Activities of Topoisomerase I in Its Complex with SRSF1

Takao Ishikawa,[†] Krystiana A. Krzysko,[‡] Barbara Kowalska-Loth,[†] Aleksandra M. Skrajna,[†] Alicja Czubaty,[†] Agnieszka Girstun,[†] Maja K. Cieplak,[†] Bogdan Lesyng,[‡,§] and Krzysztof Staron*,[†]

[†]Institute of Biochemistry, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland
[‡]Department of Biophysics, Faculty of Physics, University of Warsaw, Zwirki i Wigury 93, 02-089 Warsaw, Poland
[§]Mossakowski Medical Research Centre, Polish Academy of Sciences, Pawinskiego 5, 02-106 Warsaw, Poland

S Supporting Information

ABSTRACT: Human DNA topoisomerase I (topo I) catalyzes DNA relaxation and phosphorylates SRSF1. Whereas the structure of topo I complexed with DNA has been resolved, the structure of topo I in the complex with SRSF1 and structural determinants of topo I activities in this complex are not known. The main obstacle to resolving the structure is a contribution of unfolded domains of topo I and SRSF1 in formation of the complex. To overcome this difficulty, we employed a three-step strategy: identifying the interaction regions, modeling the complex, and validating the model with biochemical methods. The binding sites in both topo I and SRSF1 are localized in the structured regions as well as in the unfolded domains. One observes cooperation between the binding sites in topo I but not in SRSF1. Our results indicate two features of the unfolded RS domain of SRSF1 containing phosphorylated residues that are critical for the kinase activity of topo I: its spatial arrangement relative to topo I and the organization of its sequence. The efficiency of phosphorylation of SRSF1 depends on the length and flexibility of the spacer between the two RRM domains that uniquely determine an arrangement of the RS domain relative to topo I. The spacer also influences inhibition of DNA nicking, a prerequisite for DNA relaxation. To be phosphorylated, the RS domain has to include a short sequence recognized by topo I. A lack of this sequence in the mutants of SRSF1 or its spatial inaccessibility in SRSF9 makes them inadequate as topo I/kinase substrates.

The structure and a putative motif recognized by topoisomerase I during phosphorylation of SRSF1
R--------RSRS----R
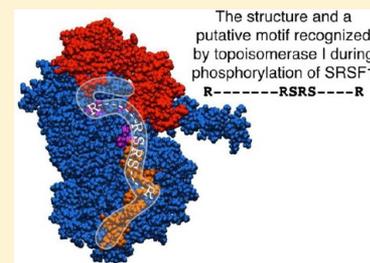
Human topoisomerase I (topo I) is a dual-activity enzyme. First, it transiently nicks double-stranded DNA. Because of this activity, topo I is a main nuclear swivelase responsible for relieving a torsional stress that appears in DNA during transcription, replication, and chromatin condensation.[1] The activity is also used to remove ribonucleotides incorporated during DNA replication and not deleted by the repair system.[2,3] Next, topo I is a kinase that phosphorylates SR proteins,[4] essential splicing factors, and regulators of splicing[5] composed of one or two RNA recognition motifs (RRMs) followed by the RS domain (RS) and containing numerous serine-arginine repeats. Because of the latter activity, topo I has been observed to influence alternative splicing of several transcripts.[6] Both DNA relaxation and phosphorylation activities of topo I have been also shown to cooperate in preventing a conflict between transcription and DNA replication.[7,8] However, both activities do not work at the same time: DNA, a substrate for DNA relaxation, inhibits the kinase reaction,[4] whereas both a protein substrate for the kinase activity[9] and ATP[10] inhibit DNA nicking.

Topo I is a single polypeptide of 765 amino acid residues composed of four domains: N-terminal domain (NT, residues 1−214), core domain (residues 215−635), linker domain (residues 636−712), and C-terminal domain (CT, residues 713−765).[11] The core domain is further divided into three subdomains that form two distinct lobes in the three-dimensional structure of topo I: the cap, including subdomains

I and II (residues 215−433), and subdomain III (residues 434−635). The molecular structure of topo I complexed with DNA has been resolved,[11] and the mechanism of DNA relaxation catalyzed by topo I is fully understood.[12] On the other hand, the structure of topo I in its complex with a protein substrate for kinase is not known. Consequently, structural determinants of the kinase activity and a molecular background of inhibition of DNA relaxation by the kinase substrate remain obscure.

Only few details about the binding of the SR proteins to topo I and its kinase activity are known. The prototypical protein substrate SRSF1 (according to the recently proposed nomenclature;[13] previously named ASF/SF2) contains two RRMs separated by a long spacer (to distinguish between fragments that separate either RRMs in SRSF1 or the core and the CT in topo I, the first one is here called the "spacer" and the other the "linker"). The structure of complete SRSF1 has not been presented except for the RRM2 remaining in the complex with the SR protein kinase SRPK1.[14] SRSF1 binds to topo I at least through two sites. The first one interacts with the cap in topo I and is built by residues of RRM1.[9,15] The other one binds to the NT in topo I[9,16] and is believed to be localized in RS.[16] Because both NT[17] and RS[18] are natively unfolded, topo I is a unique kinase that uses an unfolded site to bind an

unfolded protein substrate. With regard to the ATP binding site, it has been roughly localized in the C-terminal part of topo I.[19] Because of the lack of a classic ATP binding motif, the ATP binding site cannot be simply predicted from the amino acid sequence of topo I.

Topo I is one of several protein kinases that phosphorylate SR proteins. Main SR protein kinases compose two kinase families: SRPK and Clk/Sty.[20−22] The model enzymes of both families, SRPK1 and CLK1, respectively, have different substrate specificities that come from clearly established structural differences between both kinases. SRPK1 efficiently phosphorylates serine residues localized in the N-terminal fragment of RS in SRSF1,[23,24] while CLK1 phosphorylates all serine residues in RS[25] or serine residues in the C-terminal portion of RS in SRSF1 previously phosphorylated by SRPK1.[26] The background for different specificities of both enzymes is a docking groove that binds the N-terminal portion of RS[27] and ensures ordered and directional phosphorylation in SRPK1[28] but is inaccessible in CLK1.[26] In contrast to the the preceding discussion, no structural determinants of the efficiency of phosphorylation catalyzed by topo I are known.

DNA nicking catalyzed by topo I that accompanies transcription poses a threat to genome stability.[29] It has been proposed that inhibition of DNA nicking activity by SR proteins[9] substantially reduces this threat.[30] No details of this process are known except for the observation that inhibition of DNA nicking by SRSF1 does not require RS.[31]

Taken together, of two complexes formed by topo I, only that with DNA has been structurally and functionally characterized. The general purpose of this work was to gain a picture of topo I operating in the complex with SRSF1. The specific goals were to model a structure of the complex, to identify structural determinants of the kinase activity, and to understand how SRSF1 inhibits DNA relaxation. The main obstacle to achieving these goals was a contribution of unfolded domains of topo I and SR proteins in the formation of the complex and their participation in the enzymatic process catalyzed by topo I. To overcome this difficulty, we employed a three-step strategy of listing the interaction regions, modeling the complex, and biochemically testing its properties.

■ MATERIALS AND METHODS

Plasmids and Protein Purification. Topo I was expressed in *Saccharomyces cerevisiae* and purified as described previously[9] with minor modifications. After initial purification by nickel-nitrilotriacetic acid-agarose chromatography, the protein extract was loaded on a fast performance liquid chromatography system equipped with a Mono S column cation exchanger (Bio-Rad) equilibrated with LS buffer [25 mM Tris-HCl (pH 8.0), 0.17 M KCl, 3 mM MgCl$_2$, 10 mM β-mercaptoethanol, and 10% glycerol]. Topo I was eluted with a linear KCl gradient. Expression and purification of His-tagged topo I fragments (T[1−214] and T[215−433]) have been described previously.[31] The coding sequence of topo I fragment T[1−433] was generated by polymerase chain reaction (PCR) from the full-length cDNA of topo I and cloned into the pQE30 vector (Qiagen). Expression and purification of His-tagged T[1−433] were conducted according to the procedure for T[1−214].[31] Full-length SRSF1 was amplified by PCR from the pTrc-His vector containing cDNA of human SRSF1[9] and cloned into the pET28a vector (Novagen). To generate SRSF1[Δ], inverse PCR was used. Similarly, SRSF1[3G] and SRSF1[A1] were generated by inverse PCR using primers containing overhangs

encoding peptide spacers being introduced. In the SRSF1-[S209AS215A] mutant, the S209A mutation was introduced by inverse PCR (SRSF1[S209A]), while the S215A mutation was generated by site-directed mutagenesis using SRSF1[S209A] as a template. SRSF1[3Δ] and SRSF1[3ΔS209A] were constructed by inverse PCR using wild-type SRSF1 and SRSF1-[S209A] as templates, respectively. All SRSF1 mutants (in the spacer or RS domain) were expressed as His-tagged proteins in *Escherichia coli* strain BL21 and purified exactly as described in ref 4. GST-tagged full-length SRSF1[1−248] and fragments of SRSF1[1−119], SRSF1[120−194], and SRSF1[195−248] were described previously.[31] Coding sequences of other GST-tagged SRSF1 fragments (SRSF1[1−88] and SRSF1[89−119]) and the SRSF9[185−221] fragment were amplified by PCR and cloned into appropriate pGEX vectors (Amersham). GST-tagged SRSF1[2N] and SRSF1[2C] were generated by inverse PCR using GST-tagged full-length SRSF1[1−248] as a template.[31] To generate His-tagged SRSF1[2N] and SRSF1-[2C], GST-tagged constructs were digested with BamHI and NotI and cloned into the pET-28a vector. Expression and purification of GST, GST-tagged SRSF1 fragments, and SRSF9[185−221] were performed as described previously.[31] GST-tagged SRSF1[195−248] or GST-tagged SRSF9[185−221] bound to glutathione agarose beads was digested with thrombin to yield SRSF1[195−248] or SRSF9[185−221], respectively, deprived of GST. The coding sequences of SRSF5 and SRSF9 were amplified by PCR from cDNA generated with total RNA isolated from HeLa cells and cloned into the pQE30 vector. SRSF5 and SRSF9 were expressed as His-tagged proteins in *E. coli* strain M15 and purified according to the same procedure as His-tagged SRSF1 variants and mutants. To obtain GST-tagged CLK1, SRSF5 and SRSF9 coding sequences were amplified by PCR and cloned into the pGEX-4T1 or pGEX-4T2 vector. Overexpression and purification of CLK1, SRSF5, and SRSF9 were conducted following the procedure for GST-tagged full-length SRSF1 and its fragments as described previously.[31] Cloning, expression, and purification of UP1 have been described previously.[15] UP1[SRSF1] was generated by inverse PCR conducted using primers with overhangs encoding the spacer of SRSF1. It was purified using the same method that was used for unmodified UP1.

