Supplementary Information for Jenkins, Laird, and Kielkopf
"A Broad Range of Conformations Contribute to the Solution Ensemble of the Essential Splicing Factor U2AF^{65}"

Table S1. Boundaries of U2AF^{65} constructs used for structural studies.

Sample	Technique	Deposition ID	Reference	Residue range (human U2AF^{65}, Refseq NP_009210)
U2AF^{65}1,2	SAXS	BIOISIS ID 1U2FKP	[1]	148-336
U2AF^{65}1,2$_{FIR}$	SAXS	BIOISIS ID 2U2FKP	this work	136-347
dU2AF^{65}1,2	X-ray crystallography	PDB ID 2G4B	[2]	148-336; internal deletion of residues 238-257
'open' U2AF^{65}1,2	NMR	PDB ID 2YH1	[3]	148-342
'closed' U2AF^{65}1,2	NMR	PDB ID 2YH0	[3]	148-342

Table S2. Dimensions and discrepancy values derived from scattering data.[a]

Sample	R_G (Å)	D_{max} (Å)	$\chi^2_{structure}$ (U2AF^{65} 1,2 / U2AF^{65} 1,2$_{FIR}$) $q \leq 0.20$ Å^{-1}	$\chi^2_{structure}$ (U2AF^{65} 1,2 / U2AF^{65} 1,2$_{FIR}$) $q \leq 0.30$ Å^{-1}	χ^2_{EOM} (1-PDB / 2-PDB / 20-PDB)
U2AF^{65}1,2	24.57± 0.03	80 ± 5	-	-	2.34 / 0.94 / 0.91
U2AF^{65} 1,2$_{FIR}$	25.61 ± 0.03	85 ± 5	-	-	2.16 / 1.09 / 0.91
dU2AF^{65}1,2 PDB ID: 2G4B[b]	23.6 (20.5)	70 (65 ± 5)	2.9 / 2.5	3.5 / 3.4	-
FIR1,2 PDB ID: 2QFJ	19.6	58	19.3 / 25.7	22.1 / 23.4	-
'open' U2AF^{65}1,2 PDB ID: 2YH1	20.4	58	21.4 / 21.7	24.9 / 21.3	-
'closed' U2AF^{65}1,2 PDB ID: 2YH0	19.5	55	20.1 / 26.0	23.4 / 22.4	-

[a] R_G, radius of gyration value from *P(r)* analysis; D_{max}, maximum intraparticle distance; $\chi^2_{structure}$, discrepancy between experimental U2AF^{65} 1,2 or U2AF^{65} 1,2$_{FIR}$ SAXS data and the indicated structure calculated as described in Methods; χ^2_{EOM}, discrepancy of ensemble fit to indicated SAXS data.
[b] Experimentally-determined dU2AF^{65}1,2 dimensions based on previously reported SAXS data[1] are given in parentheses for comparison.

Figure S1. Quality of small-angle X-ray scattering data. (A) X-ray scattering profiles of the final merged U2AF^{65}1,2 (blue circles) and U2AF^{65}1,2$_{FIR}$ scattering data (green circles) compared with data calculated from the 2-PDB EOM model (black, solid lines). (B, C) Guinier plots of *ln[I(q)] versus* q^2 in the range q^*R_G <1.3 respectively for U2AF^{65}1,2 (blue) and U2AF^{65}1,2$_{FIR}$ (green).

