMR spectroscopy to characterize the oligomerization state of...
Structure and Function of Puf60 UHM

TABLE 1

Summary of crystallographic analysis

<table>
<thead>
<tr>
<th>Data collection</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2_1, 2, 2 (19)</td>
</tr>
<tr>
<td>Unit cell dimensions</td>
<td>a = 75.12 Å, b = 89.82 Å, c = 299.39 Å</td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>1.006 Å</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>4.975–2.20 (2.33–2.20 Å)</td>
</tr>
<tr>
<td>Unique reflections</td>
<td>102,915 (16,265)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>6.1 (5.8)</td>
</tr>
<tr>
<td>&lt;I&gt;/sigma(I)</td>
<td>14.54 (3.21)</td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>10.4 (60.6)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.8 (98.9)</td>
</tr>
</tbody>
</table>

Refinement

| Resolution range (Å)             | 4.975–2.20 Å                  |
| Rmerge/Rfree                     | 0.211/0.271                   |
| No. of atoms                     | 13,138                        |
| Protein                          | 12,096                        |
| Water                            | 22.44                         |
| Puf60-UHM domains                | 20.13                         |
| Water                            | 35.21                         |
| Root mean square deviations      | 0.006                         |
| Angles (°)                       | 1.012                         |
| Ramachandran disorder (°)        | 1387 (92.3%)                  |
| Additionally allowed (°)         | 10.6 (71.7%)                  |
| Generously allowed (°)           | 2.0 (1.1%)                    |
| Disallowed (°)                   | 7.0 (0.5%)                    |

As defined in XDS (37).

As defined in REFMAC (41).

Residual isotropic B-factor after TLS refinement.

As defined in PROCHECK (59).

Pu60-UHM in solution (50 mM PEP, pH 7.0, 150 mM NaCl, 5 mM dithiothreitol). NMR secondary chemical shifts (Fig. 1A) show that Pu60-UHM adopts the typical β1-αA-β2-β3-αB-β4-αC topology found for all RRM and UHMs (20, 43). The overall rotational correlation time (τ_2) of Pu60-UHM was calculated from the ratio of the trimmed mean 15N longitudinal (T_1) and transverse (T_2) relaxation times of residues with heteronuclear 1H-15N NOE values above 0.65 (Fig. 1B) (44, 45). The average 15N T_1/T_2 ratio for these residues is 7.4 (Fig. 1C), corresponding to a τ_2 of 9.7 ns. However, at 50 MHz Larmor frequency and 297 K, τ_2 values of 8.3 ns (T_1/T_2 = 5.7) and 15.8 ns (T_1/T_2 = 17.7) would be expected for a 13-kDa monomer and a 26-kDa UHM domain dimer, respectively (46, 47) (gray lines in Fig. 1C). Thus, the observed relaxation times indicate the presence of a mainly monomeric rather than a dimeric form of Pu60-UHM. The slightly increased T_1/T_2 ratio, compared with what is expected for a pure monomer, might result from some nonspecific aggregation, because the T_1/T_2 ratio of Pu60-UHM lacking an N-terminal His tag (T_1/T_2 = 4.4, τ_2 = 7.0 ns; data not shown) is consistent with a monomeric protein.

To further investigate the oligomerization state of the UHM, we used sedimentation velocity analytical ultracentrifugation (AUC). The AUC data also indicate a largely monomeric state of the UHM domain (Fig. 1D, solid gray line), whereas partial dimerization is observed at higher protein concentrations (Fig. 1D, dotted gray line). By fitting the AUC data to a monomer-dimer equilibrium model, the dimerization constant is estimated to be K_dimer = 3–4 mM. The two central RRM domains of Pu60 were reported to dimerize in the presence of DNA (48). We therefore tested whether a construct comprising RRM1-RRM2 and the C-terminal UHM had a tendency to dimerize without DNA or SDS. Our AUC data indicate that this construct is largely monomeric as well (Fig. 1D, solid black line). Taken together, these data demonstrate that Pu60-UHM (in the absence of SDS) and Pu60 RRM1-RRM2-UHM (in the absence of DNA or SDS) are monomeric in solution. Therefore, the UHM dimerization observed in denaturing and reducing SDS-PAGE (21) is presumably induced by the experimental conditions.

