# Supplementary Data

# Structure, phosphorylation and U2AF65 binding of the N-terminal domain of splicing factor 1 during 3' splice site recognition

Yun Zhang[1,2#], Tobias Madl[1,2,3,#,*], Ivona Bagdiul[4], Thomas Kern[1,2] ,Hyun-Seo Kang[1,2], Peijian Zou[1,2,5], Nina Mäusbacher[6], Stephan A. Sieber[6], Angela Krämer[4], and Michael Sattler[1,2,5,*]

# Supplementary Methods

### Measurement of proteins with mass spectrometry

2-5 µg of protein were desalted on OmixC4 Tips (Agilent Technologies). Briefly, C4 Tips were washed with 10 µl of elution buffer (50% acetonitrile, 0.1% formic acid) and twice equilibrated with 0.1% formic acid. The sample was loaded on the C4 Tips, followed by two washing steps with 0.1% formic acid and elution in a final volume of 20 µl of elution buffer. 18 µl of the sample were injected byan Ultimate 3000 HPLC (Thermo Scientific) coupled online to a LTQ-FT Ultra (Thermo Scientific). Spectra of full-length proteins were recorded from 600 – 2,000 m/z with a resolution of 100,000 at m/z 400. Protein-containing spectra were averaged and deconvoluted with the software ProMass 2.8 (Thermo Scientific). For the analysis default settings for large proteins were applied in a mass range of 10,000 to 80,000 Da.

# Supplementary Table

**Supplementary Table 1.** NMR and refinement statistics for $SF1^{HH}$ and $SF1^{NTD}$-$U2AF65^{UHM}$ ensembles*

| | $SF1^{HH}$ | $SF1^{NTD}$-$U2AF65^{UHM}$ |
|---|---|---|
| **NMR distance, dihedral and RDC restraints** | | |
| Distance restraints | | |
| Hydrogen bonds† | 38 | - |
| Intra-molecular | 2021 | 14 |
| short-range \|i-j\|<=1 | 1140 | - |
| medium-range 1<\|i-j\|<5 | 479 | 9 |
| long-range \|i-j\|>=5 | 366 | 5 |
| Inter-molecular | - | 22 |
| Dihedral angle restraints‡ | 188 | 130 |
| Protein $\phi/\psi$ | 94/94 | 65/65 |
| Residual dipolar couplings | 122 | - |
| | | |
| **Structure statistics** | | |
| Violations (mean and s.d.) | | |
| Dihedral angle violations > 5° | 0 | 0 |
| Distance violations > 0.5 Å | 0 | 0 |
| $Q^{RDC}$ | 0.27 ± 0.00 | - |
| Deviations from idealized geometry | | |
| Bond lengths (Å) | 0.010 ± 0.000 | 0.004 ± 0.000 |
| Bond angles (°) | 1.190 ± 0.028 | 0.492 ± 0.005 |
| Impropers (°) | 1.362 ± 0.109 | 0.800 ± 0.058 |
| Average pairwise r.m.s.d. (Å)§ | | |
| Heavy | 0.94 ± 0.09 | 1.38 ± 0.63 |
| Backbone | 0.40 ± 0.06 | 1.27 ± 0.62 |
| Ramachandran | | |
| Core | 98.5 | 84.7 |
| Allowed | 1.5 | 14.6 |
| Generous | 0.0 | 0.7 |
| Disallowed | 0.0 | 0.0 |
| Fit to SAXS data (c) | - | 1.29 ± 0.11 |

*Statistics are calculated for ten lowest-energy refined structures of a total of 100 structures.

