Structure, Volume *21*

Supplemental Information

Phosphorylation Drives a Dynamic Switch in Serine/Arginine-Rich Proteins

ShengQi Xiang, Vytautas Gapsys, Hai-Young Kim, Sergey Bessonov, He-Hsuan Hsiao, Sina Möhlmann, Volker Klaukien, Ralf Ficner, Stefan Becker, Henning Urlaub, Reinhard Lührmann, Bert de Groot, and Markus Zweckstetter

Supplemental Figures and Legends

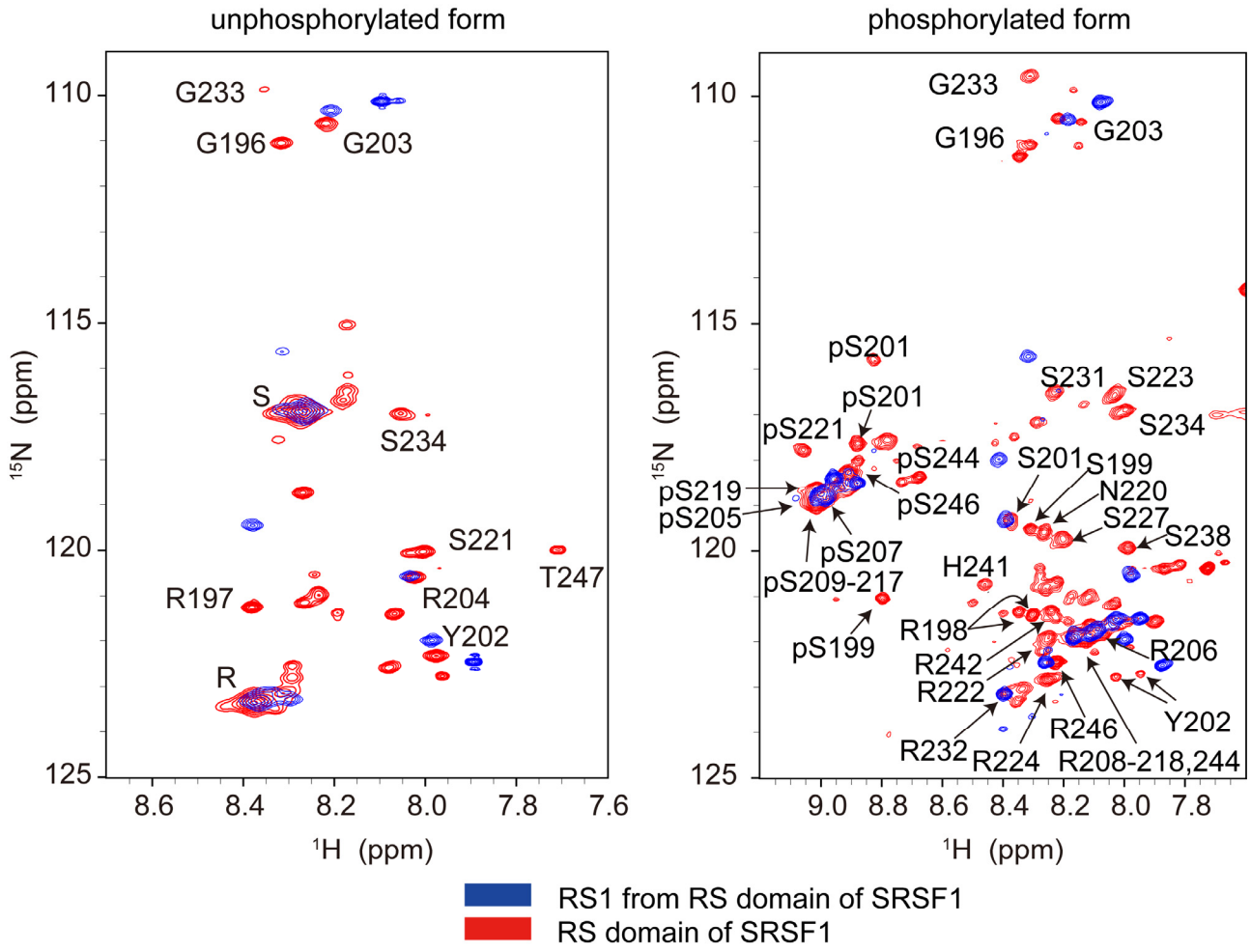

Figure S1, related to Figure 1. Superposition of ^{15}N-^{1}H HSQC spectra of SRSF1(RS) (red) and SRSF1(RS1) (blue) in the unphosphorylated state (left panel) and the phosphorylated state (right panel). Resonance assignments are indicated. Note the heterogeneity of phosphorylation, for example for S201 and S199.