Primers used in the cloning and mutagenesis are listed in Table S1 of the Supporting Information.

Pull-Down Assay. The pull-down assay was performed as described previously with some modifications.[31] Briefly, GST-tagged SRSF1, its fragments, SRSF5, SRSF9, or GST was mixed with His-tagged topo I fragments and rotated for 2 h at room temperature with glutathione-agarose beads. The beads were recovered by centrifugation and washed three times. The bound proteins were eluted by being boiled in Laemmli buffer, separated by sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE), and detected by Western blot analysis. His-tagged proteins were identified with anti-His monoclonal antibodies (Sigma).

Enzyme Activity Assays. Topo I activity and CLK1 kinase activity were assayed in a final volume of 15 μL containing 50 mM Tris-HCl (pH 8.0), 0.1 mM EDTA, 1 mM DTT, 0.01% Triton X-100, 20 μM CPT, 0.2 mg/mL bovine serum albumin, and 2% glycerol. In salt dependence tests, variable amounts of NaCl were added to the reaction mixture. [γ-^{32}P]ATP (1.7 nmol, 1.1 × 10^5 Bq) and 1.2 pmol of His-tagged substrate protein (SRSF1, its variants and mutants, SRSF5, or SRSF9) were used in each assay. The only exceptions were

SRSF1[195−248] and SRSF9[185−221] that were used as GST-tagged proteins or were deprived of any tag. The reaction was started by addition of 0.2 pmol of topo I or CLK1. After incubation at 37 °C for 20 min, the reaction was stopped by addition of Laemmli buffer. Substrate phosphorylation was visualized by autoradiography, after electrophoretic separation on 12, 15, or 20% polyacrylamide gels (SDS−PAGE). For competition assays using His-tagged fragments of topo I or GST-tagged fragments of SRSF1, appropriate proteins were added to the reaction mixtures in an amount that was 2 times the amount of the primary substrate protein. Topo I DNA cleavage activity was assayed using a ^{32}P-5′-end-labeled double-stranded oligonucleotide containing a single topo I cleavage site (suicide substrate), as previously reported.[32] Reaction mixtures (15 μL) were assembled on ice in the same buffer as in the kinase activity tests. Each reaction mixture contained the same amount of ^{32}P-5′-end-labeled double-stranded oligonucleotide (0.1 pmol, 3.4×10^6 cpm/pmol). [γ-^{32}P]ATP (1.7 nmol, 1.1×10^5 Bq) was added to the reaction mixture when kinase activity and cleavage activity were tested simultanously (double-cleavage test). The reaction was started by addition of 0.2 pmol of human topo I and conducted for 20 min at 37 °C. After termination by addition of concentrated Laemmli buffer, cleavage complexes were separated by SDS−PAGE on 12% polyacrylamide gels and visualized by autoradiography. To test the influence of SRSF1, its variants and mutants, SRSF5, or SRSF9 on the DNA cleavage activity of topo I, each was added to the reaction mixture to achieve the same concentration as that in the kinase activity assay.

Two-Dimensional (2D) Electrophoresis. The kinase assay was conducted as described above with 180 pmol of SRSF1. The reaction was stopped by precipitation with TCA and acetone. After incubation for 30 min on ice, the proteins were spun down, washed three times with cold acetone, dried, and suspended in rehydration buffer [8 M urea, 0.5% CHAPS, 0.2% DTT, 0.5% IPG buffer 3-10 (GE Healthcare), and 0.002% bromophenol blue]. Isoelectric focusing (IEF) was conducted in a Multiphor II instrument (GE Healthcare) according to the manufacturer's instructions. Linear IPG 3-10 strips (GE Healthcare) were loaded with the sample, and IEF was continued with a maximal setting of 3500 V. Next, IPG strips were incubated in equilibration buffer [6 M urea, 50 mM Tris-HCl (pH 8.8), 2% SDS, 30% glycerol, 2% DTT, and 0.002% bromophenol blue] for 15 min. Proteins were separated via SDS−PAGE, transferred onto a PVDF membrane, and probed with anti-His antibodies.

Modeling. Topo I in the model of the topo I−SRSF1 complex was built on the basis of the model of the topo I−hnRNP A1 complex from our previous study[15] and using the crystal structure of topo I[11] deposited as Protein Data Bank (PDB) entry 1k4s. Missing residues in the PDB structure were generated using the Molecular Operating Environment (MOE, 2009) of the Chemical Computing Group. NT of topo I was predicted using the I-TASSER server.[33−35] SRSF1 was built by using the MOE homology modeling module and applying the following crystal structure patterns (PDB entries 1u1k, 1l3k, 1pgz, and 3beg). The missing C-terminal region was also added. SRSF1 was docked to topo I, maintaining the reported interactions between the protein molecules. Three versions of the topo I−SRSF1 complex were created. The first consists of the RRM1 domain of SRSF1 associated with topo I (the cap); the second one consists of the RRM1 domain and RS of SRSF1 associated with topo I (the cap and NT), and the last one

accounts for possible interactions between NT of topo I and RS of SRSF1. A series of short molecular dynamics simulations (100−300 ps in a single run) were applied to relax the model complexes. Next, optimization of the whole structure was conducted via energy minimization, with the norm of the gradient being <0.15 kcal/mol, keeping selected parts frozen. The complex was soaked in a periodic box of water molecules. TIP3P-type water molecules were used. Optimization of the whole system (topo I−SRSF1 complex and water) and the following molecular dynamics simulations were conducted with the NAMD2 package[36] and the Charmm27 force field. The productive simulation period took 5 ns without any constraints, with a constant pressure of 1013 hPa and a temperature of 300 K. The simulation time step was 1 fs. The complex appeared to be stable during the simulation period. Selected thermodynamic parameters, in particular the average binding energy for the topo I−SRSF1 complex, were computed using the simulation data. The average energies were computed after thermalization and equilibration periods, using the productive MD trajectories.

Relative orientations of both unstructured RS and NT were based on the experimental pull-down results, which favored antiparallel arrangements. The lowest interaction energy of the RS domain with the topo I skeleton, including the cap and NT, determined the final structural setting being consistent with all experimental findings. In practical terms, a number of possible modes of arrangement of the RS domain were generated, followed by MD simulations to compute the lowest interaction energy. The same procedure was applied to the RRM2 domain, resulting in the most probable configurations.

The spacers of SRSF1 were built using the MOE environment in a form of linear chains. They were subjected to minimization and MD simulations to determine their flexibility. Simulations were conducted in a vacuum, to accelerate folding of the spacers. The same simulation procedures were applied to each spacer. During the simulation period, changes in the distance between the terminal α-carbons were monitored.

■ **RESULTS**

Delineation of the Sites of Interaction between Topo I and SRSF1. Before modeling, we delineated the sites involved in interactions between topo I and SRSF1 that could be found by a pull-down method using fragments of both proteins. For the sake of clarity, the regions present in topo I are called here T-sites. The first and last amino acid residues of the region are given in brackets.

In previous work,[15,31] we have found that one SRSF1 site is in RRM1 (residues 13−19 and 56−58) and interacts with the T-site localized in the cap (residues 310−319).[15] Another T-site was localized in NT;[16] however, we did not know its precise localization or the SRSF1 region that interacts with this T-site. We were also not sure whether any SRSF1 site not listed above existed. To address all these issues, several pull-down experiments using fragments of topo I and SRSF1 were conducted. The results of basic experiments are shown in Figure S1 of the Supporting Information. They identified the interacting regions of RS and NT as SRSF1[221−248] and T[171−214], respectively. They also showed that the cap interacted not only with RRM1 but also with the spacer and RS. Because both the N-terminal (SRSF1[195−220]) and C-terminal (SRSF1[221−248]) fragments similarly bound to the cap, we assumed that the fragments used were too short to provide their specific interaction. Therefore, we conducted the
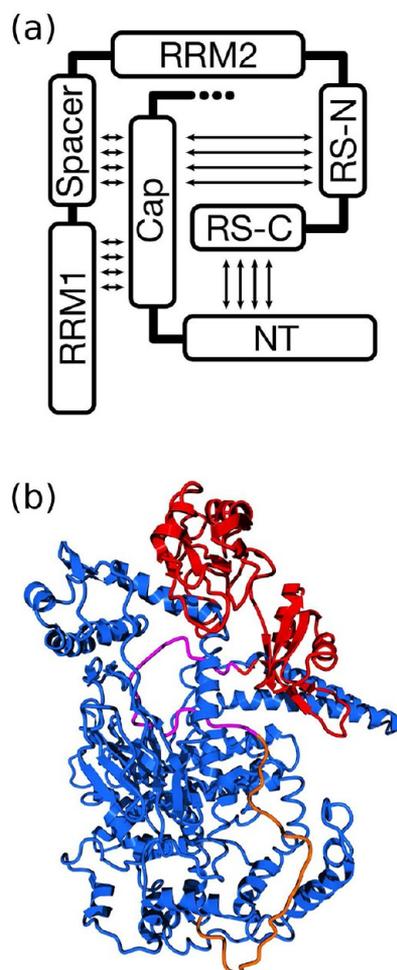
(a)

(b)

Figure 1. Complex of topo I and SRSF1. (a) Binding sites on topo I and SRSF1 revealed by the pull-down assay. Arrows denote regions whose interactions were confirmed by the pull-down assay. (b) Model of the complex. Topo I is colored blue, SRSF1 red, RS-N (residues 195−220) purple, and RS-C (residues 221−248) orange.