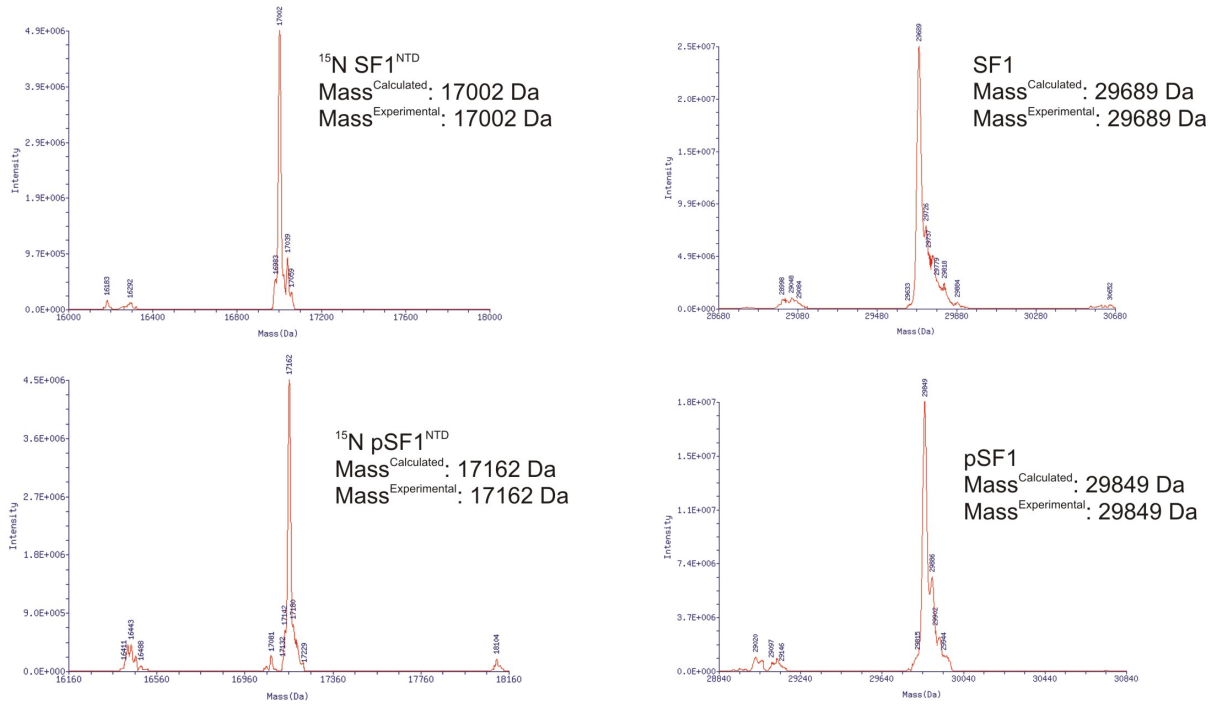
†Hydrogen bond restraints seen in the input structures were maintained by distance restraints, if they were further supported by secondary chemical shifts.

‡ Dihedral angle restraints were derived from secondary chemical shifts using TALOS+

§ Backbone atoms within $SF1^{HH}$ (45-69, 97-127), and $SF1^{NTD}$ (38-40, 45-69, 97-127) - $U2AF65^{UHM}$ (375-473)