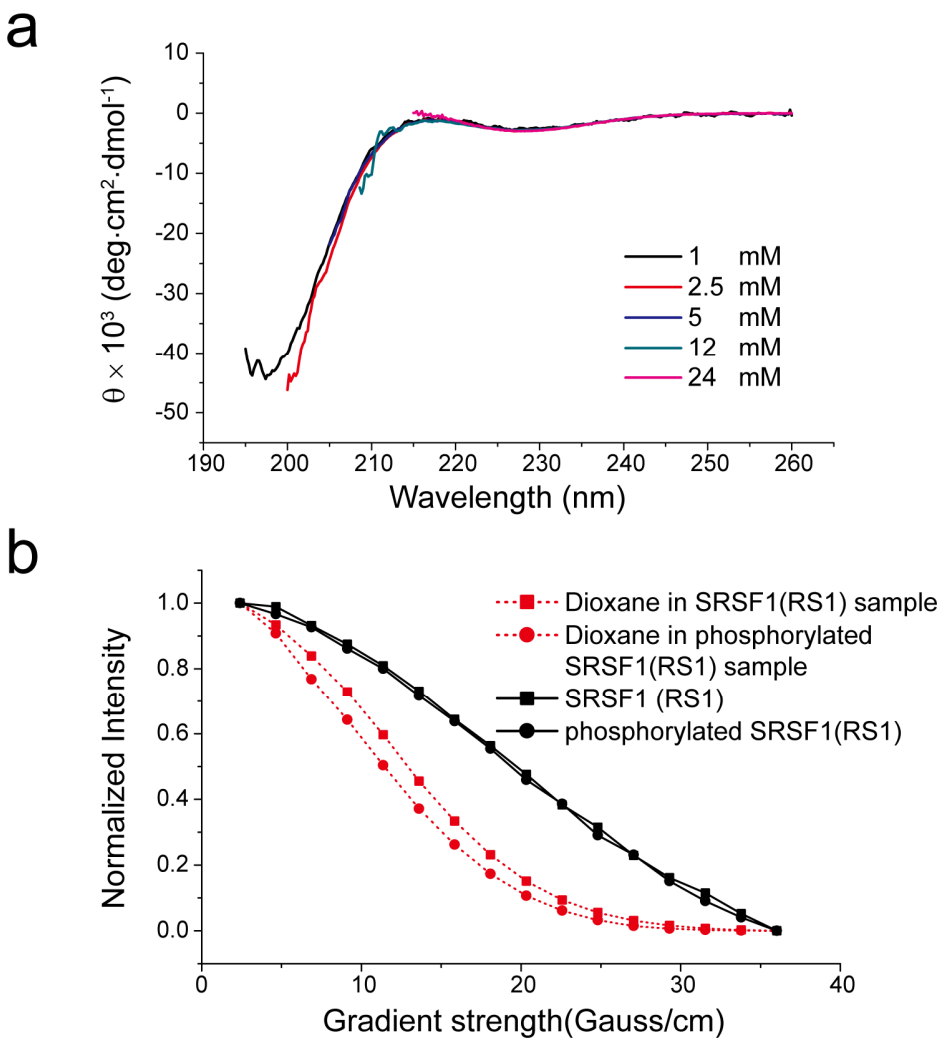

Figure S2, related to Figure 2. SRSF1(RS1) is predominantly monomeric. (a) Near UV CD spectra of phosphorylated SRSF1(RS1) at increasing peptide concentrations (50mM HEPES, pH 6.5, 300 mM NaCl). (b) NMR signal decay at increasing gradient strengths in NMR diffusion experiments of unphosphorylated and phosphorylated SRSF1(RS1) (solid line). Signal decays for the internal reference molecule of 1,4-dioxane are shown as dashed line.

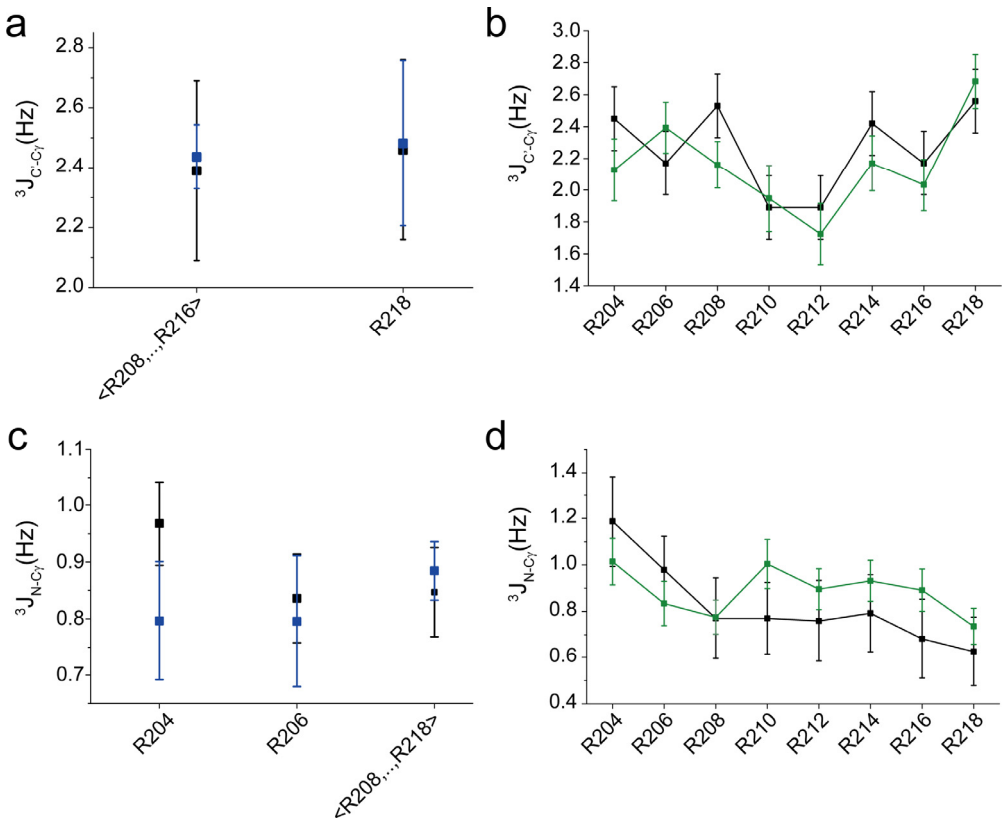

Figure S3, related to Figure 6. Comparison of experimental (black) 3-bond scalar couplings $^3J_{CO-C\gamma}$ (panels a and b) and $^3J_{N-C\gamma}$ (panels c and d) with values back-calculated from the conformer ensembles of unphosphorylated (a and c; blue) and phosphorylated (b and d; green) SRSF1(RS1). Note that $^3J_{CO-C\gamma}$ and $^3J_{N-C\gamma}$ couplings, which are sensitive to the $\chi 1$ angle, were not used for ensemble selection. Errors of experimental data were estimated based on NMR signal intensities and line widths. Errors of back-calculated couplings correspond to the standard deviations between the mean values of 100 independently selected ensembles.