Table 1. Interactions between NT and Other Domains of Topo I in Its Complex with DNA or SRSF1

pair of residues (NT:partner domain)	domain or subdomain interacting with NT residues
Topo I−DNA Complex[a]	
Q201:H346	cap
W203:H346	cap
W203:W205	NT
W203:E208	NT
W205:W206	NT
W205:E208	NT
W206:D757	CT
Topo I−SRSF1 Complex[b]	
M1:H576	subdomain III
M1:D464	subdomain III
K46:D579	subdomain III
H54:K409	cap
E65:K216	cap
R84:D762	CT
K86:W441	subdomain III
K86:E438	subdomain III
N51:E438	subdomain III
H54:R434	subdomain III
E87:Q442	subdomain III
E91:R449	subdomain III
E107:K456	subdomain III
K197:E741	CT
E199:K746	CT
Q201:S432	cap
Q201:S433	cap
K204:E348	cap
W205:W754	CT
E208:R434	subdomain III

[a]From ref 38. [b]From this work.

pull-down assay with a longer topo I fragment containing both the cap and NT (T[1−433]) as a probe, and with two pieces of bait that were complete SRSF1 but contained artificial RSs. The RS of the first contained a doubled N-terminal part (residues 195−220) of wild-type RS of SRSF1 (SRSF1[2N]), whereas the RS of the other contained a doubled C-terminal part (residues 221−248) of wild-type RS (SRSF1[2C]). We assumed that if SRSF1[221−248] really bound to the cap, SRSF1[2C] would interact with the fragment containing both the cap and NT at least like wild-type SRSF1 (SRSF1[wt]) because it also contained a fragment interacting with NT at its C-terminal end. However, this was not the case: SRSF1[2C] bound to T[1−433] significantly worse than SRSF1 and also SRSF1[2N]. Therefore, we concluded that this is N-terminal fragment of RS that binds to the cap.

The final picture of localization of the binding sites that emerges from the findings described above and previous findings[15,16] was as follows (Figure 1a): (i) the T-sites are localized only in NT and in the cap; (ii) RS interacts with the T[171−214] region of the NT through its SRSF1[221−248] and with the cap through its SRSF1[195−220]; and (iii) two additional T-sites are localized in the cap and interact with either the spacer (SRSF1[89−119]) or the SRSF1[13−19] and SRSF1[56−58] regions in RRM1.

The Model. Building the model, we considered the interactions listed in the preceding section and the results of modeling of the previously described topo I−hnRNP A1 complex.[15] As compared to hnRNP A1 in the complex with topo I, RRM2 of SRSF1 is shifted closer to the surface of topo I because of the presence of the long, glycine-rich spacer and the lack of any stable interaction between the RRMs. In the case of hnRNP A1, two salt bridges contribute to the interaction between both RRM domains in hnRNP A1, keeping them close together;[37] however, they are absent in SRSF1.

The model of the topo I−SRSF1 complex is presented in Figure 1b. It involves four contact regions between both proteins. The first one shows the interaction between the cap and RRM1, described previously.[15] The next one comprised interaction between RS and topo I. Besides interaction with the cap and NT revealed by pull-down assays, RS interacts also with three residues localized in the CT between K746 and D757 (Table S2 of the Supporting Information). Interaction between the unstructuralized RS and NT is possible because of folding of NT that forms a cavity in which the C-terminal region of RS is placed. Interaction between the spacer joining RRMs and the cap, identified by pull-down assays, is confined to two hydrogen bonds. This is because the spacer is largely pushed away from the surface of topo I by both RRMs. In addition to the contact regions presumed before modeling, one
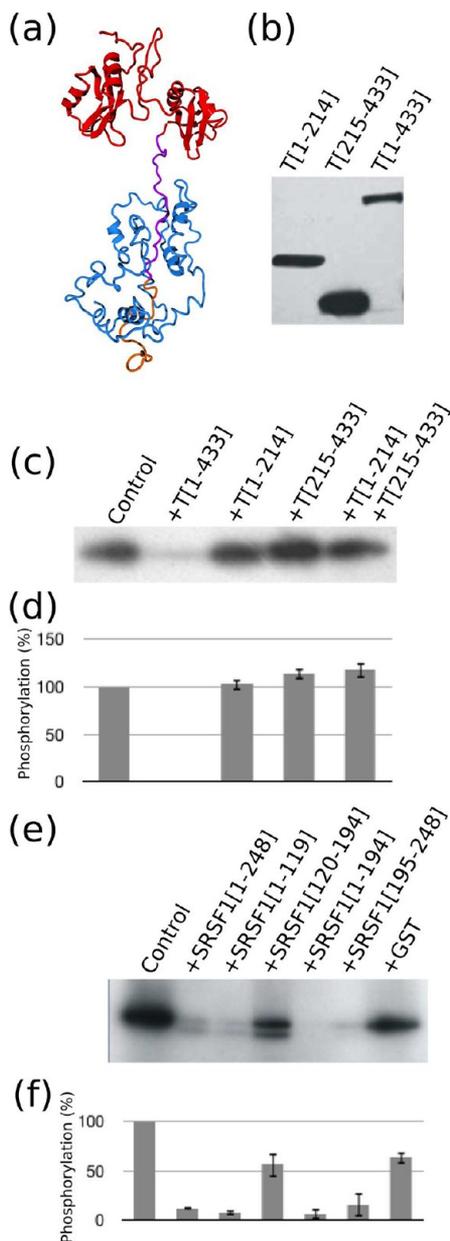
deviations were calculated from data obtained from three independent experiments.

Figure 2. Cooperation between binding sites. (a) Model of the complex of separated NT and SRSF1. Colors as in the legend of Figure 1b. (b) Stoichiometry of complexes between SRSF1 and topo I fragments. Pull-down experiments were performed with GST-tagged SRSF1 as bait and His-tagged topo I fragments as probes. The polypeptides are named according to the fragment of the full-length protein of which they are comprised; numbers in brackets refer to the first and last amino acid residues of the protein fragment, respectively (T stands for topo I). (c and d) Binding sites in topo I: competition between full-length topo I and its fragments comprising one (T[1−214] and T[215−433]) or both binding sites (T[1−433]) for SRSF1 in the kinase reaction. (e and f) Binding sites in SRSF1: competition between the full-length His-tagged substrate (SRSF1[1−248]) and the GST-tagged substrate or its fragments for topo I/kinase. (c−f) The fragments were added to the substrate prior to the enzyme in equimolar amounts with respect to SRSF1 but in 10-fold excess over topo I. A typical autoradiogram and quantitation of the results are presented. Autoradiographic bands were quantified by scanning densitometry, and data were expressed as a percentage of phosphorylation of SRSF1[1−248] used by topo I/kinase without any additions (±standard deviation). Mean values and standard
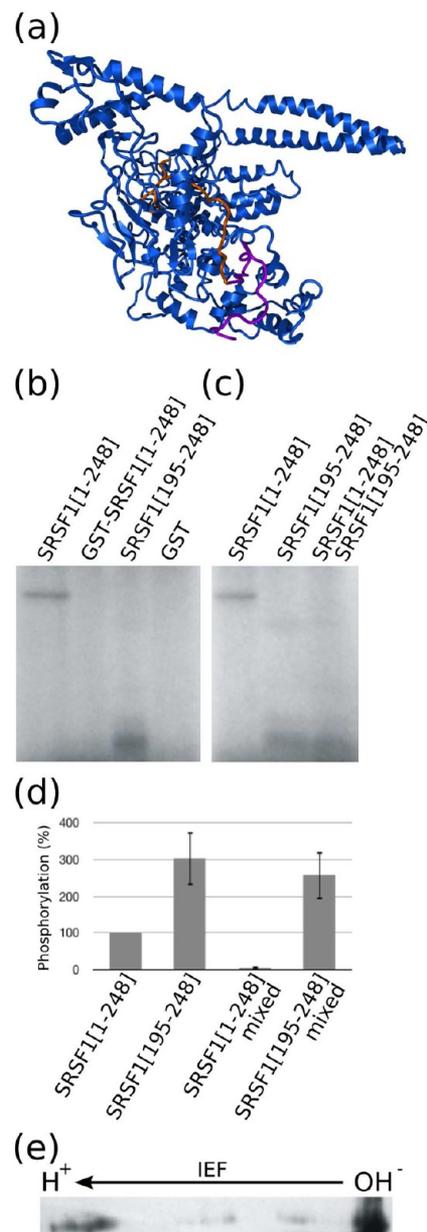
Figure 3. Phosphorylation of SRSF1 and its separated RS domain by topo I. (a) Model of the complex of separated RS and topo I. Colors as in the legend of Figure 1b. (b) Comparison of phosphorylation of complete SRSF1 (SRSF1[1−248]) and separated RS (SRSF1[195−248]). RS was obtained by proteolytic removal of the GST tag from GST-bound SRSF1[195−248]. (c) Phosphorylation of complete SRSF1 and separated RS in the 1:1 molar mixture. Typical autoradiograms are presented for panels b and c. (d) Quantitation of the results shown in panels b and c. Details as in the legend of Figure 2. (e) 2D electrophoresis of SRSF1[wt] phosphorylated by topo I/kinase. Unphosphorylated SRSF1[wt] is at the extreme right.

more region appeared between the linker and RRM2. It is established by three hydrogen bonds between residues of RRM2 placed below the nose/cone part of topo I and residues present on the top of the linker, between K650 and E696.
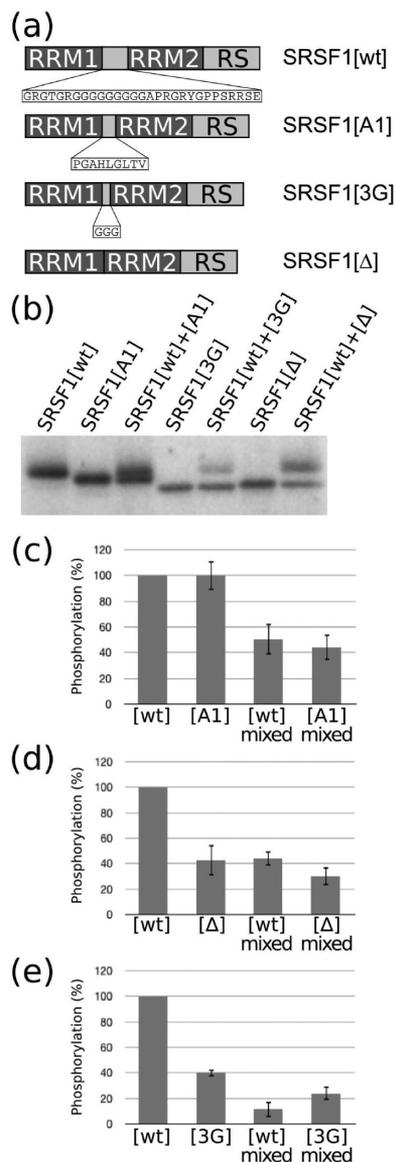
Left column

Figure 4. Phosphorylation of SRSF1[wt] and its spacer mutants. (a) Spacer mutants of SRSF1. (b) Phosphorylation of the recombinant proteins by topo I/kinase. Kinase assays were performed for the mutants added to topo I/kinase separately or in the 1:1 molar mixture with SRSF1[wt]. (c−e) Quantitative comparison of phosphorylation of SRSF1[wt] and its mutants. Details as in the legend of Figure 2.