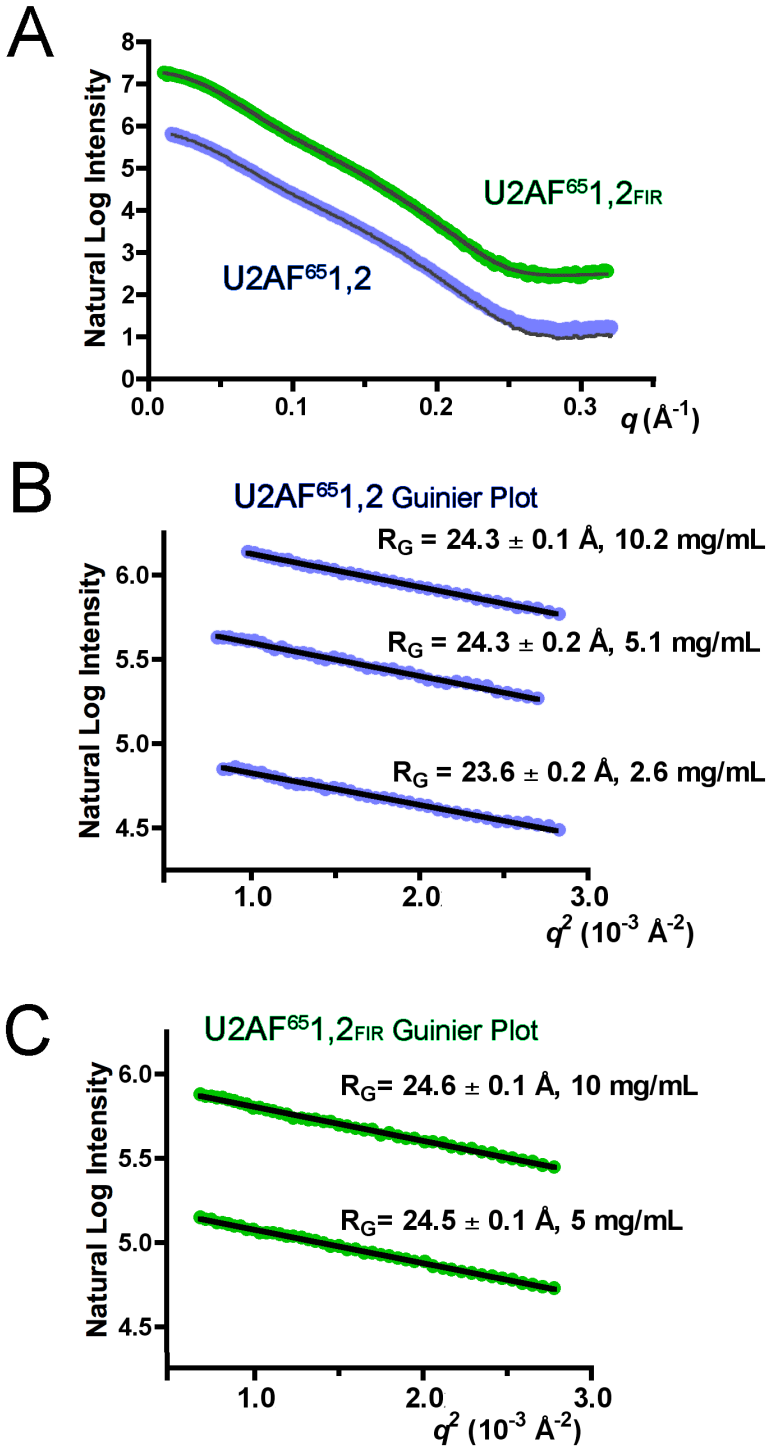
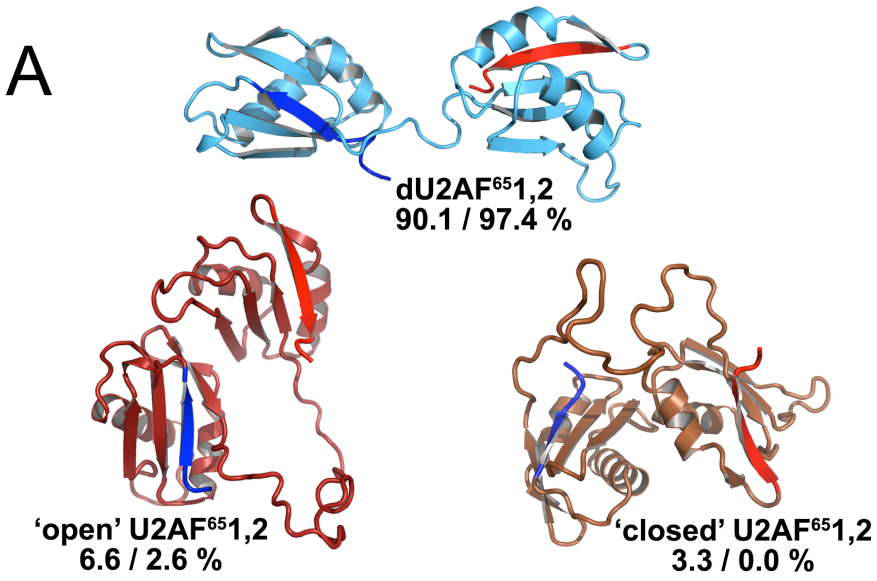