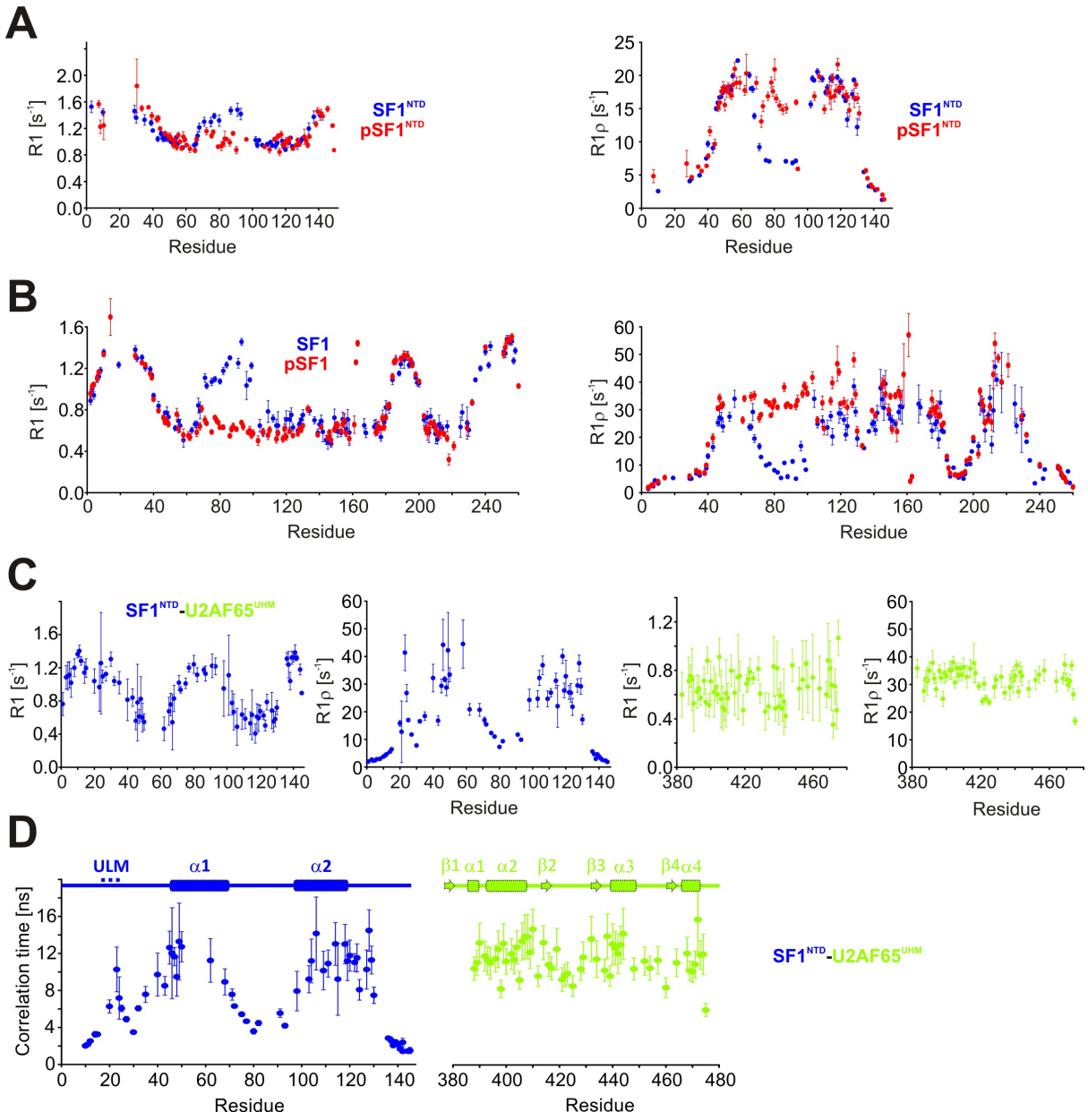
# Supplementary Figures

## Supplementary Fig. 1



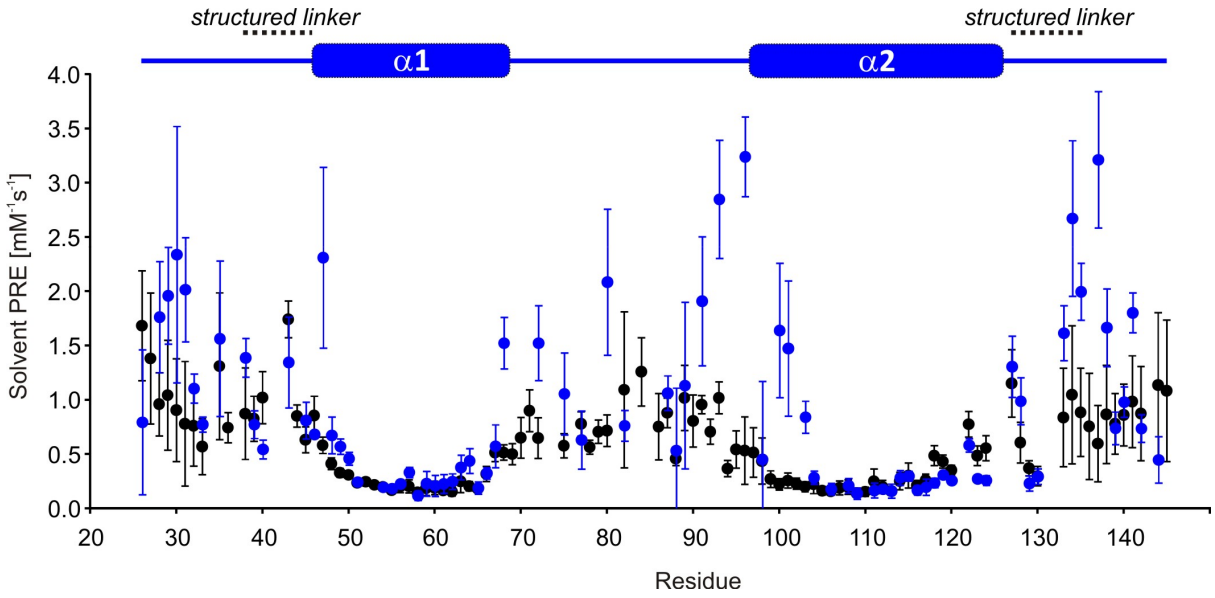

**Supplementary Figure 1: Mass spectrometry analysis of SF1 phosphorylation.** Deconvoluted spectra of non-phosphorylated $^{15}$N-SF1$^{NTD}$ and SF1 (top left and top right) are shown in comparison to their phosphorylated state (bottom left and right). The mass difference of 160 Da confirms the addition of exactly two phosphate groups.