The Three-dimensional Structure of the Pu60-UHM—Next, we determined the crystal structure of Pu60-UHM at 2.2 Å resolution. Diffracting crystals could only be obtained using a fusion protein, in which E. coli thioredoxin A (Trx) is connected to the N terminus of Pu60-UHM via a short linker sequence (49). We confirmed that the Trx-UHM construct dimerizes in SDS-PAGE similarly to what is seen for the UHM alone (data not shown).

The asymmetric unit consists of eight Trx-Pu60 fusion proteins arranged in a doughnut shape. Eight Trx molecules are stacked in two layers in the center of the doughnut, surrounded by a ring of eight Pu60-UHM domains (Fig. 2A). Consistent with the NMR secondary chemical shifts, Pu60-UHM adopts a β1-αA-β2-β3-αB-β4-αC secondary structure. A central four-stranded β-sheet is sandwiched by helices αA and αB on one side and helix αC on the other side (Fig. 2B). As seen in other UHM structures (14, 23), Pu60-UHM has an additional strand β3’ adjacent to β4, which forms a β-hairpin extension to the central four-stranded β-sheet. The β3’ strand comprises the conserved Arg-Xaa-Phe motif (RWF 535–537 in Pu60), which plays a crucial role in ULM binding in all known UHM-ULM complexes (14, 15, 23). A unique structural feature of Pu60-UHM is the presence of unusually long β2 and β3 strands, which form a β-hairpin that protrudes out of the β-sheet (Fig. 2B). In solution, the acidic β2-β3 loop is flexible, as indicated by low heteronuclear NOE values, which drop to a minimum of 0.12 for Gly<sup>401</sup> (Fig. 1B). In contrast, an average heteronuclear NOE of 0.74 for residues 462–501 and 512–559 indicates the absence of internal motion on subnanosecond time scales.

Dimerization Interface—Because Pu60-UHM crystallizes at concentrations above the dimerization constant of 3–4 mM (1.5 mM in the mother liquor, 32 mM in the crystal lattice), we expected to detect a dimeric UHM in the crystal. Analysis with PISA (50) shows that each of the eight Pu60-UHM domains contacts three other Pu60-UHM domains, two in the same and one in a symmetry-related asymmetric unit. One of the UHM- UHM interfaces within an asymmetric unit is composed of charged interactions between the residues EEE (505–507) in the β2-β3 loop of one protein monomer and Arg<sup>467</sup>, Arg<sup>540</sup>/Lys<sup>541</sup> in the adjacent strands β1 and β4 of the other dimer subunit, respectively (Fig. 2C). As shown in the electrostatic surface representation (Fig. 2C, right panel), the residues Arg<sup>467</sup>, Arg<sup>540</sup>, and Lys<sup>541</sup> form a positively charged surface, which is contacted by the negatively charged acidic β2-β3 loop. The electrostatic interactions involve the tips of the long side chains of arginine/lysine, which contact glutamate/aspartate residues in the mobile β2-β3 loop. Of the six salt bridges that can be formed, electron density is visible for a maximum of four contacts in any of the putative dimer interfaces in the asymmet-
Structure and Function of Puf60 UHM

A. \(\Delta \delta(C_a) - \Delta \delta(C_p)\) secondary chemical shift

B. \(\langle H \rangle ^{15}N\) NOE

C. \(T_1/T_2\) of residues with \(\langle H \rangle ^{15}N\) NOE > 0.65

D. Analytical Ultracentrifugation

FIGURE 1. NMR analysis of Puf60-UHM. All of the experiments were recorded at 297 K, on a Bruker DRX500 spectrometer using standard experiments (33). A, secondary chemical shifts \(\Delta \delta(C_a-C_p)\) reveal secondary structure as indicated below the graph with the primary sequence of Puf60-UHM. B, \(\langle H \rangle ^{15}N\) heteronuclear NOE of Puf60-UHM. C, ratio of \(\langle H \rangle ^{15}N\) NOE and \(T_1/T_2\) relaxation times for residues with \(\langle H \rangle ^{15}N\) NOE > 0.65. The gray horizontal lines depict expected average values for a 12.6-kDa monomer and a 25-kDa dimer at 297 K and 50.68 MHz Larmor frequency, calculated as in Refs. 44 and 45. D, analytical ultracentrifugation analysis of Puf60-UHM (solid gray line, concentration 29 \(\mu\)M; dotted gray, 377 \(\mu\)M) and a construct comprising the two central RRM domains and the UHM of Puf60 (black line, 11 \(\mu\)M). The expected theoretical molecular masses are given on the right.