Figure S2. Minimal Ensemble Search (MES)[6] starting with the three known structures composed of tandem RRMs as the starting pool for selection of a mixture that best fits the scattering curve. (A) The selected percentages of each structure in the final mixture are given for the U2AF^{65}1,2 / U2AF^{65}1,2$_{FIR}$ SAXS data. (B) X-ray scattering profiles of the final merged U2AF^{65}1,2 (blue circles) and U2AF^{65}1,2$_{FIR}$ scattering data (green circles) compared with data calculated from the selected MES ensembles (black, solid lines). The χ^2 values calculated using MES with default values for solvent density / atomic radii were 3.60 / 4.14 (against U2AF^{65}1,2 data) or 5.17 / 5.20 (against U2AF^{65}1,2$_{FIR}$ data) respectively for the MES-selected ensembles composed of the indicated percentages of each structure *versus* the dU2AF^{65}1,2 crystal structure alone.



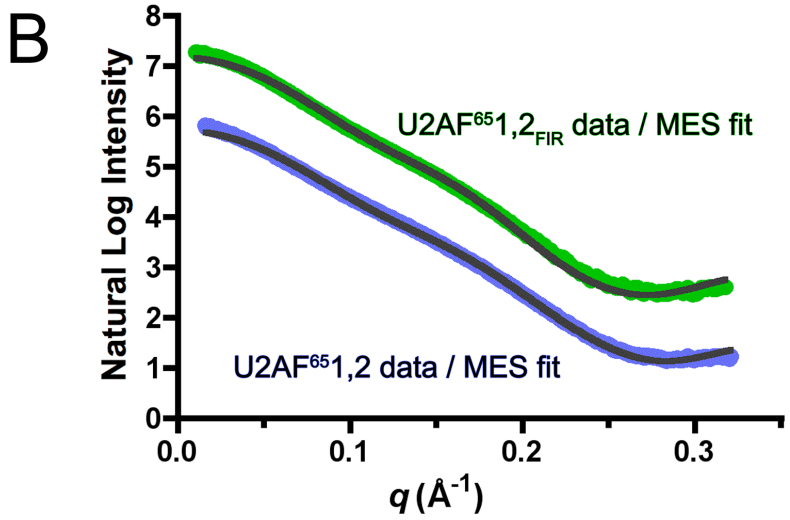


Figure S3. Selection of solution ensembles to fit the U2AF^{65}1,2 (blue) or U2AF^{65}1,2$_{FIR}$ (green) SAXS data and a randomized starting pool composed of rigid body RRM1 and RRM2 connected in a variety of proximities by *ab initio* linkers. (A) Plot of χ^2 values resulting from GAJOE selection[5] *versus* number of PDB coordinate sets in the ensemble. Similar results were obtained using MES[6] (χ^2 3.0 / 2.3 for one PDB decreased to 1.4 / 1.7 for two or 1.1 / 1.1 for 20-PDB ensemble fits of U2AF^{65}1,2 / U2AF^{65}1,2$_{FIR}$ using an identical randomized input pool generated using RANCH). (B, C) The 2-PDB ensemble fits of (B) U2AF^{65}1,2 and (C) U2AF^{65}1,2$_{FIR}$ selected using GAJOE[5]. Representative selected structures with the given R_G are inset. The radii of gyration (R_G) of the structures are plotted on the x-axis, against the frequency with which a structure of a given R_G occurs in the selected ensemble on the y-axis. Gray dashed lines represent the randomized starting pool of structures. Solid lines plot the R_G of the selected pool.