Individual contact regions do not participate equally in the total average binding energy of the topo I−SRSF1 complex. The main contribution to the total binding energy (−2331.5 kcal/mol) comes from the interaction between RS and topo I (−1605.4 kcal/mol), especially from the interaction between both unfolded domains: RS and NT (−1038.1 kcal/mol). Interactions between topo I and SRSF1 are listed in Table S2 of the Supporting Information.

Interactions between NT and other domains of topo I are different for the protein in a complex with either DNA or SRSF1 (Table 1). In the first case, a majority of the interactions involve three tryptophan residues: W203, W205, and W206.[38,39] None of these interactions is saved in the topo I−SRSF1 complex. Thus, a change in the partner (from DNA in the relaxation activity to the protein substrate that is phosphorylated) is accompanied by rearrangement of intra-
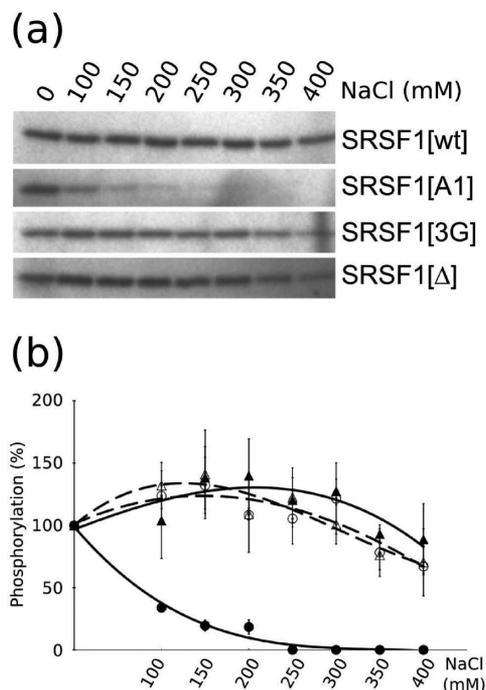
Right column

Figure 5. Salt dependence of phosphorylation of the wild type and spacer mutants of SRSF1. (a) Typical autoradiograms. (b) Quantitation of the results expressed as a percentage of protein phosphorylation in the absence of salt: (▲) SRSF1[wt], (●) SRSF1[A1], (○) SRSF1[3G], and (△) SRSF1[Δ]. Details as in the legend of Figure 2.

molecular interactions between NT and the remaining domains of topo I.

Cooperation between the Cap and NT inside Topo I. A direct structural conclusion from the model is that NT builds a binding site for RS of SRSF1 because of interaction of NT with the remaining regions of topo I: the cap, subdomain III, and CT. In particular, the cap not only comprises two binding sites for SRSF1 but also is necessary for structuralization of NT. Thus, we expected that the latter effect should be reflected by a cooperative binding of SRSF1 to sites localized in the cap and in NT, not observed for these sites analyzed individually. To examine it, we performed additional modeling studies, pull-down experiments, and the enzymatic tests.

Modeling of interactions of separated NT and SRSF1 confirmed that the ability to adopt a structure capable of binding to RS was not an intrinsic property of NT. An example of one of several possible complexes is shown in Figure 2a. Although some local secondary structures appeared along separated NT following its binding to SRSF1, they did not form a cavity in which RS could be placed as in the case of the NT remaining in the complete topo I (Figure 1b).

In the pull-down experiments, wild-type SRSF1 (SRSF1[1−248]) and three kinds of topo I fragments were used: T[1−214], which comprised NT; T[215−433], which comprised the cap; and T[1−433], which comprised both NT and the cap (Figure 2b). Only the T[1−433]−SRSF1[1−248] complex exhibited 1:1 stoichiometry, whereas both remaining fragments bound SRSF1[1−248] with an excess, which suggested an unspecific binding (stoichiometry of ∼1:3 and ∼1:5 for T[1−214] and T[215−433], respectively).

Direct evidence of cooperation between binding sites localized in the cap and in NT came from enzymatic tests in

(a)

(b)
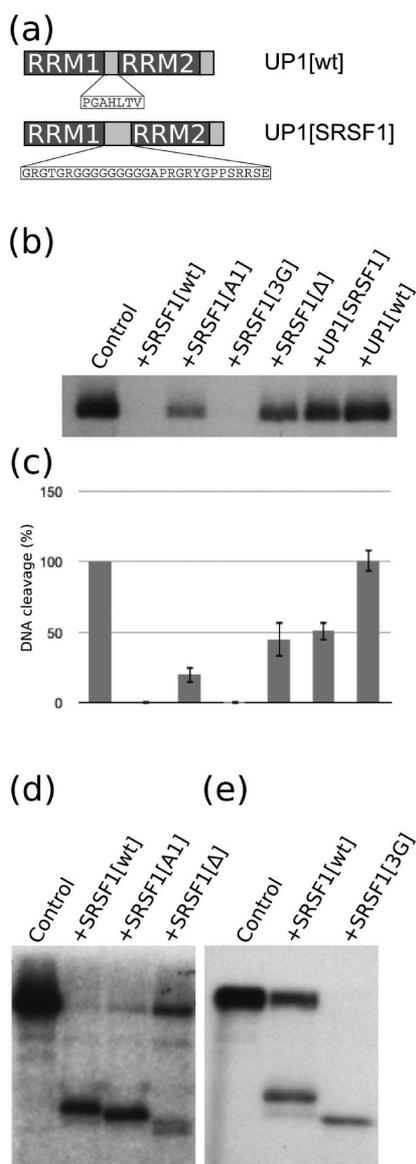

(c)

(d)   (e)

Figure 6. Inhibition of DNA nicking activity of topo I by RRM proteins with different spacers. (a) Spacers of UP1[wt] and UP1[SRSF1]. (b) Effect of the proteins on the DNA nicking activity of topo I. (c) Quantitation of the results presented in panel b. Details as in the legend of Figure 2. (d) Comparison of DNA nicking and kinase activity of topo I for SRSF1[wt], SRSF[A1], and SRSF1[Δ]. (e) Comparison of DNA nicking and kinase activity of topo I for SRSF1[wt] and SRSF1[3G]. To disclose more efficient binding of SRSF1[3G] to topo I with reference to SRSF1[wt], 5 times smaller amounts of both proteins, compared to other results presented in this work, were used (i.e., 0.2 pmol).

which recombinant fragments of the topo I protein competed with the complete topo I/kinase for the substrate. As one can see in panels c and d of Figure 2, neither T[1−214] nor T[215−433] inhibited phosphorylation of SRSF1, in a manner independent of whether they were added to the reaction medium separately or as a mixture of the fragments. In contrast, a strong inhibitory effect was observed for the fragment that covered all binding sites (T[1−433]). The structural background of the cooperation between the cap and NT may be explained by interactions between NT and the cap formed in the presence of SRSF1 that facilitate organization of the C-

terminal region of NT that binds the C-terminal part of RS (Figure 1b and Table 1). Because T[1−433] with tryptophan residues present in the C-terminal part of the NT mutated (W203A, W205A, and W206A) competed with topo I for a substrate as well as the original fragment (not shown), this additionally confirmed the conclusion drawn from the model that interaction formed by a tryptophan residue with the cap[38] is not present in the topo I−SRSF1 complex (Table 1).

To examine whether cooperation similar to that described above could occur between binding sites present in RRM1 and RS, we used GST-tagged polypeptides in the competition experiments. This was because they bind to topo I,[31] although they cannot be phosphorylated. As one can see in panels e and f of Figure 2, phosphorylation of the His-tagged SRSF1 was effectively inhibited by GST-tagged fragments that contained either RRM1 (SRSF1[1−119]) or RS (SRSF1[195−248]). These results indicate that no cooperation occurs between binding sites present in RRM1 and RS. Although no binding of GST to NT or the cap was revealed under conditions used in the pull-down assay (Figure S1 of the Supporting Information), a moderate inhibitory effect of the tag on phosphorylation was observed (Figure 2e,f). Most possibly, this effect contributed to the inhibition. Taking it into account, one can calculate that the inhibition of phosphorylation by the separated RRM2 (SRSF1[120−194]) was negligible.

In summary, the results presented in this section show that the sites localized in NT cooperate with those in the cap involved in the binding of SRSF1. This is in agreement with the model in which the cap facilitates the rearrangement of NT that provides a proper binding site for RS (Figure 1b and Table 1).

Role of the Interaction between the Cap and RRM1. Modeling of interactions of the separated RS domain with topo I revealed that several arrangements of the RS relative to topo I were possible (an example of RS bound in the reverse orientation to topo I is shown on Figure 3a). Therefore, in the complex formed by the complete SRSF1 and topo I (Figure 1b), RRM1 pins up the remaining domains to topo I and excludes in this way many possible arrangements of RS.

To test whether constraining of spatial arrangements of RS by binding of RRM1 to the cap influenced the efficiency of phosphorylation, we performed phosphorylation experiments with separated RS. Separated RS (SRSF1[195−248]) was phosphorylated much better than complete SRSF1[1−248] (Figure 3b,d). Moreover, when SRSF1[195−248] competed for the enzyme with complete SRSF1 (SRSF1[1−248]) added to the reaction mixture in an equimolar amount, phosphorylation of the wild-type protein was totally suppressed, pointing to SRSF1[195−248] as a preferable substrate (Figure 3c,d). In all the experiments described above, a short RS polypeptide was obtained directly by a proteolytic cleavage of GST-bound SRSF1[195−248]. The presence of a GST tag at the N-terminus of RS completely prevented phosphorylation of this fragment (Figure 3b).

The results presented in this section indicate that exclusion of several possible arrangements of RS relative to topo I results in a decrease in the level of phosphorylation. Topo I phosphorylates more than one serine residue in complete SRSF1 (Figure 3e). Therefore, a decreased level of phosphorylation may result from a reduced number of serine residues that are phosphorylated and/or an altered efficiency of the phosphoryl transfer.