To determine which residues are involved in the dimerization of Puf60 in SDS-PAGE, we introduced amino acid changes for the residues in the \(\beta 2-\beta 3\) loop (E501A/K502A/Q503A, E505A/E506A/E507A, and D508A/A509A/E510A) and of the positively charged residues that contact the acidic loop (R540A/K541A, R467A, and RK540–541AA+R467A). The structural integrity of the E505A/E506A/E507A mutant was confirmed by comparison of the HSQC spectra (supplemental Fig. S1). The integrity of the other mutants was confirmed by one-dimensional NMR (data not shown).

In denaturing SDS-PAGE, Puf60-UHM wild type (12.6 kDa), E501A/K502A/Q503A, R540A/K541A, and R467A run at an apparent molecular mass of 28 kDa (Fig. 2D, lanes 1, 2, 5, and 6, respectively), as expected for a dimer. In contrast, the mutants E505A/E506A/E507A (lane 3), D508A/A509A/E510A (lane 4), and R540A/K541A+R467A (lane 7) run at lower molecular masses, indicating that their dimerization is impaired.

These findings indicate that the dimerization of Puf60 UHM involves the acidic residues \(505\text{EEDEA}^{510}\) in the flexible \(\beta 2-\beta 3\) loop and the basic residues Arg\(^{467}\) and Arg\(^{540}\)-Lys\(^{541}\). Salt bridges and electrostatic contacts between these regions thus mediate dimerization of Puf60-UHM in the presence of SDS and presumably also contribute to the small population of dimeric species in physiological buffers (Fig. 1D).

To confirm that the observed bands indeed correspond to dimerization of the UHM in SDS-PAGE and that the observed positions of the bands do not fortuitously appear at unusual positions, we mixed recombinant, purified ZZ-tagged wild type UHM (28.4/56.8 kDa for monomer/dimer; Fig. 2E, lane 1) with untagged wild type UHM (12.6/25.2 kDa; Fig. 2E, lane 6). Because the protein species at 41.2 kDa in lanes 2 and 3 is not contained in either pure ZZ-tagged UHM (lane 1) or untagged UHM (lane 6), the appearance of a mixed dimer species of the type ZZ-UHM-UHM at 41.2 kDa (lanes 2 and 3) proves the formation of a mixed dimer. The UHM mutant E505A/E506A/E507A does not form the mixed dimer species, confirming that the mutations impair the dimerization of the UHM in SDS-PAGE (Fig. 2E, lanes 4, 5, and 7).
The 15N simple equimolar binding of SDS does not induce dimerization of Puf60-UHM. However, intermediate SDS concentrations (>0.02% (w/v) SDS, ~3-fold molar excess) induce dimerization of the wild type protein, as indicated by the appearance of new signals in the NMR spectra that have 15N NMR relaxation properties expected for a dimer (supplemental Fig. S3). With further increasing SDS concentrations, both Puf60-UHM wild type as well as the mutant protein are denatured (Fig. 3B).

Binding of Puf60 UHM to Tandem ULMs—The dimerization propensity of Puf60-UHM opens the possibility that simultaneous binding of two UHM domains to two adjacent ULMs on the same peptide chain (tandem ULM motif) could cooperatively induce the dimerization also in the absence of SDS. Based on the distance of the ULM-binding sites of the Puf60 homodimer in the crystal structure, we estimated that the ULMs should be separated by a minimum of 15–20 residues in an extended conformation. We identified evolutionarily conserved tandem ULMs in intrinsically disordered regions of several proteins with the program SIRW (51). Of these, tandem ULMs in SF3b155 (194–229, 210–251 and 229–269) and in the nuclear RNA helicase Prp16 (201–238) (Fig. 4, A and B) were tested experimentally for binding to Puf60-UHM.