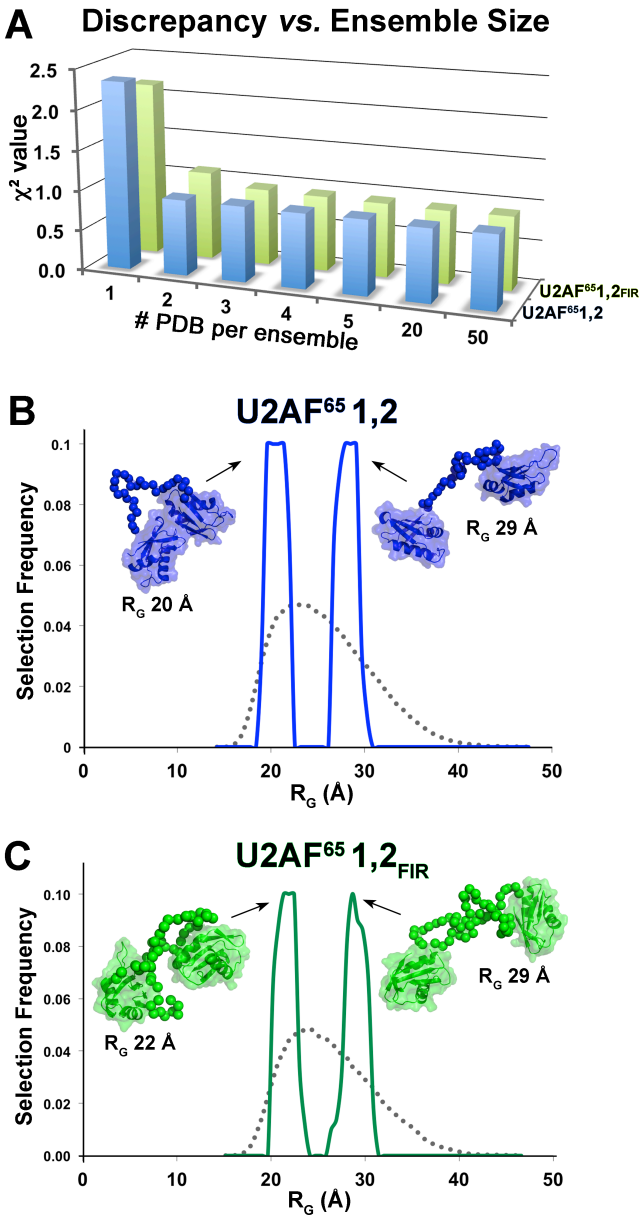
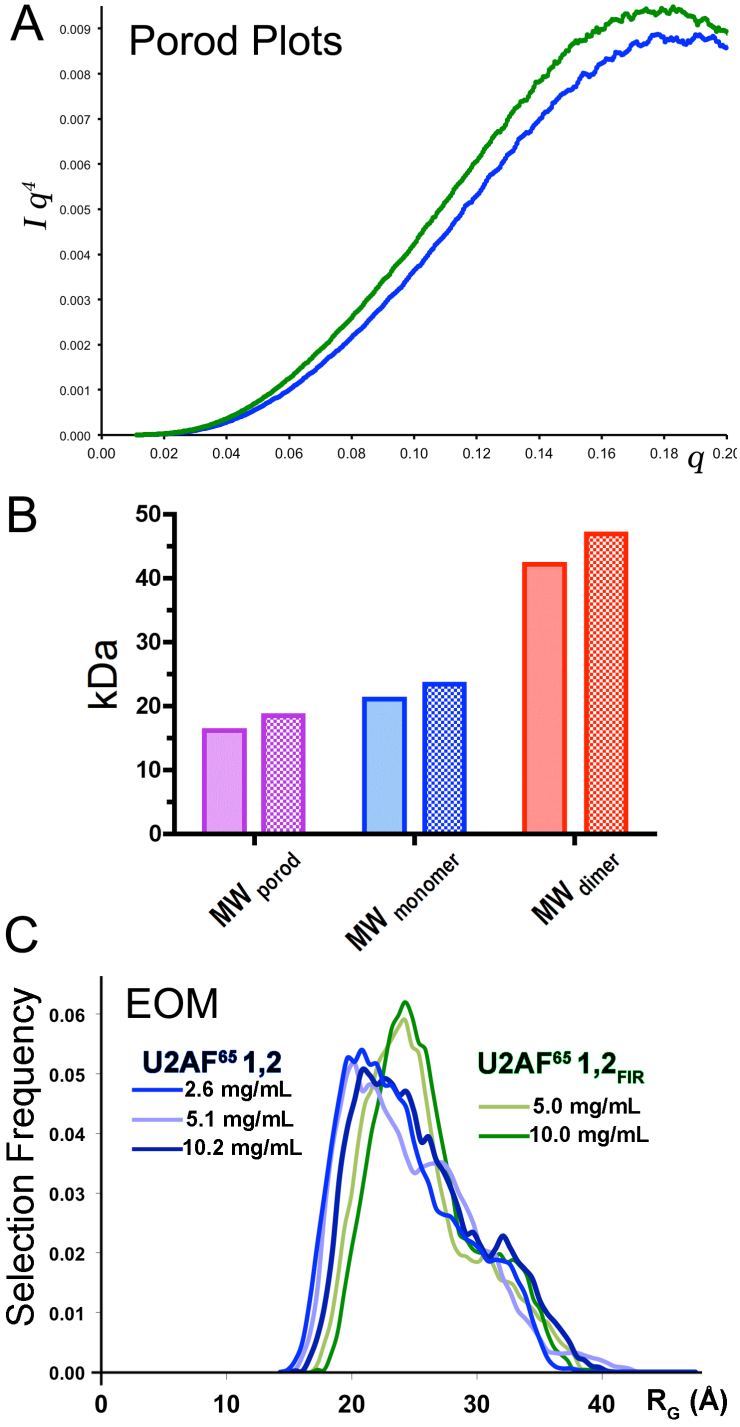