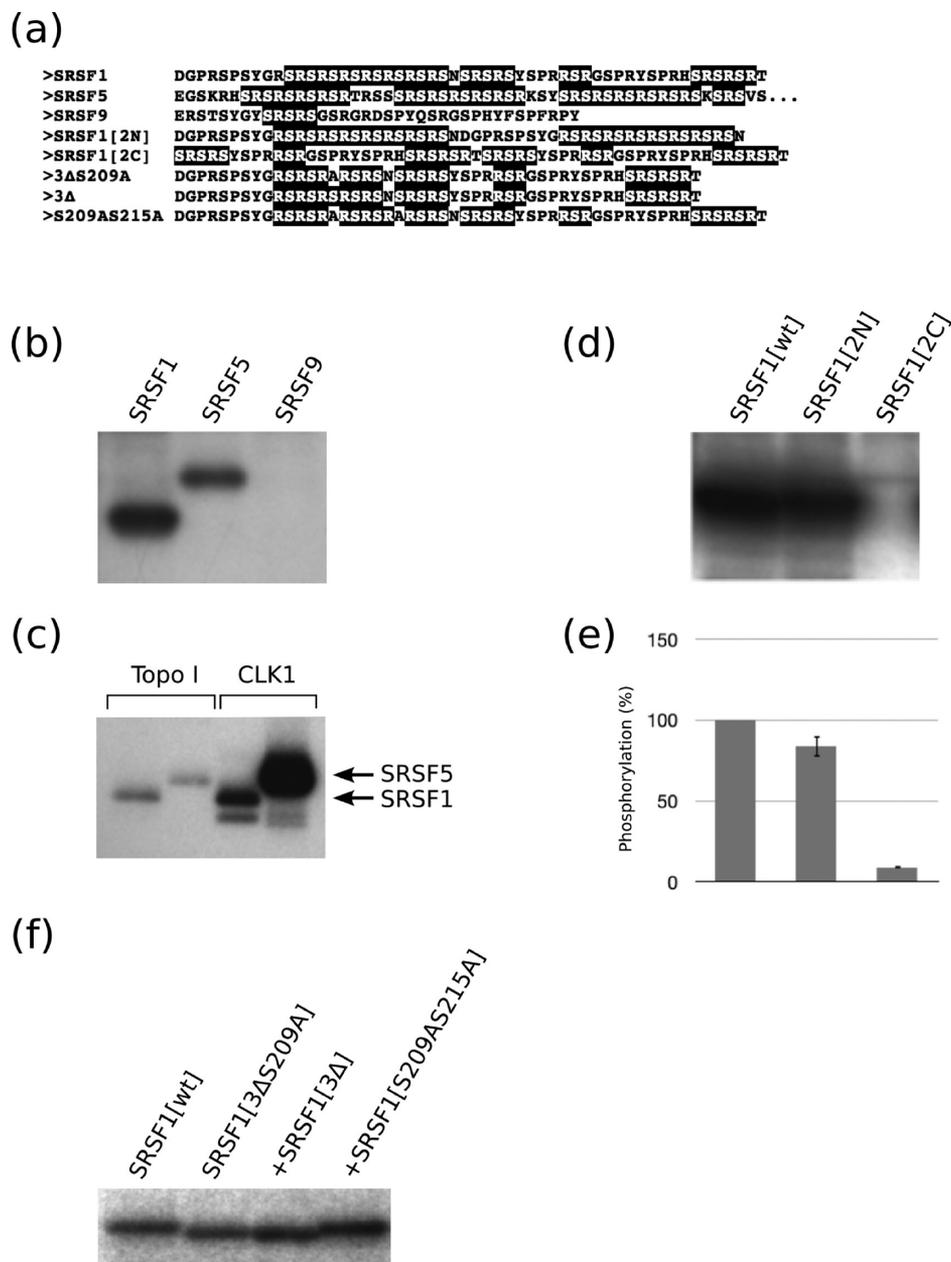**Phosphorylation of the Spacer Mutants.** In complete SRSF1, the constraint for the interaction between RS and topo

(a)

(b)

(c)

(d)

(e)

(f)

Figure 7. SRSF5, SRSF9, and the RS mutants of SRSF1 as substrates for topo I kinase. (a) Arrangement of RS/SR repeats in RS domains of the substrates. RS/SR repeats are highlighted in black. (b) SRSF5 and SRSF9 as substrates for topo I/kinase. (c) Comparison of phosphorylation of SRSF1 and SRSF5 by topo I and CLK1. (d) SRSF1[2N] and SRSF1[2C] as substrates for topo I/kinase. (e) Quantitation of the results presented in panel d. (f) Phosphorylation of SRSF1[3ΔS209A], SRSF1[3Δ], and SRSF1[S209AS215A] mutants by topo I/kinase. For panels b−f, details as in the legend of Figure 2.

I is imposed on SRSF1 by holding the first residue of RS in a fixed position. This position is directly determined by RRM2 laying on the linker of topo I (Figure 1b). However, RRM2 is only loosely bound to topo I, and the N- and C-termini of RRM2 are close to one another.[14] As a result, if SRSF1 is attached to the cap by RRM1, the position of the first residue of RS is practically determined by the length and flexibility of the spacer. Thus, we expected that we could change this position simply by modifying the properties of the spacer.

The wild-type spacer of SRSF1 is a glycine-rich polypeptide (Figure 4a). It contains 54% glycine residues, and nine of them are organized in a continuous tract that provides an unperturbed hinge region. Despite this, a free wild-type spacer did not behave as a flexible polypeptide. This was due to

positive charges of five arginine residues distributed along the polypeptide chain, because in silico substitution of all arginine residues with threonines made the wild-type spacer as flexible as the glycine homopolymer with the same length (Figure S2 of the Supporting Information). As positive charges of arginine residues may be screened under ionic conditions, the wild-type spacer is therefore expected to increase its flexibility with an increase in ionic strength.

SRSF1[A1] contained a spacer of a moderate length. The spacer was derived from the hnRNP A1 protein that is also composed of two RRM domains and binds to the cap, like SRSF1 does.[15] Besides the hnRNP A1 sequence, the spacer in SRSF1[A1] contained two additional residues not originally present in hnRNP A1 ([96]LG[97]), which appeared there because
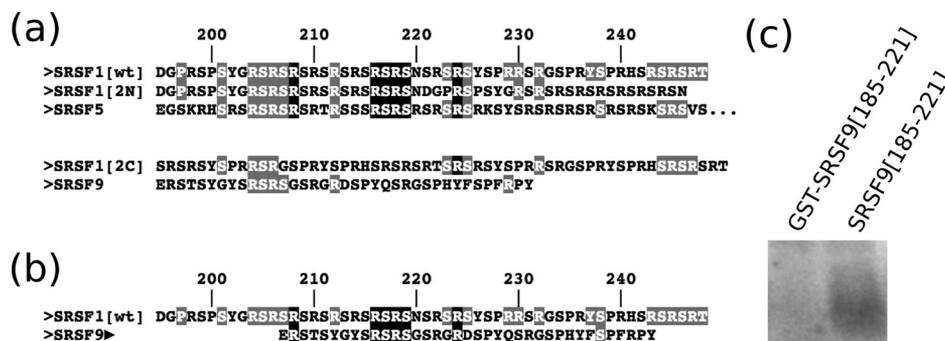
Figure 8. Identification of the motif recognized by topo I/kinase. (a) Distribution of the interacting residues in phosphorylated and nonphosphorylated proteins examined in this work. The putative motif recognized by topo I/kinase is highlighted in black, and other residues interacting with the enzyme are highlighted in gray. (b) Appearance of the recognized motif in separated RS of SRSF9. (c) Phosphorylation of separated RS of SRSF9 (SRSF9[185−221]) by topo I.

of the self-ligation requirements. Computations indicated that the central region of the [A1] spacer (amino acid residues 91−97) is folded into an α-helix. As a result, the free spacer folded down to approximately half of its initial length (Figure S2 of the Supporting Information). Although two glycine residues are present in the spacer, they are randomly distributed and constitute ∼20% of the residues. Because of all the properties mentioned above, the [A1] spacer was treated as moderately rigid.[40] As the [A1] spacer did not contain any charged amino acid residue, it was expected not to change its rigidity with an increase in ionic strength.

SRSF1[3G] contained a very short spacer composed of three glycine residues. It was considered as a very short glycine hinge that provides RRMs a possibility for any change of their direction as well as rotation around the spacer's axis.[40] Because SRSF1[3G] is very short, it also allows the RRM2 to adopt fewer conformations than the much longer wild-type spacer.

SRSF1[Δ] was completely devoid of the spacer, and their two RRMs were rigidly joined together.

All constructed mutants were phosphorylated by topo I/ kinase (Figure 4b). To gain insight into both, the reaction catalyzed by topo I/kinase and binding of the substrate to the enzyme, we performed two kinds of assays. In the first, the mutant was the only substrate for topo I/kinase, whereas in the other, it competed for the enzyme with SRSF1[wt] added in an equimolar amount to the reaction mixture. Both assays were conducted under conditions under which the enzyme was saturated with the substrate, so that the amount of the phosphorylated protein was independent of the amount of the subsequently added substrate. We also checked that no selective precipitation of the particular mutant occurred, when it was the only substrate or as it was mixed with another (not shown).

The first assay showed that SRSF1[A1] was phosphorylated to an extent similar to that of SRSF1[wt], but the mutants with a shorter spacer or without a spacer, i.e., SRSF1[3G] and SRSF1[Δ], were ∼2−3 times worse substrates than a wild-type one (Figure 4b−e). In a mixture, SRSF1[wt] and SRSF1[A1] were phosphorylated to the same level that was equal to one-half of the extent achieved when each of them was the only substrate for topo I (Figure 4c). It could be thus concluded that they were not only similarly good substrates but they also similarly bound to topo I. Within an experimental error, the latter conclusion roughly applied also to the SRSF1[wt] and SRSF1[Δ] pair. Although both substrates were phosphorylated with a different efficiency, a total level of their phosphorylation

in the mixture made exactly one-half of the sum of extents achieved by the substrates phosphorylated separately (Figure 4d). A different picture was observed for the SRSF1[wt] and SRSF1[3G] pair, where the presence of SRSF1[3G] significantly impaired phosphorylation of SRSF1[wt] (Figure 4e). Thus, we concluded that although the latter mutant is phosphorylated with a lower efficiency, it bound to topo I more strongly than the wild-type protein. This conclusion is in agreement with experiments examining inhibition of DNA nicking activity of topo I by SRSF1[3G] presented below (see Figure 6e).