## Supplementary Fig. 2



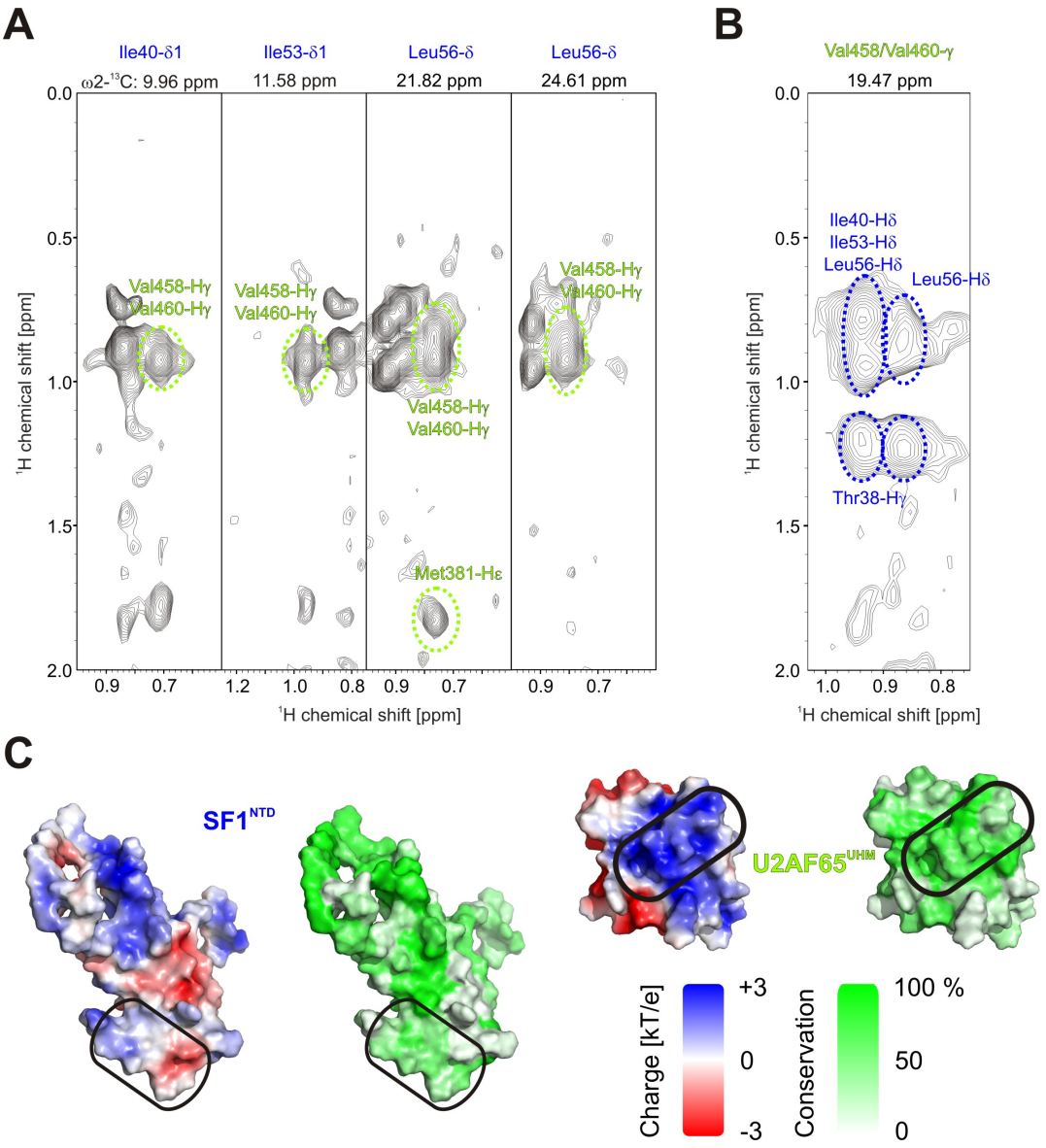

**Supplementary Figure 2: NMR relaxation data.** $^{15}$N backbone amide longitudinal ($R_1$) and transverse relaxation rates $R_{1\rho}$ for non-phosphorylated (blue) and phosphorylated (red) SF1$^{NTD}$ (A) and SF1 (B), respectively. Relaxation data was measured at 600 and 800 MHz proton larmor frequency for (p)SF1$^{NTD}$ and (p)SF1, respectively. (C) SF1$^{NTD}$ (blue) and U2AF65$^{UHM}$ (green) in the SF1$^{NTD}$-U2AF65$^{UHM}$ complex at 750 MHz proton larmor frequency. (D) Residue-specific local correlation times $\tau_c$ for SF1$^{NTD}$ (blue) and U2AF65$^{UHM}$ (green) in the SF1$^{NTD}$-U2AF65$^{UHM}$ complex. Correlation time values were determined from $^{15}$N relaxation data as described in Methods. The secondary structures are depicted above the diagram.