Using Western blot detected GST pulldown experiments, we found that the tandem ULM sequence of SF3b155 (194–229) binds Puf60-UHM (supplemental Fig. S4A). However, ITC (supplemental Fig. S4, C and D) and NMR (supplemental Fig. S5A) data show that the binding of Puf60-UHM to SF3b155 (194–229) is mediated by the ULM around Trp300 and that the ULM at Trp216 does not contribute to the binding cooperatively. NMR titrations reveal a weak interaction of Puf60-UHM with Prp16 (201–238), which was not detected in the GST pulldown experiments. However, the two ULMs (Trp215 and Trp230) in Prp16 do not mutually enhance each other’s binding cooperatively (supplemental Fig. S5B). Thus, dimerization of Puf60-UHM is not induced upon binding to these tandem ULMs in the absence of SDS. ULM binding in the presence of SDS (350 μM to 1.4 mM) was not observed in GST pulldown experiments (data not shown). This is consistent with the observation that the SDS interaction
but reproducible (supplemental Fig. S4B) and has been described previously (18).

**ULM-binding Site Mapping on Puf60-ULM**—We used NMR titrations to map the ULM-binding site onto the structure of Puf60-ULM. Unlabeled peptides comprising the ULMs in SF1(1–25), SF3b155(194–229), SF3b155(317–357), and U2AF65(85–112) were titrated to 15N-labeled Puf60-ULM, and chemical shift perturbations were monitored in 1H-15N correlation spectra (Fig. 5A). Mapping the chemical shift perturbations onto the crystal structure of Puf60-ULM shows that for all four ULMs, the strongest chemical shift perturbations cluster around the β1-αA loop, helix αA and around the RWF motif (part of helix αB and strands β3’ and β4) of Puf60-ULM (Fig. 5B).

The ULM interaction interface of Puf60-ULM is thus analogous to the interfaces of the U2AF35UHM-U2AF65ULM (20), the U2AF65ULM-SF1ULM (15), and the SPF45SF3b155(330–342) (23) complexes, indicating a similar mode of molecular recognition.

**Structural Basis for ULM Specificity**—To further characterize the binding specificity of Puf60-ULM for distinct ULMs we compared its structure with the structures of the UHM-ULM complexes of U2AF65SF1 (Protein Data Bank code 1O0P) and SPF45SF3b155 (Protein Data Bank code 2PEH). As shown in Fig. 5C, the ULM-binding region of Puf60-ULM, defined by the NMR titrations (Fig. 5B), is structurally more similar to SPF45 than to U2AF65. Helix αA in U2AF65ULM is N-terminally extended by four additional residues compared with the αA helices in Puf60-ULM and SPF45-ULM. As a consequence, the conformation of the β1-αA loops in U2AF65ULM differs considerably from Puf60-ULM or SPF45ULM. It is likely that these differences, in combination with amino acid variations in the ULM sequences (Fig. 4A), determine the specificity of the UHM-ULM complexes. For example, SF1 has a longer stretch of positively charged residues preceding the ULM-tryptophan than the SF3b155 ULMs. In the U2AF65SF1 structure, this region contacts the highly negatively charged helix αA of U2AF65ULM (10 Glu/Asp residues). Because the length helix αA of Puf60 is shorter and because it is less negatively charged (5 Glu/Asp residues), ionic interactions involving these residues should contribute less to the ULM binding by Puf60-ULM. A second specificity-mediating region in the SF1 and SF3b155 ULMs involves
the residue C-terminal of the tryptophan. The aspartate flanking Trp338 in SF3b155 forms a salt bridge with the arginine in the SPF45-UHM RYF motif (23). Because of its structural similarities to the UHM of Puf60, we speculate that analogous interactions might stabilize the complexes of Puf60-UHM with the SF3b155-ULMs around tryptophans 200 and 338. Because SF1 has an asparagine instead of aspartate at this position, a similar salt bridge cannot be formed, which may contribute to the weaker interaction of Puf60-UHM with SF1-UHM.

**DISCUSSION**

Puf60 was repeatedly found to interact with itself in yeast two-hybrid assays (26, 27), and the Puf60 UHM domain was reported to be necessary and sufficient for the dimerization of Puf60 in SDS-PAGE (21). Our analytical ultracentrifugation
and NMR data show that the UHM domain is mainly monomeric in physiological buffer, whereas SDS is required for dimerization. A crystal structure and mutational analysis reveal a dimer interface, which is stabilized by electrostatic interactions and involves the acidic β2-β3 loop of one subunit and basic residues (Lys467, Arg540, and Lys541) in the β-sheet surface of the other subunit of the dimer. The acidic β2-β3 loop is conserved in all higher eukaryotic orthologs of Puf60 but is distinct in other UHM or RRM domains (20). This suggests that Puf60 orthologs may have a similar dimerization mode, which is unique for Puf60 and not found in other UHMs.