Figure S4. SAXS data are consistent with monomeric protein samples. (A) Porod plots of $I(q) * q^4$ *versus* q in the Porod-Debye region (q <0.2) for U2AF^{65}1,2 (blue) and U2AF^{65}1,2$_{FIR}$ (green). (B) Molecular weights estimated from the Porod volumes ($V_{porod} \approx 1.5 * MW_{porod}$) (purple) [7, 8] compared with molecular weights of the monomeric (blue) or dimeric (red) U2AF^{65}1,2 (filled) and U2AF^{65}1,2$_{FIR}$ (hatched) proteins. (C) Comparison of 20-PDB EOM fits of unmerged SAXS data demonstrates concentration-independent distributions.



Supplementary Methods

Protein Purification and RNA Preparation
The boundaries of the U2AF^{65} constructs (NCBI Refseq NP_009210) used in this study include U2AF^{65}1,2 (residues 148-336), U2AF^{65}1,2$_{FIR}$ (residues 136-347) (Figure 1A and Table S1). All proteins were expressed and purified as glutathione-*S*-transferase (GST) fusion proteins from pGEX-6p vectors followed by GST cleavage as previously described [9]. The elution volumes in a final step of size exclusion chromatography (Superdex-75, GE Healthcare) and dynamic light scattering analyses were consistent with monodisperse monomeric proteins. Protein concentrations were estimated using the absorbance at 280 nm and calculated molar extinction coefficients [10].

Fluorescence Anisotropy
Fluorescence anisotropy changes were measured following the addition of U2AF^{65}1,2$_{FIR}$ protein to a solution of 30 nM 5'-fluorescein-labeled 20-uridine RNA in 100 mM NaCl, 25 mM Hepes, pH 6.8. Data were fit to obtain the apparent equilibrium dissociation constant (K_D) as described [1].

Small-Angle X-ray Scattering Data Collection and Analyses
The SAXS data for U2AF^{65}1,2$_{FIR}$ were collected at the SIBYLS Beamline 12.3.1 of the Advanced Light Source (Lawrence Berkeley National Laboratory) by the same procedure as the U2AF^{65}1,2 SAXS data described previously [1, 11]. SAXS data were collected at two concentrations for 6- and 60-s exposures followed by a 6 s exposure to check for radiation damage. Buffer corrections used the size exclusion chromatography buffer, which was optimized to eliminate detectable sample aggregation. The sample buffers for U2AF^{65}1,2 and U2AF^{65}1,2$_{FIR}$ were respectively 100 mM NaCl, 15 mM HEPES pH 7.4, 0.2 mM TCEP and 250 mM NaCl, 15 mM HEPES pH 7.4, 3% v/v glycerol, 0.2 mM TCEP. The nearly identical results with the two samples indicate that the difference in salt concentration and addition of 150 mM NaCl had little effect on the conformational pool. Data from the low and high resolution ranges of the respective short and long exposures and the two different U2AF^{65}1,2$_{FIR}$ concentrations were scaled and merged using the Primus suite to obtain the final data files for each sample [12].

Scattering data was calculated from PDB files using CRYSOL [4] using data in the indicated q range. For consistency, the B-factors of NMR PDB files were reset to 20 Å^2 and explicit hydrogen atoms were excluded from the calculations. Only polypeptide coordinates were used in cases where coordinate files contained water molecules or ligands. In all cases, 3.2 Å maximum atomic radii, which correspond to the average radius of protein C, N or O atoms plus the radius of a water molecule (1.4 Å), improved agreement between the calculated and experimental data as compared to the 1.8 Å default. For most cases, 0.38 e Å^{-3} solvent density improved agreement with the SAXS data compared with the default value (0.33 e Å^{-3} for pure water) or slightly higher density (0.42 e Å^{-3}), consistent with the composition of the SAXS buffer solution. Exceptions were found for the NMR structures, for which the default solvent density of water (0.33 e Å^{-3}) slightly improved the fits with $q \leq 0.3$ Å^{-2} data. Setting the solvent density to 0.33 e Å^{-3} also slightly improved the $q \leq 0.2$ Å^{-1} fit of 'closed' NMR structure (PDB ID 2YH0).