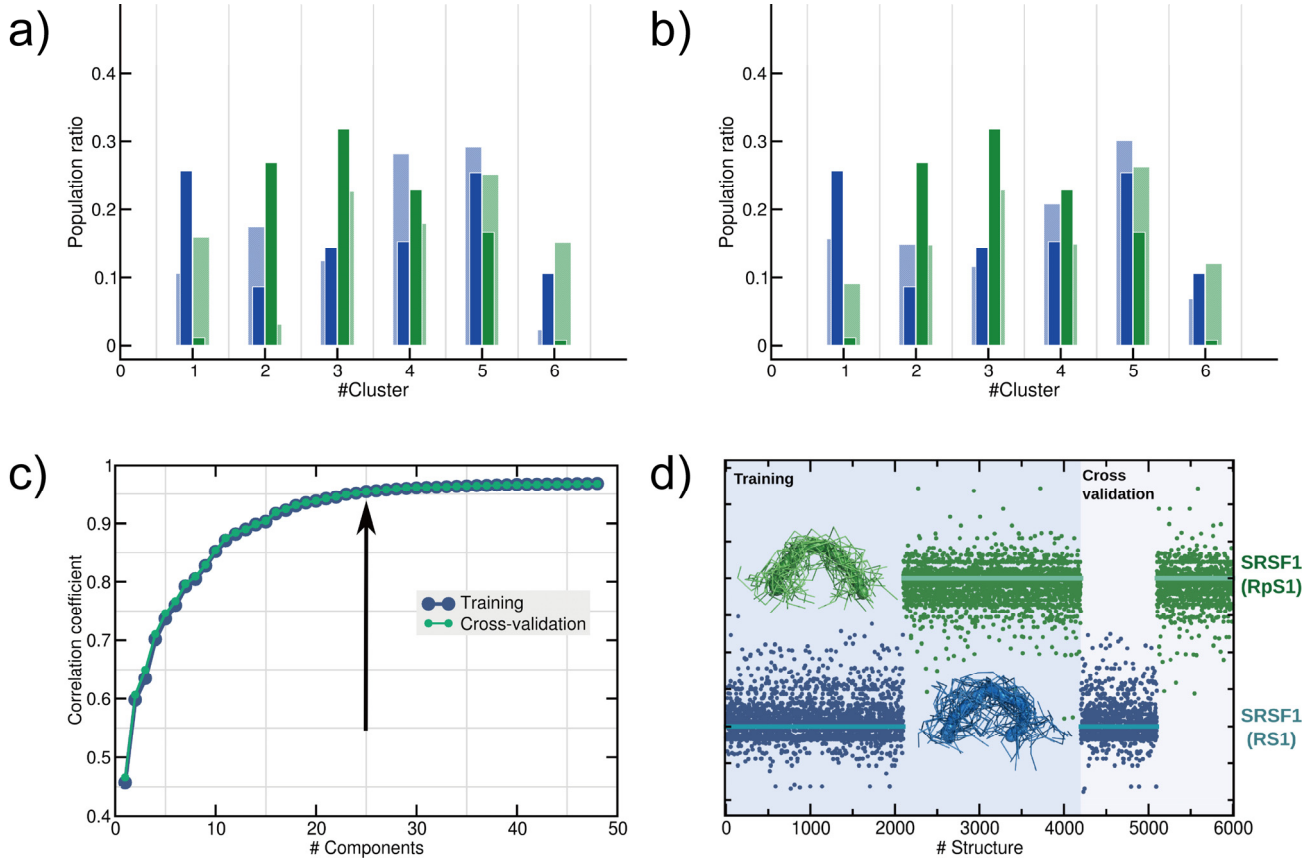

Figure S4, related to Figure 7. Cluster and partial least squares (PLS) analysis of the backbone conformation of unphosphorylated and phosphorylated SRSF1(RS1). **(a, b)** Cluster analysis of unphosphorylated and phosphorylated SRSF1(RS1) backbone conformations. 3000 structures of each type were used in the analysis. Principal component analysis identified six clusters in the plane of the first two eigenvectors. For each cluster the population ratio for phosphorylated vs unphosphorylated SRSF1(RS1) is shown. Compared to Figure 7b in the main text, structures were selected from either **(a)** an inverted pool (i.e. conformations were selected on the basis of experimental data of phosphorylated SRSF1(RS1) from the pool of structures generated by MD simulation for unphosphorylated SRSF1(RS1) and vice versa) or **(b)** a combined conformation pool, which contained structures generated by MD simulation of both phosphorylated and unphosphorylated SRSF1(RS1) from. The results of these selections are shown in light blue and green. The green as blue bars are the same as in Figure 7b. **(c)** PLS analysis of the backbone conformation of unphosphorylated and phosphorylated SRSF1(RS1). Correlation coefficient plotted against number of latent vectors. The arrow indicates the position 25, which is the number of latent vectors used in partial least squares regression. **(d)** PLS analysis: the left part with the dark background marks the training phase with 4200 structures (2100 conformers of SRSF1(RS1) and 2100 conformers of SRSF1(RpS1)). The right part with the light background shows the cross validation result with 1800 structures (half of SRSF1(RS1) and of SRSF1(RpS1)). Every point corresponds to one structure.