## Supplementary Fig. 3



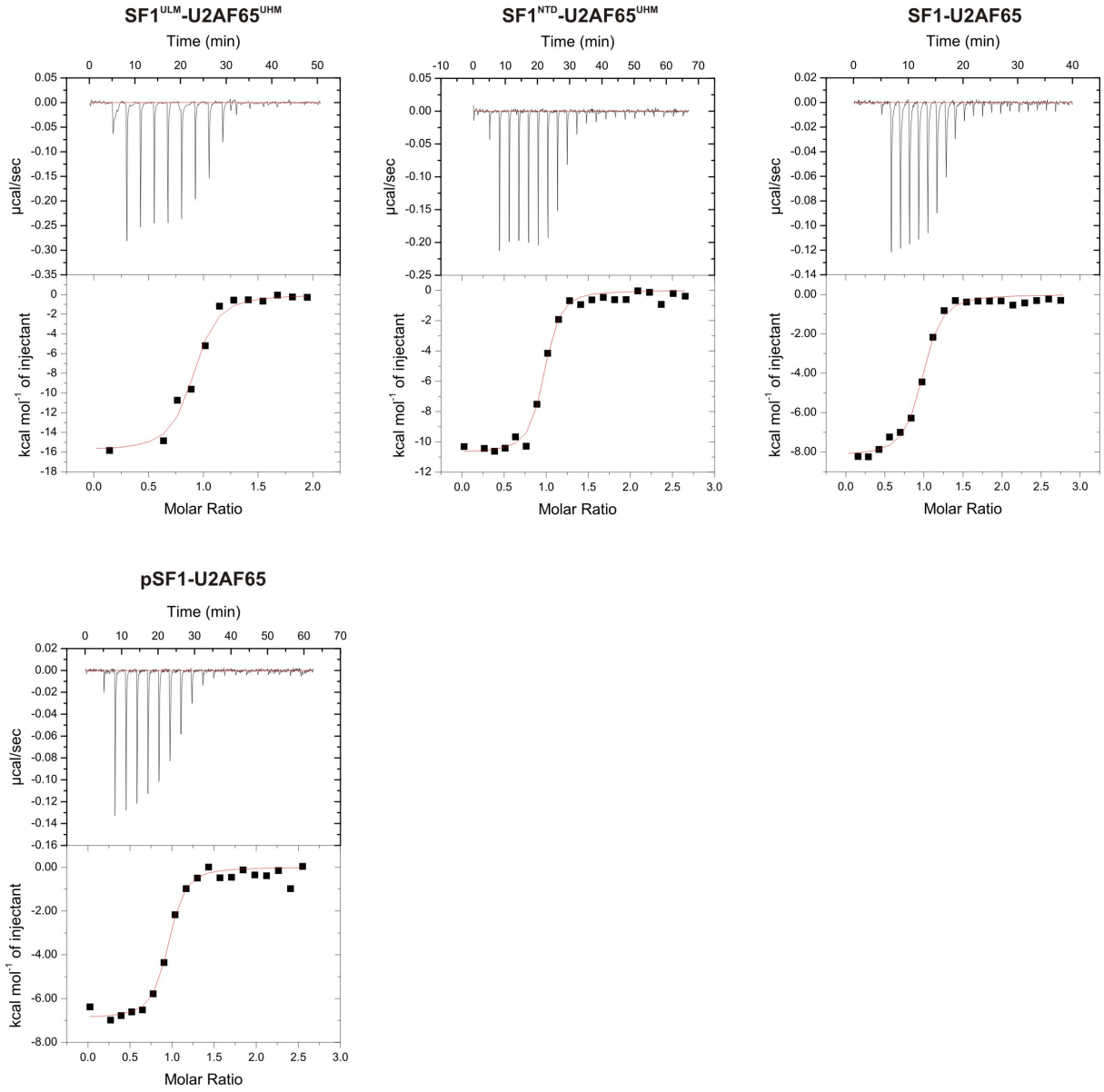

**Supplementary Figure 3: Comparison of back-calculated and experimental solvent PREs.** Experimental solvent PRE rates of backbone amide protons of $SF1^{HH}$ (blue) were compared to values back-calculated from the ensemble of NMR structures (black). Values for the structured region including helices α1-α2 are in excellent agreement. For flexible regions, experimental values are higher than the back-calculated values due to chemical exchange between amide protons in these parts with water molecules which experience a large relaxation enhancement because of transient binding to Gd(DTPA-BMA).The secondary structure is depicted above the diagram.