Several observations ruled out the possibility that the poor fits of the SAXS data with the compact NMR models was produced by aggregates in the SAXS samples (Figure S1, S4): (*i*) plateau shapes of the scattering curves at low q values (Figure S1A); (*ii*) absence of concentration dependent differences in the Guinier radii of gyration (R_G) (Figure S1B,C); (*iii*) Porod volumes consistent with monomeric proteins (Figure S4A,B); (*iv*) Concentration-independent EOM fits (Figure S4C). Ionic strengths also were unlikely to cause the large differences between the SAXS data and the PRE

structures, since the U2AF^{65}1,2 and U2AF^{65}1,2$_{FIR}$ SAXS samples contained different salt concentrations due to lower solubility of the latter protein (Supporting Methods) yet produced comparable *P(r)* plots and discrepancies values (Figure 1C,D).

For the input pool of know structures, the polypeptide coordinates of the 'closed', 'open' NMR structures and dU2AF^{65}1,2 crystal structure with B-factors reset to 20 Å^2 were used in a minimal ensemble search (MES)[6] that selects the fractional populations by the FoXS server (http://modbase.compbio.ucsf.edu/foxs/). To generate the randomized input ensembles, a pool of 10,000 input conformations was generated using the RANCH component of EOM[5]. The separated RRM structures from PDB 2G4B (RRM1: residues 148:228; RRM2: residues 260:334) were connected in randomized configurations by *ab initio* linkers (for residues 229-259). When fitting the U2AF^{65}1,2 data, the C-terminal two residues (335 and 336) were assumed to be flexible in the absence of RNA and included in the *ab initio* portion of the models. When fitting U2AF^{65}1,2$_{FIR}$ data, the additional N- and C-terminal residues flanking the RRMs were included as *ab initio* models (residues 136-147 and 335-347). Ensembles were fit against the entire q-range (0.01-0.32) of the final merged scattering data using either MES[6] or the GAJOE component of EOM[5] as indicated. The number of structures per ensemble was altered by systematically testing 1, 2, 3, 4, 5, 20, and 50 'number of curves per ensemble' in GAJOE or 'subset-size' in MES. Default values were input with the exception that the number of fitting runs in MES was increased from the default of two to five as recommended[6].

Supplementary References

(1) Jenkins, J. L., Shen, H., Green, M. R., and Kielkopf, C. L. (2008) Solution conformation and thermodynamic characteristics of RNA binding by the splicing factor U2AF^{65}, *J Biol Chem 283*, 33641-33649.

(2) Sickmier, E. A., Frato, K. E., Shen, H., Paranawithana, S. R., Green, M. R., and Kielkopf, C. L. (2006) Structural basis of polypyrimidine tract recognition by the essential pre-mRNA splicing factor, U2AF^{65}, *Mol Cell 23*, 49-59.

(3) Mackereth, C. D., Madl, T., Bonnal, S., Simon, B., Zanier, K., Gasch, A., Rybin, V., Valcarcel, J., and Sattler, M. (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF, *Nature 475*, 408-411.

(4) Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) CRYSOL- a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. , *J Appl Crys 28*, 768-773.

(5) Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M., and Svergun, D. I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering, *J Am Chem Soc 129*, 5656-5664.

(6) Pelikan, M., Hura, G. L., and Hammel, M. (2009) Structure and flexibility within proteins as identified through small angle X-ray scattering, *Gen Physiol Biophys 28*, 174-189.

(7) Mertens, H. D., and Svergun, D. I. (2010) Structural characterization of proteins and complexes using small-angle X-ray solution scattering, *J Struct Biol 172*, 128-141.

(8) Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law, *Biopolymers 95*, 559-571.

(9) Sickmier, E. A., Frato, K. E., and Kielkopf, C. L. (2006) Crystallization and preliminary X-ray analysis of a U2AF^{65} variant in complex with a polypyrimidine-tract analogue by use of protein engineering, *Acta Crystallogr Sect F 62*, 457-459.

(10) Cavaluzzi, M. J., and Borer, P. N. (2004) Revised UV extinction coefficients for nucleoside-5'-monophosphates and unpaired DNA and RNA, *Nucleic Acids Res 32*, e13.

(11) Gupta, A., Jenkins, J. L., and Kielkopf, C. L. (2010) RNA induces conformational changes in the SF1-U2AF^{65} splicing factor complex, *J Mol Biol 405*, 1128-1138.

(12) Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J., and Svergun, D. I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J Appl Crys 36*, 1277-1282.