Table S1, related to Figure 2 | NMR parameters of phosphorylated SRSF1(RS1).

residue	D_{HN} (Hz)	D_{CH} (Hz)	$D_{C\alpha Co}$ (Hz)	CS Cα (ppm)	CS Co (ppm)	$^{1}J_{C\alpha-H\alpha}$ (Hz)	$^{1}J_{CaCb}$ (Hz)	$^{3}J_{NH-H\alpha}$ (Hz)	$^{3}J_{NC\gamma}$ (Hz)	$^{3}J_{CoC\gamma}$ (Hz)
G	-	-	-	-	-	-	-	-	-	-
A	-	-	-	-	-	144.84	-	-	-	-
M	-	-	-	55.23	176.42	141.61	34.54	7.61	-	-
G	-	-	-	44.6	-	140.82	-	-	-	-
P200	-	-	-	-	-	148.24	-	-	-	-
S201	3.24	8.24	-0.86	58.33	174.35	142.32	39.58	-	-	-
Y202	-2.82	2.12	0.16	58.22	176.33	143.44	35.12	7.1	-	-
G203	-1.8	3.39	0.31	45.25	173.79	141.05	-	-	-	-
R204	1.04	3.87	-0.21	55.9	176.36	142.27	34.71	7.22	1.19	2.45
S205	0.7	5.85	-0.69	58.35	174.7	142.38	38.84	6.68	-	-
R206	-0.18	-1.61	0.24	56.24	176.59	141.8	35.07	7.41	0.98	2.17
S207	-1.3	5.75	-0.61	58.35	174.38	143.76	38.91	6.18	-	-
R208	3.76	6.12	-0.35	55.72	176.14	142.27	34.75	6.97	0.77	2.53
S209	3.4	7.87	-1.25	58.2	174.19	143.47	37.9	5.77	-	-
R210	7.28	10.05	-1.71	55.73	176.13	143.14	35.34	6.61	0.77	1.89
S211	7.62	11.71	-1.5	58.18	174.16	143.41	39.87	5.58	-	-
R212	9.24	16.86	-1.71	55.76	176.13	140.87	33.62	7.09	0.76	1.89
S213	9.16	11.71	-1.01	58.18	174.2	143.56	38.97	5.58	-	-
R214	7.88	17.1	-1.33	55.79	176.3	142.41	34.32	6.83	0.79	2.42
S215	7.56	11.73	-1.19	58.18	174.3	143.61	39.25	5.51	-	-
R216	5.62	11.55	-1.5	55.91	176.17	142.15	34.27	6.67	0.68	2.17
S217	8.04	9.97	0.26	58.19	174.36	144.53	39.03	6.18	-	-
R218	4.38	7.07	-1.63	55.92	175.44	141.07	36.07	6.44	0.63	2.56
S219	2.22	7.98	-	59.93	-	141.81	38.51	7.54	-	-

Table S2, related to Figure 7 | Number of arginine residues involved in a hydrogen bond with a phosphoserine in the selected ensembles of phosphorylated SRSF1(RS1).

Number of arginines to which pSer has at least one hydrogen bond	Ratio with respect to all possible Arg-pSer hydrogen bonds (%)
0	60.88
1	28.83
2	9.88
3	0.53
4	0.01
5	0

Table S3, related to Figure 8 | Phosphorylation sites (marked in red) in the RS domain of hPrP28 as identified by mass spectrometry in native U5 snRNP and spliceosomal B complex.

U5 snRNP	B Complex	Begin aa position	End aa position	Sequence	Site
X	X	10	20	DRDA**S**PSKEER	S14
X		10	20	DRDASP**S**KEER	S16
X	X	23	30	**S**R**T**PDRER	S23, T25
	X	23	30	SR**T**PDRER	T25
X	X	37	44	KS**S**PSKDR	S38
X	X	37	44	KS**S**PSKDR	S39
X	X	37	44	KSSP**S**KDR	S41
	X	65	70	**S**K**S**AER	S65,S67
X		63	72	**S**R**S**K**S**AERER	S63, S65, S67
X	X	104	112	KR**SS**LSPGR	S106, S107
X	X	104	112	KRS**S**L**S**PGR	S106, S109
X	X	106	117	S**S**LSPGRGKDFK	S107
	X	105	112	RS**S**L**S**PGR	S107, S109
X	X	106	112	SSL**S**PGR	S109

Supplemental Experimental Procedures

Sample preparation. A phosphorylated $(RpS)_8$ peptide as well as a $(RE)_8$ peptide were prepared by solid-phase synthesis. SRSF1(RS) and SRSF1(RS1) were cloned into a pETZ2_1a plasmid(Hammarstrom et al., 2002) using NcoI and BamH1 restriction sites. Recombinant SRSF1(RS) and SRSF1(RS1) were produced in *E.coli* BL21 (DE3) induced by 1mM IPTG at 37 °C for 4 hours. Uniformly ^{13}C,^{15}N-labeled proteins were produced by growing cells in M9 minimal medium using ^{15}N-NH_4Cl, $^{13}C_6$-glucose as sole nitrogen and carbon sources. hPrp28 sequences were cloned into a modified pET-16 plasmid using NdeI and BamHI restriction sites. Recombinant hPrp28 was produced in the same way as SRSF1. Phosphorylated hPrp28 variants were obtained by co-expression with SRPK1.

SRSF1(RS), SRSF1(RS1) and hPrp28 variants were purified using a NTA-Ni^{2+} column followed by TEV cleavage for His-tag removal. Phosphorylation of various proteins by SRPK1 was preformed as described previously (Mathew et al., 2008). Phosphorylated proteins were purified by reverse phase HPLC and analyzed by mass spectrometry. SRSF1(RS1) species with different phosphorylation numbers could be well separated. Purified proteins were lyophilized and dissolved in 50 mM sodium phosphate buffer, pH 7.0, 100 mM NaCl. For pH titration, pH values were adjusted by addition of NaOH and HCl. For GdnHCl denaturing experiments, 6M GdnHCl was added to the samples.