We also examined the properties of the spacers resulting from their sensitivity to ionic strength. As mentioned above, a relative rigidity of the wild-type spacer observed at low ionic strengths could be largely neutralized in the moderate-ionic strength environment by screening of the arginine residues charges. On the other hand, ionic conditions should also weaken numerous interactions between RRM2 and RS from one side and topo I from the other because many of them are formed by acidic and basic residues (Table S2 of the Supporting Information). If both effects compensated one another, only minor changes in the arrangement of RS relative to topo I would be expected with an increase in ionic strength. In fact, phosphorylation of SRSF1[wt] by topo I/kinase was only weakly sensitive to ionic strength (Figure 5). In contrast to the discussion above, the level of phosphorylation of SRSF1-[A1] rapidly dropped with a growing concentration of NaCl (Figure 5), while no changes in the solubility of SRSF1[A1] accompanied this effect (not shown). The only difference between the [wt] and [A1] spacer is that the latter should not undergo a change in its rigidity with an increasing ionic strength. Thus, an explanation of the unique salt dependence of phosphorylation of SRSF1[A1] may be that the moderately rigid spacer did not allow RRM2 to accommodate its arrangement relative to the linker when interactions between RRM2 and the linker and between RS and the cap or NT were weakened by salt. It is worth noting that a similar high sensitivity to salt has been observed for RNA binding by hnRNP A1,[41] which was a source of the spacer introduced into SRSF1[A1].

Salt dependence curves similar to that described for SRSF1[wt] were observed when SRSF1[3G] and SRSF1[Δ] were used as substrates (Figure 5). In the first case, the spacer is devoid of any charged residues like [A1]; however, in a manner independent of ionic strength, it is too flexible to be any obstacle for arrangement of RRM2 and RS domains. The same

concerns SRSF1[Δ] where the arrangement of both domains does not depend on the spacer.

A general conclusion drawn from experiments described in this section is that to be efficiently phosphorylated SRSF1 requires a flexible spacer that allows the arrangement of RRM2 and, as a consequence, RS on the surface of topo I. Moreover, the spacer must not be shorter than a border length that is between three and nine residues. The latter requirement may result from at least two properties of shorter spacers. The first possibility is that proteins with shorter spacers bind to topo I too strongly to be efficiently phosphorylated, which was observed for SRSF1[3G]. The other one is that shorter spacers are not able to accept all conformations necessary during multisite phosphorylation of the substrate. As shown above (Figure 3e), several serine residues were phosphorylated in SRSF1.

Inhibition of the DNA Nicking Activity of Topo I by Spacer Mutants. As the inhibition of DNA nicking by SRSF1 depends on RRMs joined by a spacer,[31] we expected that different spacer mutants should differently influence the DNA nicking activity of topo I.

We have previously shown that the DNA nicking activity of topo I is effectively inhibited by SRSF1[9] but not by UP1,[15] another protein that contains two RRMs and binds to the cap. UP1 is a shortened form of hnRNP A1, devoid of the C-terminal domain.[37] To clarify the role of the spacer in the inhibition of DNA nicking activity, we added here a new protein to the previously described set of spacer mutants. It was composed of the RRMs of UP1 joined together by the spacer coming from SRSF1[wt] [UP1[SRSF1] (Figure 6a)]. We confirmed a strong inhibition of topo I DNA nicking activity for SRSF1[wt] and a lack of any inhibitory effect for UP1[wt] (Figure 6b,c). However, we observed that UP1 acquired some inhibitory properties when the spacer from SRSF1[wt] was introduced between its RRMs [UP1[SRSF1] (Figure 6b,c)]. On the other hand, SRSF1 exhibited a less pronounced inhibitory effect when the wild-type spacer was substituted with shorter and less flexible [A1] or the protein was devoid of any spacer. A strong inhibitory effect of SRSF1[3G] indicated that there is no simple relationship between the length of the linker and inhibition of DNA nicking activity. Instead, the flexibility of the linker seems to be an important factor that promotes the inhibition. It was clearly visible in the double-cleavage assay in which both DNA nicking and the kinase activities of topo I could be observed simultaneously.[32] For SRSF1[wt], SRSF1-[A1], and SRSF1[Δ], a roughly reverse relationship was observed between both activities: the lower the level of phosphorylation of the mutated substrate, the higher the level of inhibition of DNA nicking (Figure 6d). However, a different picture was observed for SRSF1[3G]. Under conditions that allowed for both phosphorylation of SRSF1[wt] and nicking of the DNA substrate, SRSF1[3G] completely inhibited DNA nicking, although it was phosphorylated to the lower extent than the wild-type protein (Figure 6e). The latter observation showed that a short but flexible spacer of SRSF1[3G] makes it a better inhibitor of DNA nicking activity but a worse phosphorylation substrate than SRSF1[wt].

Although the flexibility of the spacer seems to be essential for inhibition of DNA nicking, the contribution of the length of the spacer is observed in the case of UP1[SRSF1]. As mentioned above, two salt bridges stabilize interaction between both RRM domains in hnRNP A1 keeping them close together.[37] Therefore, moderate inhibitory properties exhibited by UP1-

[SRSF1] at low ionic strengths cannot come from the modified arrangement of RRM2 on topo I but rather from the presence of a longer spacer.

Role of Organization of the RS Domain in Phosphorylation of the Substrate. As it has been shown for SRPK1, phosphorylation of SRSF1 is highly dependent on the number and distribution of RS/SR repeats in the RS.[23] Here, we examined whether both parameters influenced the efficiency of phosphorylation catalyzed by topo I/kinase. Because SRSF1 contains 14 RS/SR repeats gathered in one long cluster of eight repeats and three shorter regions, we compared its phosphorylation with (i) two human SR proteins, SRSF5 and SRSF9 (according to the recently proposed nomenclature;[13] previously named SRp40 and SRp30c, respectively), that had a total number of RS/SR repeats different from that of SRSF1; (ii) two previously described RS mutants of SRSF1, SRSF1[2N] and SRSF1[2C], that had differently clustered RS/SR repeats; and (iii) point mutants of SRSF1 in which a long cluster of RS/SR repeats was split into shorter ones (Figure 7a).

(i) Both SRSF5 and SRSF9 bound to the cap and to NT (Figure S3 of the Supporting Information), like SRSF1. Analysis of flexibility indicated that the spacers of SRSF5 and SRSF9 folded like the spacer of SRSF1: they were flexible upon neutralization of positive charges on arginine residues, mimicked by substitution of arginines with threonines (Figure S3 of the Supporting Information). The similarity of the regions of SRSF1, SRSF5, and SRSF9 comprising the RRMs linked by the spacer was further confirmed because of the same ability of the all proteins to inhibit DNA nicking (Figure S3 of the Supporting Information). As we have demonstrated previously for SRSF1,[31] the inhibition depends on the fragment composed of the RRMs and the spacer but not on the RS.

RS domains of SRSF1, SRSF5, and SRSF9 differ in the number of serine residues in RS/SR repeats [15, 28, and 5, respectively (Figure 7a)]. Besides SRSF1, topo I phosphorylated SRSF5 but not SRSF9 (Figure 7b). However, although SRSF5 was phosphorylated by topo I, it was a poorer substrate than SRSF1. The lower extent of phosphorylation of SRSF5 is unique for topo I/kinase because CLK1 phosphorylated SRSF5 better than SRSF1 (Figure 7c). This observation indicated that there was no simple relationship between a total number of RS/SR repeats and phosphorylation efficiency. It also suggested that a maximal number of repeats that was 8 for SRSF1 might be a key factor.

(ii) To verify the hypothesis described above, we used the RS mutants that contained differently clustered RS/SR repeats. Two long clusters of SRSF1[2N] were composed of eight RS/SR repeats each, whereas SRSF1[2C] included four short clusters built of three RS/SR repeats each, distributed along the RS (Figure 7a). Only SRSF1[2N] was phosphorylated by topo I, at a level slightly lower than that of SRSF1[wt] (Figure 7d,e). As all the proteins contained a similar total number of serine residues in RS/SR repeats (14, 16, and 12 in SRSF1[wt], SRSF1[2N], and SRSF1[2C], respectively), a number of the RS/SR repeats in the cluster might be considered as a factor that facilitates phosphorylation. It would be in agreement with a previous suggestion that a continuous tract composed of at least five RS/SR repeats is indispensable for SRSF1 to be a substrate of topo I/kinase.[16]

(iii) This idea failed when directly tested using mutants of SRSF1 lacking one or more serine residues exclusively present in the long RS/SR cluster (names of mutations are given in brackets). We present here three of them that differ in the

number of repeats in clusters (Figure 7f). The SRSF1[3Δ] mutant was devoid of three RS repeats and retained exactly five successive RS/SR repeats in one cluster, whereas serine → alanine mutants SRSF1[S209AS215A] and SRSF1[3ΔS209A] had only two successive RS/SR repeats in small clusters that derived from the original long one. All the mutants were phosphorylated by topo I with a similar efficiency (Figure 7f), indicating that the length of the RS/SR cluster is not critical for phosphorylation of the substrate.

Therefore, we asked whether topo I recognized a specific pattern of residues in the RS that allowed it to bind and phosphorylate the substrate. To address this question, we compared amino acid sequences of phosphorylated and nonphosphorylated proteins examined here and applied the results to residues involved in the interaction with topo I previously identified in the model (Table S2 of the Supporting Information). The comparison revealed a pattern localized in the region of residues 208−224 of SRSF1 that is present in all the phosphorylated but not in the nonphosphorylated proteins and might be a motif recognized by topo I (Figure 8a). Of six residues present in the pattern, R208 is placed at the top of the loop that enters into topo I under the nose cone of the cap[11] and therefore determines a depth of penetration of the loop and a layout of the remaining part of RS on topo I (Figure 1b). Among the other residues of the pattern, S217 and R224 provide interaction of RS with the C-terminal region of topo I where the ATP binding site is localized.[19]

Because of the constraint imposed on RS by the fixed position of its N-terminal end (Figure 1b), the residues of the recognition motif have to fit to their partners on topo I. Such a constraint could be of key importance for the RS of SRSF9, which unveiled the recognition motif if it was shifted toward the N-terminal part of topo I (Figure 8b). To examine this assumption, we removed the constraint using a separated RS of SRSF9 (SRSF9[185−221]) as a substrate, and we found that under such conditions, it was phosphorylated by topo I (Figure 8c). This result confirmed that to phosphorylate the substrate topo I/kinase requires both the recognition motif in the RS and a structural fitting of RS to the critical binding sites in topo I. Identification of exact residues composing the recognition motif requires further detailed studies.

■ DISCUSSION

Interaction between the unstructured NT[17] and RS[18] domains in the topo I−SRSF1 complex was a serious ambiguity for determining structures by crystallographic studies. This paper presents an alternative approach that accounts for the modeling of the complex. It resulted in a structure that well agrees with biochemical properties of topo I bound with SRSF1. They are (i) cooperation of T-sites in SRSF1 binding, (ii) suppression of phosphorylation by binding of RRM1 to the cap, and (iii) the influence of the spacer's properties on phosphorylation and inhibition of DNA nicking.