## Supplementary Fig. 4



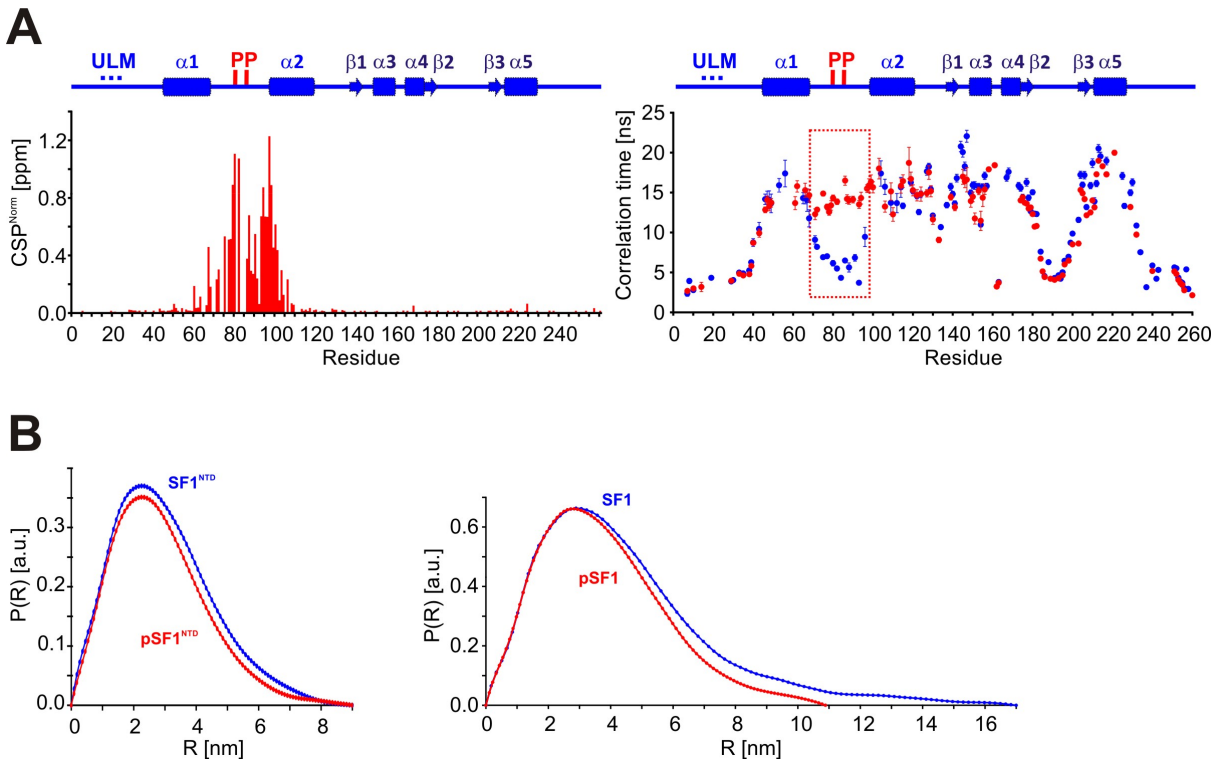

**Supplementary Figure 4: Analysis of the SF1$^{NTD}$-U2AF65$^{UHM}$ complex.** Strip plots from filtered/edited NOE experiments to detect SF1$^{NTD}$-U2AF65$^{UHM}$ NOEs (A,B).Strip plots are shown for ILV methyl-protonated [U-$^{2}$H,$^{13}$C,$^{15}$N] SF1$^{NTD}$ in complex with unlabeled U2AF65$^{UHM}$ (A) and *vice versa* (B). The methyl group of the labelled subunit is shown on top and the cross peaks to the unlabeled binding partner are shown in the strip plots. The colours are according to the colour coding used in Fig. 1. (C) Surface representation of the individual subunits of the SF1$^{NTD}$-U2AF65$^{UHM}$ complex coloured according to electrostatic surface potential and conservation. The electrostatic surface potential is coloured at 3 kT/e$^{-}$ for positive (blue) or negative (red) charge potential using the program APBS (1). For the surface representations of the individual subunits, they are rotated by +/- 90° with respect to the orientation in Fig. 4A to expose the binding interface (black ellipse). Sequence conservation of SF1$^{NTD}$ and U2AF65$^{UHM}$ was generated sing the ConSurf and ConSeq webservers (2-3).