NMR spectroscopy. NMR experiments were measured at 288 K and 298K on Bruker 900, 800, 700 and 600 MHz spectrometers equipped with cryoprobes, and a 600 MHz spectrometer equipped with a room temperature probe. Spectra (except of APSY) were processed using NMRPipe(Delaglio et al., 1995) and analyzed using Sparky(Goddard and Kneller) . 2D ^{15}N-^{1}H HSQC and 2D ^{13}C-^{1}H HSQC experiments were used to investigate the influence of pH on the structure of SRSF1(RS1).

APSY experiments, 5D CBCACONH (25 projection angles) and 6D HNCOCANH (23 projection angles) (Fiorito et al., 2006; Hiller et al., 2005) were used for backbone assignment of phosphorylated and unphosphorylated SRSF1(RS), as well as for phosphorylated hPrp28(1-138). Backbone assignments of phosphorylated and unphosphorylated SRSF1(RS1) were obtained using APSY 7D HNCOCACBNH experiments (33 projection angles) (Hiller et al., 2005; Hiller et al., 2007). APSY spectra were processed using PROSA (Güntert et al., 1992). Peaks on each projection spectrum were picked and the final peak list was calculated using GAPRO (Hiller et al., 2005). Assignment was performed in an iterative manner using MARS (Narayanan et al., 2010) and manual inspection.

$^{3}J_{HNH\alpha}$ scalar couplings were measured using intensity-modulated HSQCs with mixing times τ of 40 ms, 45 ms and 50 ms (Permi et al., 2000a). Hα magnetization was flipped by Q3 shaped pulse centered at 4 ppm and

covering a width of 2 ppm. Coupling values were calculated from the intensity ratios using the relation $S_{cross}/S_{diag} = \cos(\pi^3 J_{HNH\alpha} 2\tau)$. $^1J_{C\alpha\text{-}C\beta}$ couplings were measured using coupled 3D HNCA and 3D HNCOCA experiments executed in a 2D manner (sweep width in ppm:10(H)/10(C), carrier in ppm: 4.7(H)/117(N)/55(C), number of data points: 1024(H)*1(N)*256(C)). $^3J_{N\text{-}C\gamma}$ and $^3J_{CO\text{-}C\gamma}$ coupling constants of arginine side chains of SRSF1(RS1) were measured using two-dimensional spin-echo difference experiments(Hu et al., 1997) with mixing times of 100 ms for N-Cγ and 90 ms for Co-Cγ, respectively. In N-Cγ coupling measurements, Cγ was flipped by a Q3 shaped pulse centered at 25 ppm and covering 24 ppm. In CO-Cγ coupling measurements, Cγ was flipped by a 500 μs sinc pulse.

Hα-Cα couplings were measured using 3D (HA)CANH experiments manner (sweep widths in ppm: 11(H)/16(N)/80(C), carrier in ppm: 4.7(H)/117(N)/53(C), number of data points: 1024(H)×96(N)×96(C)) (Zweckstetter and Bax, 2001), Cα-Co couplings by spin-state selective HNCO experiments carried out in a 2D manner (sweep widths in ppm: 11(H)/8(C), carrier in ppm: 4.7(H)/117(N)/174(C), number of data points: 1024(H)×1(N)×256(C)) (Permi et al., 2000b), and HN-N splittings using 2D TROSY-HSQC interleaved experiments (sweep widths in ppm: 11(H)/16(N), carrier in ppm: 4.7(H)/117(N), number of data points: 1024(H)×352(N)) (Kontaxis et al., 2000). Residual dipolar couplings were calculated as the difference between splittings measured in the isotropic phase and in a sample, in which SRSF1(RS1) had been aligned in 5% C8E5/n-octanol liquid crystalline(Ruckert and Otting, 2000). In addition, a sample was prepared that contained both SRSF1(RS1) and protein G in C8E5/n-octanol. The HN-N dipolar couplings observed for protein G and SRSF1(RS1) in this sample were then used to calibrate the effective concentration of the liquid crystal in prediction of RDCs using the PALES steric alignment mode (Zweckstetter and Bax, 2000). This optimization resulted in a liquid crystal concentration of 0.093 mg/ml.

Steady-state heteronuclear $^{15}N\text{-}^1H$-NOE were calculated as the intensity ratios of the $^{15}N\text{-}^1H$ correlation peaks from pairs of interleaved 2D $^{15}N\text{-}^1H$ correlation spectra acquired at 298K with and without ^1H presaturation during the recycle time of 7.5 s(Ferrage et al., 2010).

Mass spectrometry of in vivo phosphorylation of hPrp28. U5 snRNPs were purified from HeLa nuclear extract essentially as described by Bach et al.(Bach et al., 1990),and human spliceosomal B complex was assembled and purified as described earlier(Bessonov et al., 2008). Proteins in the purified spliceosomal (sub-) complexes were digested with endoproteinase trypsin in solution and after separation of the proteins by 1D PAGE. Peptides were enriched for phosphopeptides by TiO_2 and analyzed by liquid chromatography (LC) coupled ESI-MSMS on an Orbitrap XL (U5 snRNPs) and an Orbitrap Velos (spliceosomal B complex) mass spectrometer (ThermoFisherScientific) as described earlier(Schneider et al., 2010). Phosphopeptides including its

phosphorylation site were identified by database search against NCBInr database using MASCOT as search engine(Schneider et al., 2010).

Molecular dynamics simulations. The topology and force field parameters for phosphoserine were created according to Homeyer et al.(Sousa da Silva and Vranken, 2012). For production runs the Amber99sb* force field(Best and Hummer, 2009; Hornak et al., 2006) was employed. The topology of phosphoserines was modified by including additional backbone dihedral angles in accordance to the topology of a serine residue to adapt it to the Amber99sb* force field.