The flexibility of the β2-β3 loop in solution (indicated by the NMR relaxation data) and the variability of the electrostatic contacts seen in the crystal structure suggest that the dimer interface is dynamic. The electrostatic nature of the dimer interface presumably contributes to the stability of the Puf60-UHM dimer in SDS-PAGE (21) (Fig. 2, D and E). Because the dimerization interface is stabilized by electrostatic contacts, the SDS alkyl chains might not be able to energetically favor the solvation of the UHM monomer.

We found that a longer construct, comprising the two central RRM domains and the UHM, is also largely monomeric in the

![FIGURE 5. NMR chemical shift perturbation of Puf60-UHM/ULM interactions. A, 1H,15N HSQC spectra of free 15N-labeled Puf60-UHM (black) and upon addition of ULM peptides derived from SF1, SF3b155, and U2AF65 (red). B, ribbon representation of the Puf60-UHM structure colored according to the extent of chemical shift perturbation induced by addition of the peptide ligands indicated above (white, no perturbation; red, strong perturbation). C, superposition of the Puf60-UHM structure (blue) onto the structures of SPF45-UHM/SF3b-ULM (left panel, gray; Protein Data Bank code 2PEH) and U2AF65-UHM/SF1-ULM (right panel, gray; Protein Data Bank code 1O0P).]
absence of SDS. Crichlow et al. (48) report a weak tendency of the two central RRM domains to dimerize in the presence of single-stranded DNA. Therefore, it is likely that the oligomerization of Puf60 observed in yeast two-hybrid experiments involves the UHM and the two central RRM domains of Puf60 and additionally requires binding of a ligand. Potentially, ligand-induced dimerization of Puf60-UHM could involve binding of tandem ULMs. However, our experiments with the tandem ULM motifs in SF3b155 and Prp16 did not provide evidence for such a mechanism.

Detergent-induced oligomerization has been reported for several membrane-associated proteins (52–56). No experimental evidence for a functional role of the SDS-induced dimerization of Puf60-UHM is known. However, it is possible that detergent-induced (or lipid-induced) dimerization might play a role for the molecular functions of Puf60. Alternatively, SDS may resemble a putative, as yet unknown ligand of Puf60.

Puf60 was reported to interact directly with U2AF65 (26, 27). Our data provide a rationale for how the two proteins interact and suggest that a minimal binding interface involves the ULM sequence of U2AF65 and the UHM domain of Puf60. Note that binding of Puf60 to the U2AF65 ULM can only occur if this ULM is not already bound by U2AF35-UHM, which has a significantly higher affinity. Thus, for this interaction to occur in vivo, there should be a population of U2AF65 molecules that is not bound to U2AF35 in the nucleus.

It was shown in pulldown experiments from nuclear extract that Puf60 associates with SF3b155 (28). We suggest that this interaction likely involves direct binding of Puf60-UHM to ULM sequences in the N terminus of SF3b155. Interestingly, the Puf60-UHM and U2AF65-UHM have distinct binding affinities for ULMs. Puf60-UHM binds only weakly to SF1-ULM, whereas this ULM strongly interacts with U2AF65-UHM. Furthermore, Puf60-UHM has a stronger affinity to SF3b155(194–229) than to SF3b155(317–357), whereas the opposite is found for U2AF65-UHM (Table 2). The affinity differences of these two UHM domains are rationalized by comparing structural models of these interactions. As shown in Fig. 5C, the ULM-binding region of the two UHMs is significantly different, which may be linked to the distinct binding preferences.

Our biochemical data imply that Puf60 and U2AF65 can bind to the N terminus of SF3b155 simultaneously and noncompetitively (Fig. 6). The mutual enhancement of splicing activation by these two splicing factors (28) could thus involve simultane-

References

Dimerization and Protein Binding Specificity of the U2AF Homology Motif of the Splicing Factor Puf60
Lorenzo Corsini, Michael Hothorn, Gunter Stier, Vladimir Rybin, Klaus Scheffzek, Toby J. Gibson and Michael Sattler

doi: 10.1074/jbc.M805395200 originally published online October 29, 2008

Access the most updated version of this article at doi: 10.1074/jbc.M805395200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2008/10/29/M805395200.DC1

This article cites 58 references, 13 of which can be accessed free at
http://www.jbc.org/content/284/1/630.full.html#ref-list-1