# Supplementary Fig. 5



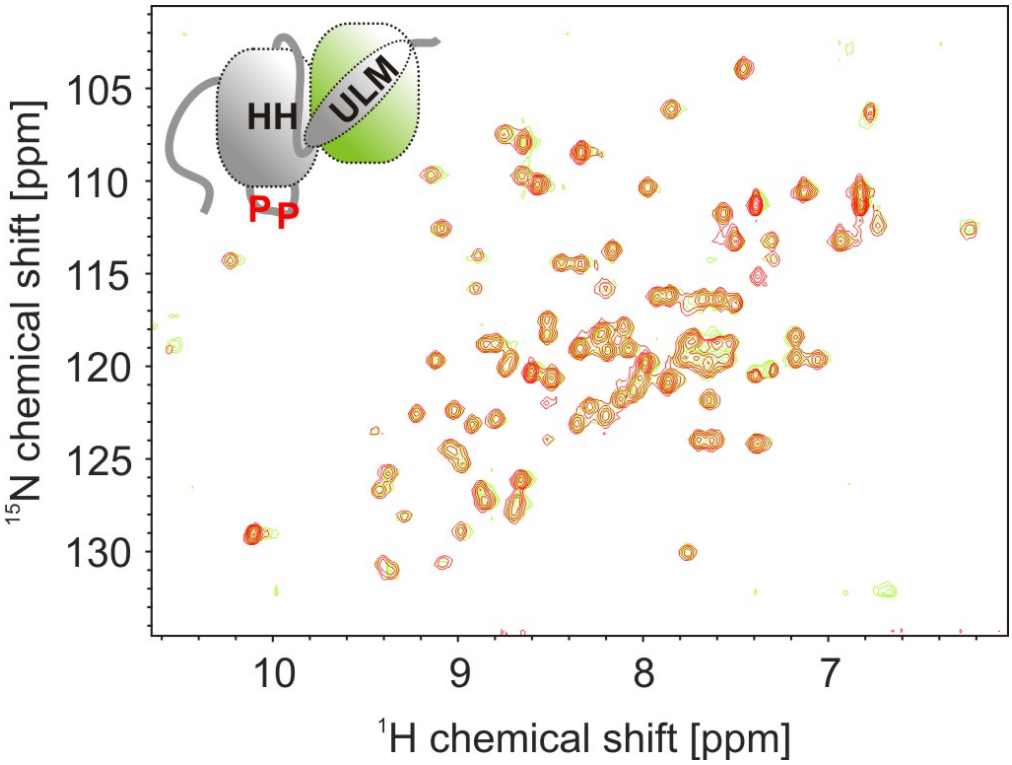

**Supplementary Figure 5: Isothermal titration calorimetry data.** Measurements were carried out at 25 °C. Dilution heats determined by titrating proteins or RNA into the corresponding buffer were in the range of the heat effects observed at the end of the titration (data not shown) and were subtracted for the analysis.

## Supplementary Fig. 6




**Supplementary Figure 6: Tandem serine phosphorylation of SF1 at Ser80/82.** (A) Chemical shift perturbations (red) and residue-specific local correlation times $\tau_c$ are shown for SF1 (blue) and pSF1 (red). The secondary structures and the phosphorylation sites are depicted above the diagram. The tandem phosphorylated linker region which rigidifies upon phosphorylation is framed.(B) SAXS data showing comparisons of radial density distributions of non-phosphorylated and phosphorylated $SF1^{NTD}$ and SF1, respectively.

## Supplementary Fig. 7




**Supplementary Figure 7: Effect of SF1$^{NTD}$ phosphorylation on U2AF65$^{UHM}$.** Superposition of $^{1}$H,$^{15}$N HSQC NMR spectra of U2AF65$^{UHM}$ in complex with unlabeled non-phosphorylated (green) and phosphorylated SF1$^{NTD}$ (red), respectively. No residues are shifted upon SF1$^{NTD}$phosphorylation, indicating that the phosphorylated residues in SF1 are remote from the SF1$^{NTD}$/U2AF65$^{UHM}$ interface.

## Supplementary References


1. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U S A*, **98**, 10037-10041.
2. Ashkenazy, H., Erez, E., Martz, E., Pupko, T. and Ben-Tal, N. (2010) ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res.*, **38**, W529-533.
3. Berezin, C., Glaser, F., Rosenberg, J., Paz, I., Pupko, T., Fariselli, P., Casadio, R. and Ben-Tal, N. (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics*, **20**, 1322-1324.