Randomized starting structures for MD simulations were generated employing the tConcoord(Seeliger et al., 2007) algorithm. 20 independent runs were set up by placing a peptide in a dodecahedron box and solvating it with the SPC/E(Berendsen et al., 1987) water. Sodium and chloride ions were added to neutralize the system and reach a salt concentration of 150 mM. The system was coupled to an external heat bath using the velocity rescaling thermostat(Bussi et al., 2007) with a time constant of 0.1 ps at a temperature of 298 K. The Parrinello-Rahman barostat(Parrinello and Rahman, 1981) with a time constant of 5 ps was used to retain a pressure of 1 bar.

All peptide bonds were constrained with the LINCS(Hess et al., 1997) algorithm of order 6. The fastest degrees of freedom in the simulation were removed by replacing hydrogens with virtual sites(Feenstra et al., 1999). This allowed an integration timestep of 4 fs. Bond lengths and bond angles of water molecules were constrained using the SETTLE(Miyamoto and Kollman, 1992) algorithm. Electrostatics were treated using the particle-mesh Ewald (PME)(Darden et al., 1993; Essmann et al., 1995) method with a real space cutoff of 1.2 nm, a Fourier grid spacing of 0.14 nm and a PME order of 4. Van der Waals interactions were smoothly switched to zero in the interval from 1 to 1.1 nm. Prior to starting the MD simulations a steepest descent energy minimization was performed. Every MD run started with 250 ps of simulated annealing during which system was heated up from 0 to 298 K. The duration of each independent MD simulation was set to 52 ns, leaving the first 2 ns for equilibration and using 50 ns for subsequent analysis. In total, 1 μs of simulation time was reached for phosphorylated and unphosphorylated SRSF1(RS1).

Ensemble selection and comparison. To derive ensembles that best match the experimental NMR data, structural ensembles generated by MD were subjected to a sub-ensemble selection procedure. The developed ensemble selection algorithm consists of two main parts: Monte-Carlo based simulated annealing(Kirkpatrick et al., 1983) optimization and a brute force search. The objective function that the method aims to minimize was constructed as an RMSD between the mean value over a selected ensemble and an experimental observation. Prior to performing the selection, variables were normalized (z-scored) to avoid bias towards any particular data set. In case of unphosphorylated SRSF1(RS1), averaged values over several amino acids were used for

selection, when overlapping peaks in the NMR spectra could not be assigned to individual residues. For the selection procedure of the phosphorylated SRSF1(RpS1), structures of both simulated protonation states (PO_4^- and PO_4^{2-}) were pooled together.

The algorithm is initialized by a random ensemble of a defined size from the pool of MD structures. During a Monte-Carlo step a random structure from the sub-ensemble was replaced with another randomly picked structure from the MD pool. An update was accepted in case the agreement with the experimental data set increased. Otherwise, the threshold acceptance method(Dueck and Scheuer, 1990) was applied by directly comparing the objective energy function to the temperature. Multiple heuristic switches were included to avoid getting trapped in local minima, for example, raising the temperature, setting the current ensemble to the best intermediate result, forgetting the best intermediate result which allowed navigating further from the already explored regions. Every 100 million steps the simulated annealing mode was switched off and a brute force search was performed: replacement of every structure of the selected sub-ensemble was attempted with every structure from an MD pool. Only single structure replacements were performed. The brute force search was iterated until no replacement was accepted in one full iteration over the whole sub-ensemble, subsequently, the algorithm switched back into the simulated annealing mode. The overall procedure was repeated for a fixed number of steps ($5x10^9$) after which the best solution found was written out.

Ensemble comparison. To avoid artifacts related to superposition of flexible disordered structures, internal distances between C atoms were used. For dimensionality reduction and clustering half of the internal distance matrix (discarding the diagonal) was calculated for each structure of an ensemble. Principal component analysis was performed by calculation and subsequent diagonalization of a covariance matrix of internal distances. Comparison of the SRSF1(RS1) and SRSF1(RpS1) ensembles was performed by adapting an ensemble similarity score based on dimensionality reduction(Lindorff-Larsen and Ferkinghoff-Borg, 2009). Structures of the respective ensembles were projected onto the two principal components with the largest eigenvalues. Distributions of the projections were smoothed by applying the kernel density estimation algorithm(Duong, 2007) as implemented in the statistical software package R(R Development Core Team, 2010). Jensen-Shannon divergence (Lin, 1991) was used as a metric to compare the smoothed distributions.

Conformational entropies of the selected sub-ensembles were estimated using Schlitter's formula(Schlitter, 1993). Changes in the backbone conformation upon phosphorylation were analysed by means of partial least squares-based functional mode analysis(Krivobokova et al., 2012). Ensemble weighted maximally correlated motion(Hub and de Groot, 2009) was used to capture the difference between SRSF1(RS1) and SRSF1(RpS1) ensembles. Side chain densities were calculated with the VMD(Humphrey et al., 1996) tool VolMap.

Conformer ensembles were superimposed by minimizing the variance and RMSDs between the molecules in their local neigbourhood(Gapsys and de Groot, 2013). Procedures previously described for clustering structural ensembles were used(Matthes et al., 2011; Matthes et al., 2012). Clustering was based on the Hartigan-Wong k-means algorithm(Hartigan and Wong, 1979), for which the cluster centers were determined using the global k-means method(Likas et al., 2003), and the Krzanowski-Lai criterion(Krzanowski and Lai, 1988) was used to determine the optimal number of clusters.