(i) The model infers that a direct interaction between the two unstructured domains, NT and RS, does not provide a structure present in the topo I−SRSF1 complex (Figures 1b and 2a). Instead, this structure appears upon binding of RS because of interactions of NT with structuralized regions of topo I. It is confirmed for sites present in the fragment comprising the cap and NT (T[1−433]) as cooperation in the binding of SRSF1 revealed by the ability of the fragment to maintain a 1:1 stoichiometry in the complex and an effective competition with the complete topo I for SRSF1 binding.

However, the model shows that complete topo I forms many more interactions with SRSF1 than fragment T[1−433]. In particular, all interactions formed by RRM2 with the linker and between NT and subdomain III and CT (Table 1) are lost in the complex of SRSF1 and T[1−433]. To explain it, we consider the entropic expense paid by topo I binding to SRSF1 because of the reduction in the flexibility of the linker domain as a reason for the lowered affinity of the complete topo I for SRSF1. The linker domain is the most pliable part of topo I.[42−44] We assume that the flexibility of the linker domain has to be reduced upon binding with RRM2. The range of displacement of the three residues in the flexible linker slightly exceeds the length of hydrogen bonds formed because of the interaction with RRM2.[43] Therefore, formation of the contact between RRM2 and the linker domain may result in exclusion of several conformations of the linker present in free topo I but not remaining in the complex.

(ii) The model infers restrictions in possible arrangements of the RS imposed by binding of RRM1 to the cap (Figures 1b and 3a). Consistent with this, phosphorylation of the RS was shown here to be inhibited in SRSF1 compared to that in free RS. Additional binding sites distal from the site where phosphoryl transfer occurs are often used by protein kinases, usually to increase their binding affinity for the protein substrate.[45] A reason of suppression of phosphorylation suggested here is an exclusion of several arrangements of RS that promote phosphorylation. Indeed, shifting of separated RS of SRSF1 toward the N-terminal part of topo I unveils two additional patterns of residues that can be recognized by topo I, one starting from R198 and the other from R206. One more pattern starting from R214 appears upon shifting the RS toward the C-terminal part of topo I. Similarly, removal of the constraint introduced by RRM1 allowed separated RS of SRSF9 to shift toward the N-terminal part of topo I and to make use of the otherwise covered recognition motif (Figure 8c).

(iii) The model infers that the spacer determines the arrangement of RS and RRM2 relative to topo I. It was confirmed here as the influence of the spacer's properties on phosphorylation that depends on the arrangement of RS, and inhibition of DNA nicking that depends on the arrangement of RRM2. Consistent with this, somewhat different structural features of modified spacers determined the activity of the mutants as substrates for phosphorylation or inhibitors of DNA nicking. Although both activities need the flexibility of the SRSF1 molecule to arrange it on the surface of topo I, the kinase activity additionally requires the minimal length of the spacer that is necessary to establish the exact position of the N-terminal end of RS relative to its binding site in the cap region. On the other hand, the efficiency of phosphorylation was not dependent on the length of the spacer exceeding the minimal value, whereas a longer spacer localized in the cavity normally occupied by DNA influenced the ability of topo I to nick DNA.

We identified here a motif in the RS that might be recognized by topo I/kinase. We think that the amino acid sequence recognized by topo I/kinase might be necessary for a proper positioning of serine residues relative to ATP. Localization of the ATP binding site in topo I is not known except for identification of the region necessary for the binding that has been roughly localized in the C-terminal part of topo I.[19] Preliminary studies performed by us suggest that ATP might bind to the pocket localized on the border of core subdomain III and the C-terminal domain (Krzyśko and

Article

Bukowicki, unpublished results). According to the model presented here, serine residues held close to the γ-phosphate of ATP bound to this putative pocket are in the middle region of RS.

The overall picture of the topo I−SRSF1 complex appears as follows (Figure 9a). Interaction between RRM1 and the cap ensures the RS domain a fixed orientation relative to topo I. The properties of the linker determine the position of RRM2 and, as a consequence, the position of the N-terminal residue of the RS relative to the cap. It allows the RS residues in the recognition motif to bind their partners in topo I. If the recognition motif is too close to RRM2, which is the case for SRSF9, it does not bind to the partners in topo I (Figure 9b). The interaction between topo I and SRSF1 is completed when the C-terminal fragment of RS meets NT and eventually zips up the complex (Figure 9a). It is accompanied by structuralization of NT because of its binding to remaining regions of topo I. Phosphorylation efficiency strongly depends on the spacer flexibility that determines an arrangement of RS on topo I (Figure 9c). An arrangement of RRM2 in the cavity normally occupied by DNA results in the inhibition of DNA nicking activity. If RRM2 is placed outside the cavity, which is the case for UP1, DNA nicking is not inhibited (Figure 9d).

Experiments conducted with the spacer mutants indicated that the flexibility of the spacer contributes to the ability of SRSF1 to be phosphorylated and to inhibit DNA nicking. Especially, the salt dependence of phosphorylation efficiency (Figure 5) directly shows that efficiency of phosphorylation can be easily modified by a change in the flexibility of the spacer. These observations might be biologically relevant because at least two cellular factors have been identified to modify properties of the spacer and thus may be considered as potential regulators of the effects of SRSF1 on the activities of topo I. First, arginine residues present in the spacer have been shown to provide a binding site for mRNA export factor TAP.[46] Next, it has recently been demonstrated that arginine residues of the spacer are methylated and control in this way the subcellular localization of SRSF1.[47] The simulation conducted by us indicates that methylation of the arginine residues may subtly regulate the flexibility of the spacer of SRSF1 (Figure S4 of the Supporting Information).

ASSOCIATED CONTENT

Supporting Information

A list of primers used in this work, detailed data for interactions between topo I and SRSF1, mapping of the binding sites between topo I and SRSF1 by a GST pull-down assay, folding properties of the intact SRSF1 spacer and its engineered variants, binding of SRSF5 and SRSF9 to topo I and the influence of the binding on DNA nicking activity, and a simulation indicating that methylation of the arginine residues in the spacer of SRSF1 may regulate its flexibility. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Institute of Biochemistry, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland. Phone: (+4822)5543114. Fax: (+4822)5543221. E-mail: staron@biol.uw.edu.pl.
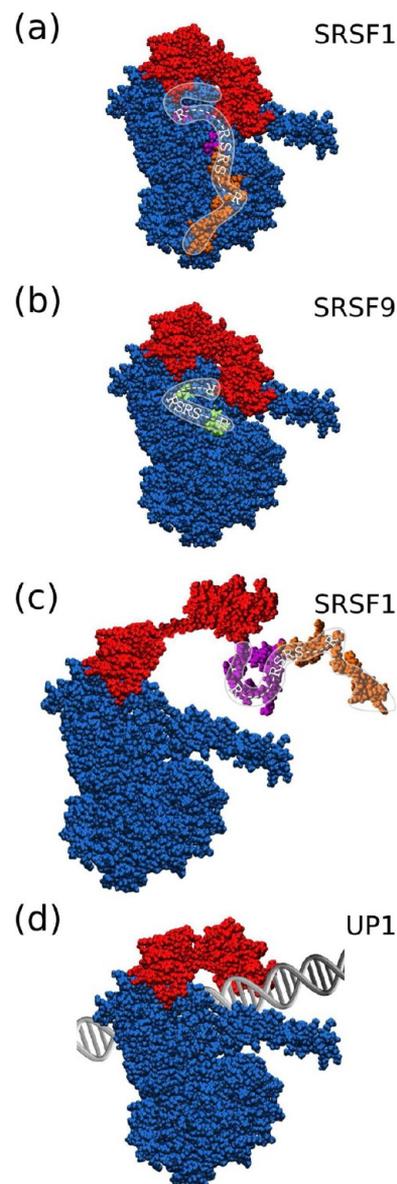

Figure 9. Cartoon depicting activities of topo I in the complex with RRM proteins. The gray line indicates the location of RS of proteins interacting with topo I. (a) Kinase activity of the topo I−SRSF1 complex. The recognition motif in RS of SRSF1 is positioned at its binding site in topo I and allows phosphorylation. (b) Kinase activity of the topo I−SRSF9 complex. Although the recognition motif exists in RS, structural constrains prevent it from being positioned at the binding site and preclude phosphorylation. (c) Regulation of the kinase activity by changes in the flexibility of the spacer of SRSF1. A decrease in flexibility leads to disruption of the interaction between the recognition motif and its binding site and precludes phosphorylation. (d) DNA nicking activity of the topo I−UP1 complex. Because of the arrangement of RRM2 outside the cavity wherein DNA binds, DNA nicking activity is observed for the complex.

Notes

The authors declare no competing financial interest.

■ ABBREVIATIONS

CT, C-terminal domain of topo I; hnRNP A1, heterogeneous nuclear ribonucleoprotein A1; NT, N-terminal domain of topo I; RRM, RNA recognition motif; RS, domain of SRSF1, SRSF5, or SRSF9 that contains RS/SR repeats; SRSF1, alternative splicing factor/splicing factor 2; SRSF5, splicing factor SRp40; SRSF9, splicing factor SRp30c; topo I, human DNA topoisomerase I; UP1, unwinding protein 1; wt, wild-type.

■ REFERENCES

(1) Leppard, J. B., and Champoux, J. J. (2005) Human DNA topoisomerase I: Relaxation, roles, and damage control. *Chromosoma* 114, 75–85.

(2) Kim, N., Huang, S.-Y. N., Williams, J. S., Li, Y. C., Clark, A. B., Cho, J.-E., Kunkel, T. A., Pommier, Y., and Jinks-Robertson, S. (2011) Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science* 332, 1561–1564.

(3) Kim, E., Goren, A., and Ast, G. (2008) Alternative splicing and disease. *RNA Biol 5*, 17–19.

(4) Rossi, F., Labourier, E., Forné, T., Divita, G., Derancourt, J., Riou, J. F., Antoine, E., Cathala, G., Brunel, C., and Tazi, J. (1996) Specific phosphorylation of SR proteins by mammalian DNA topoisomerase I. *Nature* 381, 80–82.

(5) Graveley, B. R. (2000) Sorting out the complexity of SR protein functions. *RNA 6*, 1197–1211.

(6) Soret, J., Gabut, M., Dupon, C., Kohlhagen, G., Stévenin, J., Pommier, Y., and Tazi, J. (2003) Altered serine/arginine-rich protein phosphorylation and exonic enhancer-dependent splicing in mammalian cells lacking topoisomerase I. *Cancer Res. 63*, 8203–8211.

(7) Tuduri, S., Crabbé, L., Conti, C., Tourrière, H., Holtgreve-Grez, H., Jauch, A., Pantesco, V., de Vos, J., Thomas, A., Theillet, C., Pommier, Y., Tazi, J., Coquelle, A., and Pasero, P. (2009) Topoisomerase I suppresses genomic instability by preventing interference between replication and transcription. *Nat. Cell Biol. 11*, 1315–1324.

(8) Tuduri, S., Crabbé, L., Tourrière, H., Coquelle, A., and Pasero, P. (2010) Does interference between replication and transcription contribute to genomic instability in cancer cells? *Cell Cycle 9*, 1886–1892.

(9) Kowalska-Loth, B., Girstun, A., Piekielko-Witkowska, A., and Staron, K. (2002) SF2/ASF protein inhibits camptothecin-induced DNA cleavage by human topoisomerase I. *Eur. J. Biochem. 269*, 3504–3510.

(10) Chen, H. J., and Hwang, J. (1999) Binding of ATP to human DNA topoisomerase I resulting in an alteration of the conformation of the enzyme. *Eur. J. Biochem. 265*, 367–375.

(11) Redinbo, M. R., Stewart, L., Kuhn, P., Champoux, J. J., and Hol, W. G. (1998) Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science 279*, 1504–1513.

(12) Baker, N. M., Rajan, R., and Mondragón, A. (2009) Structural studies of type I topoisomerases. *Nucleic Acids Res. 37*, 693–701.

(13) Manley, J. L., and Krainer, A. R. (2010) A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes Dev. 24*, 1073–1074.

(14) Ngo, J. C. K., Giang, K., Chakrabarti, S., Ma, C.-T., Huynh, N., Hagopian, J. C., Dorrestein, P. C., Fu, X.-D., Adams, J. A., and Ghosh, G. (2008) A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Mol. Cell 29*, 563–576.

(15) Trzcińska-Daneluti, A. M., Górecki, A., Czubaty, A., Kowalska-Loth, B., Girstun, A., Murawska, M., Lesyng, B., and Staron, K. (2007) RRM proteins interacting with the cap region of topoisomerase I. *J. Mol. Biol. 369*, 1098–1112.

(16) Labourier, E., Rossi, F., Gallouzi, I. E., Allemand, E., Divita, G., and Tazi, J. (1998) Interaction between the N-terminal domain of human DNA topoisomerase I and the arginine-serine domain of its substrate determines phosphorylation of SF2/ASF splicing factor. *Nucleic Acids Res. 26*, 2955–2962.

(17) Vassallo, O., Castelli, S., D'Annessa, I., della Rocca, B. M., Stella, L., Knudsen, B. R., and Desideri, A. (2010) Evidences of a natively unfolded state for the human topoisomerase IB N-terminal domain. *Amino Acids 41*, 945–953.

(18) Haynes, C., and Iakoucheva, L. M. (2006) Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Res. 34*, 305–312.

(19) Rossi, F., Labourier, E., Gallouzi, I. E., Derancourt, J., Allemand, E., Divita, G., and Tazi, J. (1998) The C-terminal domain but not the tyrosine 723 of human DNA topoisomerase I active site contributes to kinase activity. *Nucleic Acids Res. 26*, 2963–2970.

(20) Gui, J. F., Tronchère, H., Chandler, S. D., and Fu, X. D. (1994) Purification and characterization of a kinase specific for the serine- and arginine-rich pre-mRNA splicing factors. *Proc. Natl. Acad. Sci. U.S.A. 91*, 10824–10828.

(21) Colwill, K., Pawson, T., Andrews, B., Prasad, J., Manley, J. L., Bell, J. C., and Duncan, P. I. (1996) The Clk/Sty protein kinase phosphorylates SR splicing factors and regulates their intranuclear distribution. *EMBO J. 15*, 265–275.

(22) Ding, J.-H., Zhong, X.-Y., Hagopian, J. C., Cruz, M. M., Ghosh, G., Feramisco, J., Adams, J. A., and Fu, X.-D. (2006) Regulated cellular partitioning of SR protein-specific kinases in mammalian cells. *Mol. Biol. Cell 17*, 876–885.

(23) Ma, C.-T., Hagopian, J. C., Ghosh, G., Fu, X.-D., and Adams, J. A. (2009) Regiospecific phosphorylation control of the SR protein ASF/SF2 by SRPK1. *J. Mol. Biol. 390*, 618–634.

(24) Ghosh, G., and Adams, J. A. (2011) Phosphorylation mechanism and structure of serine-arginine protein kinases. *FEBS J. 278*, 587–597.

(25) Velazquez-Dones, A., Hagopian, J. C., Ma, C.-T., Zhong, X.-Y., Zhou, H., Ghosh, G., Fu, X.-D., and Adams, J. A. (2005) Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *J. Biol. Chem. 280*, 41761–41768.

(26) Bullock, A. N., Das, S., Debreczeni, J. E., Rellos, P., Fedorov, O., Niesen, F. H., Guo, K., Papagrigoriou, E., Amos, A. L., Cho, S., Turk, B. E., Ghosh, G., and Knapp, S. (2009) Kinase domain insertions define distinct roles of CLK kinases in SR protein phosphorylation. *Struct. Folding Des. 17*, 352–362.

(27) Ngo, J. C. K., Chakrabarti, S., Ding, J.-H., Velazquez-Dones, A., Nolen, B., Aubol, B. E., Adams, J. A., Fu, X.-D., and Ghosh, G. (2005) Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Mol. Cell 20*, 77–89.

(28) Hagopian, J. C., Ma, C.-T., Meade, B. R., Albuquerque, C. P., Ngo, J. C. K., Ghosh, G., Jennings, P. A., Fu, X.-D., and Adams, J. A. (2008) Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1. *J. Mol. Biol. 382*, 894–909.

(29) Lippert, M. J., Kim, N., Cho, J.-E., Larson, R. P., Schoenly, N. E., O'Shea, S. H., and Jinks-Robertson, S. (2011) Role for topoisomerase 1 in transcription-associated mutagenesis in yeast. *Proc. Natl. Acad. Sci. U.S.A. 108*, 698–703.

(30) Niu, D.-K., and Yang, Y.-F. (2011) Why eukaryotic cells use introns to enhance gene expression: Splicing reduces transcription-associated mutagenesis by inhibiting topoisomerase I cutting activity. *Biol. Direct 6*, 24.

(31) Kowalska-Loth, B., Girstun, A., Trzcińska-Daneluti, A. M., Piekielko-Witkowska, A., and Staron, K. (2005) SF2/ASF protein binds to the cap region of human topoisomerase I through two RRM domains. *Biochem. Biophys. Res. Commun. 331*, 398–403.

(32) Malanga, M., Czubaty, A., Girstun, A., Staron, K., and Althaus, F. R. (2008) Poly(ADP-ribose) binds to the splicing factor ASF/SF2 and regulates its phosphorylation by DNA topoisomerase I. *J. Biol. Chem.* *283*, 19991−19998.

(33) Zhang, Y. (2007) Template-based modeling and free modeling by I-TASSER in CASP7. *Proteins 69* (Suppl.8), 108−117.

(34) Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction. *BMC Bioinf. 9*, 40.

(35) Wu, S., Skolnick, J., and Zhang, Y. (2007) Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biol. 5*, 17.

(36) Kalé, L., Skeel, R., Bhandarkar, M., Brunner, R., Gursoy, A., Krawetz, N., Phillips, J., Shinozaki, A., Varadarajan, K., and Schulten, K. (1999) NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys. 151*, 283−312.

(37) Xu, R. M., Jokhan, L., Cheng, X., Mayeda, A., and Krainer, A. R. (1997) Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs. *Structure 5*, 559−570.

(38) Laco, G. S., and Pommier, Y. (2008) Role of a tryptophan anchor in human topoisomerase I structure, function and inhibition. *Biochem. J. 411*, 523−530.

(39) Fiorani, P., Tesauro, C., Mancini, G., Chillemi, G., D'Annessa, I., Graziani, G., Tentori, L., Muzi, A., and Desideri, A. (2009) Evidence of the crucial role of the linker domain on the catalytic activity of human topoisomerase I by experimental and simulative characterization of the Lys681Ala mutant. *Nucleic Acids Res. 37*, 6849−6858.

(40) Wriggers, W., Chakravarty, S., and Jennings, P. A. (2005) Control of protein functional dynamics by peptide linkers. *Biopolymers 80*, 736−746.

(41) Nadler, S. G., Merrill, B. M., Roberts, W. J., Keating, K. M., Lisbin, M. J., Barnett, S. F., Wilson, S. H., and Williams, K. R. (1991) Interactions of the A1 heterogeneous nuclear ribonucleoprotein and its proteolytic derivative, UP1, with RNA and DNA: Evidence for multiple RNA binding domains and salt-dependent binding mode transitions. *Biochemistry 30*, 2968−2976.

(42) Chillemi, G., Redinbo, M., Bruselles, A., and Desideri, A. (2004) Role of the linker domain and the 203-214 N-terminal residues in the human topoisomerase I DNA complex dynamics. *Biophys. J. 87*, 4087−4097.

(43) Fiorani, P., Bruselles, A., Falconi, M., Chillemi, G., Desideri, A., and Benedetti, P. (2003) Single mutation in the linker domain confers protein flexibility and camptothecin resistance to human topoisomerase I. *J. Biol. Chem. 278*, 43268−43275.

(44) Szklarczyk, O., Staroń, K., and Cieplak, M. (2009) Native state dynamics and mechanical properties of human topoisomerase I within a structure-based coarse-grained model. *Proteins 77*, 420−431.

(45) Lieser, S. A., Aubol, B. E., Wong, L., Jennings, P. A., and Adams, J. A. (2005) Coupling phosphoryl transfer and substrate interactions in protein kinases. *Biochim. Biophys. Acta 1754*, 191−199.

(46) Tintaru, A. M., Hautbergue, G. M., Hounslow, A. M., Hung, M.-L., Lian, L.-Y., Craven, C. J., and Wilson, S. A. (2007) Structural and functional analysis of RNA and TAP binding to SF2/ASF. *EMBO Rep. 8*, 756−762.

(47) Sinha, R., Allemand, E., Zhang, Z., Karni, R., Myers, M. P., and Krainer, A. R. (2010) Arginine methylation controls the subcellular localization and functions of the oncoprotein splicing factor SF2/ASF. *Mol. Cell. Biol. 30*, 2762−2774.