Validation of structural ensembles. To validate the structural ensembles of unphosphorylated and phosphorylated SRSF1(RS1), NMR observables that were not included in the ensemble selection were back-calculated from the selected ensembles and compared to experimental values. The Karplus equation parameters provided in(Vuister and Bax, 1993) were used for $^{3}J_{HNH\alpha}$ coupling calculations. Parameters for back-calculation of $^{3}J_{NC\gamma}$ and $^{3}J_{CoC\gamma}$ couplings were taken from(Perez et al., 2001). For $^{1}J_{C\alpha\text{-}H\alpha}$ coupling calculations the Karplus equation was newly parameterized due to an observed discrepancy between values reported in the literature(Vuister et al., 1993; Zweckstetter and Bax, 2001). Details of the parameterization will be published elsewhere. $^{1}J_{C\alpha\text{-}H\alpha}$ couplings back-calculated from phosphorylated SRSF1(RS1) ensembles were in close agreement with experimental measurements with a RMSD of 0.86 Hz. The unphosphorylated SRSF1(RS1) ensemble also showed a good match with a RMSD of 0.99 Hz from the measured data. $^{1}J_{C\alpha\text{-}C\beta}$ couplings were back-calculated according to equation (1) in(Cornilescu et al., 2000) and were in agreement with experiment with RMSDs of 1.29 Hz and 1.21 Hz for SRSF1(RS1) and SRSF1(RpS1), respectively. A larger RMSD in case of $^{1}J_{C\alpha\text{-}C\beta}$ may be due to an offset in the Karplus equation of $^{1}J_{C\alpha\text{-}C\beta}$, which had been parameterized using $^{1}J_{C\alpha\text{-}C\beta}$ couplings of free amino acids(Cornilescu et al., 2000). In the above calculations, random coil values of serine were used for phosphoserine residues, due to a lack of random coil values for phosphoserines.

For the back-calculation of chemical shifts two different programs were used. Chemical shifts predicted from the conformational ensembles by both SPARTA+(Shen and Bax, 2010) and SHIFTX2(Han et al., 2011), for which the sequence-based prediction component was switched off, were in good agreement with experimental values. Predicted Cα and CO chemical shifts showed high Pearson correlation coefficients (>0.9) for both unphosphorylated and phosphorylated SRSF1(RS1). Chemical shift prediction for phosphoserines is complicated since the empirical predictors are not trained against phosphorylated amino acids, potentially leading to the systematic offset observed for CO chemical shifts of phosphoserines in SHIFTX2 predictions. In addition, due to the high density of phosphorylated serine residues in SRSF1(RS1), the chemical shift prediction of neighbouring arginine residues might also be influenced. Despite these difficulties, however, the agreement between experimental NMR observables and those back-calculated from the ensembles of SRSF1(RS1) was favorable.

Supplemental References

Bach, M., Bringmann, P., and Luhrmann, R. (1990). Purification of small nuclear ribonucleoprotein particles with antibodies against modified nucleosides of small nuclear RNAs. Methods Enzymol *181*, 232-257.

Berendsen, H.J.C., Grigera, J.R., and Straatsma, T.P. (1987). The Missing Term in Effective Pair Potentials. J Phys Chem *91*, 6269-6271.

Bessonov, S., Anokhina, M., Will, C.L., Urlaub, H., and Luhrmann, R. (2008). Isolation of an active step I spliceosome and composition of its RNP core. Nature *452*, 846-850.

Best, R.B., and Hummer, G. (2009). Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. J Phys Chem B *113*, 9004-9015.

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. J Chem Phys *126*.

Cornilescu, G., Bax, A., and Case, D.A. (2000). Large variations in one-bond (C alpha-13C beta)-C-13 J couplings in polypeptides correlate with backbone conformation. J Am Chem Soc *122*, 2168-2171.

Darden, T., York, D., and Pedersen, L. (1993). Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. J Chem Phys *98*, 10089-10092.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. J Biomol NMR *6*, 277-293.

Dueck, G., and Scheuer, T. (1990). Threshold Accepting - a General-Purpose Optimization Algorithm Appearing Superior to Simulated Annealing. JCoPh *90*, 161-175.

Duong, T. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in R. Journal of Statistical Software *21*, 1-16.

Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A Smooth Particle Mesh Ewald Method. J Chem Phys *103*, 8577-8593.

Feenstra, K.A., Hess, B., and Berendsen, H.J.C. (1999). Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem *20*, 786-798.

Ferrage, F., Reichel, A., Battacharya, S., Cowburn, D., and Ghose, R. (2010). On the measurement of 15N–{1H} nuclear Overhauser effects. 2. Effects of the saturation scheme and water signal suppression. J Magn Reson *207*, 294-303.

Fiorito, F., Hiller, S., Wider, G., and Wuthrich, K. (2006). Automated resonance assignment of proteins: 6D APSY-NMR. J Biomol NMR *35*, 27-37.

Gapsys, V., and de Groot, B.L. (2013). Optimal Superpositioning of Flexible Molecule Ensembles. Biophys J *104*, 196-207.

Goddard, T.D., and Kneller, D.G. SPARKY 3 (University of California, San Francisco).

Güntert, P., Dötsch, V., Wider, G., and Wüthrich, K. (1992). Processing of multi-dimensional NMR data with the new software PROSA. J Biomol NMR *2*, 619-629.

Hammarstrom, M., Hellgren, N., van Den Berg, S., Berglund, H., and Hard, T. (2002). Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci *11*, 313-321.

Han, B., Liu, Y., Ginzinger, S.W., and Wishart, D.S. (2011). SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR *50*, 43-57.

Hartigan, J.A., and Wong, M.A. (1979). Algorithm AS 136: A K-Means Clustering Algorithm. Journal of the Royal Statistical Society Series C (Applied Statistics) *28*, 100-108.

Hess, B., Bekker, H., Berendsen, H.J.C., and Fraaije, J.G.E.M. (1997). LINCS: A linear constraint solver for molecular simulations. J Comput Chem *18*, 1463-1472.

Hiller, S., Fiorito, F., Wuthrich, K., and Wider, G. (2005). Automated projection spectroscopy (APSY). Proc Natl Acad Sci U S A *102*, 10876-10881.

Hiller, S., Wasmer, C., Wider, G., and Muthrich, K. (2007). Sequence-specific resonance assignment of soluble nonglobular proteins by 7D APSY-NMR Spectroscopy. J Am Chem Soc *129*, 10823-10828.

Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., and Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. Proteins *65*, 712-725.

Hu, J.S., Grzesiek, S., and Bax, A. (1997). Two-dimensional NMR methods for determining (chi 1) angles of aromatic residues in proteins from three-bond J(C'C gamma) and J(NC gamma) couplings. J Am Chem Soc *119*, 1803-1804.

Hub, J.S., and de Groot, B.L. (2009). Detection of Functional Modes in Protein Dynamics. PLoS Comp Biol *5*.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. J Mol Graphics Model *14*, 33-38.

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P. (1983). Optimization by Simulated Annealing. Science *220*, 671-680.

Kontaxis, G., Clore, G.M., and Bax, A. (2000). Evaluation of cross-correlation effects and measurement of one-bond couplings in proteins with short transverse relaxation times. J Magn Reson *143*, 184-196.

Krivobokova, T., Briones, R., Hub, J.S., Munk, A., and de Groot, B.L. (2012). Partial Least-Squares Functional Mode Analysis: Application to the Membrane Proteins AQP1, Aqy1, and CLC-ec1. Biophys J *103*, 786-796.

Krzanowski, W.J., and Lai, Y.T. (1988). A Criterion for Determining the Number of Groups in a Data Set Using Sum-of-Squares Clustering. Biometrics *44*, 23-34.

Likas, A., Vlassis, N., and J. Verbeek, J. (2003). The global k-means clustering algorithm. Pattern Recogn *36*, 451-461.

Lin, J.H. (1991). Divergence Measures Based on the Shannon Entropy. ITIT *37*, 145-151.

Lindorff-Larsen, K., and Ferkinghoff-Borg, J. (2009). Similarity Measures for Protein Ensembles. Plos One *4*.

Mathew, R., Hartmuth, K., Mohlmann, S., Urlaub, H., Ficner, R., and Luhrmann, R. (2008). Phosphorylation of human PRP28 by SRPK2 is required for integration of the U4/U6-U5 tri-snRNP into the spliceosome. Nat Struct Mol Biol *15*, 435-443.

Matthes, D., Gapsys, V., Daebel, V., and de Groot, B.L. (2011). Mapping the Conformational Dynamics and Pathways of Spontaneous Steric Zipper Peptide Oligomerization. Plos One *6*.

Matthes, D., Gapsys, V., and de Groot, B.L. (2012). Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations. J Mol Biol *421*, 390-416.

Miyamoto, S., and Kollman, P.A. (1992). Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. J Comput Chem *13*, 952-962.

Narayanan, R.L., Durr, U.H.N., Bibow, S., Biernat, J., Mandelkow, E., and Zweckstetter, M. (2010). Automatic Assignment of the Intrinsically Disordered Protein Tau with 441-Residues. J Am Chem Soc *132*, 11906-11907.

Parrinello, M., and Rahman, A. (1981). Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. J Appl Phys *52*, 7182-7190.

Perez, C., Lohr, F., Ruterjans, H., and Schmidt, J.M. (2001). Self-consistent Karplus parametrization of (3)J couplings depending on the polypeptide side-chain torsion chi(1). J Am Chem Soc *123*, 7081-7093.

Permi, P., Kilpelainen, I., Annila, A., and Heikkinen, S. (2000a). Intensity modulated HSQC and HMQC: Two simple methods to measure (3)J(HNH alpha) in proteins. J Biomol NMR *16*, 29-37.

Permi, P., Rosevear, P.R., and Annila, A. (2000b). A set of HNCO-based experiments for measurement of residual dipolar couplings in N-15, C-13, (H-2)-labeled proteins. J Biomol NMR *17*, 43-54.

R Development Core Team (2010). R: A Language and Environment for Statistical Computing.

Ruckert, M., and Otting, G. (2000). Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc *122*, 7793-7797.

Schlitter, J. (1993). Estimation of Absolute and Relative Entropies of Macromolecules Using the Covariance-Matrix. Chem Phys Lett *215*, 617-621.

Schneider, M., Hsiao, H.H., Will, C.L., Giet, R., Urlaub, H., and Luhrmann, R. (2010). Human PRP4 kinase is required for stable tri-snRNP association during spliceosomal B complex formation. Nat Struct Mol Biol *17*, 216-221.

Seeliger, D., Haas, J., and de Groot, B.L. (2007). Geometry-based sampling of conformational transitions in proteins. Structure *15*, 1482-1492.

Shen, Y., and Bax, A. (2010). SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR *48*, 13-22.

Sousa da Silva, A., and Vranken, W. (2012). ACPYPE - AnteChamber PYthon Parser interfacE. BMC Research Notes *5*, 367.

Vuister, G.W., and Bax, A. (1993). Quantitative J Correlation - a New Approach for Measuring Homonuclear 3-Bond J(H(N)H(Alpha) Coupling-Constants in N-15-Enriched Proteins. J Am Chem Soc *115*, 7772-7777.

Vuister, G.W., Delaglio, F., and Bax, A. (1993). The Use of 1jc-Alpha-H-Alpha Coupling-Constants as a Probe for Protein Backbone Conformation. J Biomol NMR *3*, 67-80.

Zweckstetter, M., and Bax, A. (2000). Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. J Am Chem Soc *122*, 3791-3792.

Zweckstetter, M., and Bax, A. (2001). Single-Step Determination of Protein Substructures Using Dipolar Couplings: Aid to Structural Genomics. J Am Chem Soc *123*, 9490-9491.