Phosphorylation tunes electrostatically driven protein-RNA interactions

Marie Synakewicz,[1,*] Adhvitha Premanand,[2,1] Héloïse Bürgisser,[2] Sarah Habeler,[1] Lucia R. Franchini,[1] Noémie Kociolek,[3] Mark Nüesch,[1] Antoine Cléry,[3] Daniel Nettels,[1] Frédéric Allain,[3] Nina Hartrampf,[2,*] and Benjamin Schuler[1,4,*]

[1] *Department of Biochemistry, University of Zurich, Switzerland*
[2] *Department of Chemistry, University of Zurich, Switzerland*
[3] *Department of Biology, Institute of Biochemistry, ETH Zurich, Switzerland*
[4] *Department of Physics, University of Zurich, Switzerland*

(Dated: April 27, 2026)

Abstract

Phosphorylation of intrinsically disordered proteins (IDPs) is essential for regulating biomolecular interactions in many cellular processes. However, a quantitative understanding of how phosphorylation tunes the affinity between highly charged IDPs and nucleic acids is lacking. Here, we show that multi-site phosphorylation of the disordered arginine/serine-rich (RS) domain of the splicing factor SRSF1 acts as an electrostatic rheostat that governs RNA binding. By combining enzymatic phosphorylation, phosphomimetic variants, and chemically synthesised phosphopeptides with single-molecule Förster resonance energy transfer measurements, we reveal the RS domain to be a potent driver of protein–RNA association. Increasing phosphorylation progressively reduces this interaction, and extensive phosphorylation eliminates detectable RNA binding. Remarkably, the binding free energy depends linearly on RS-domain net charge, regardless of whether the charge arises from phosphorylation or acidic residues introduced as phosphomimetics. Together, our findings uncover a quantitative framework for how phosphorylation tunes the interactions of charged IDPs and rationalize why two acidic residues are required to mimic a single phosphorylation event.

Keywords. RNA binding | SR protein | phosphorylation | phosphomimetic | automated fast-flow peptide synthesis (AFPS) | single-molecule spectroscopy | Förster resonance energy transfer (FRET)

* To whom correspondence should be addressed: m.synakewicz@bioc.uzh.ch, nina.hartrampf@chem.uzh.ch, schuler@bioc.uzh.ch

INTRODUCTION

Cellular communication and homeostasis are often regulated by post-translational modifications [1]. An important example is phosphorylation, the effects of which range from the modulation of enzymatic activity through conformational changes triggered by highly specific single-site phosphorylation to an overall influence on the electrostatic interactions between charged biomolecules. Examples include the binding interactions between proteins, especially for intrinsically disordered proteins (IDPs)[2, 3], which are particularly prone to post-translational modification[4]. Many IDPs are highly charged, especially those involved in binding nucleic acids[5], and we thus expect an important role of phosphorylation on such interactions[6]. However, a quantitative understanding of the influence of phosphorylation on the affinity between IDPs and nucleic acids is currently lacking.

Here, we address this question for an important example of an RNA-binding protein, the highly positively charged serine/arginine-rich splicing factor 1 (SRSF1). As a member of the Ser-Arg (SR) protein family, it regulates exon inclusion in the constitutive and alternative splicing of pre-messenger RNAs (pre-mRNAs), as well as other crucial biological functions, such as nuclear export of RNAs, translation, mRNA homeostasis, and mRNA decay [7–9]. Notably, the function and sub-cellular localization of SRSF1 are regulated by massive phosphorylation of its RS domain, a disordered Arg-Ser-rich region. Previous work has qualitatively shown an influence of phosphorylation on RNA binding [10]. To be able to quantitatively investigate the role of phosphorylation and charge on the interactions between SRSF1 — and especially the RS domain — and RNA, we use single-molecule Förster resonance energy transfer (FRET) spectroscopy combined with enzymatic phosphorylation, site-specific phosphoserine incorporation by chemical synthesis, and glutamate-based phosphomimetics. By varying the phosphorylation state and the net charge over a broad range, the affinity of the RS domain for RNA can be tuned across several orders of magnitude, ranging from nanomolar affinity to complete loss of binding. Phosphorylation of the RS domain is thus a major determinant of the interactions between SRSF1 and RNA, and its net charge is related roughly linearly with the binding free energy.

RESULTS

SRSF1 comprises two structured RNA recognition motifs (RRMs) and a C-terminal intrinsically disordered RS domain (Figure 1a) [11–16]. The RRMs interact specifically with the bases of the target sequences in single-stranded RNA (ssRNA), and the positively charged RS domain is thought to interact largely non-specifically with RNA [15–18]. The RS domain contains 19 Arg residues and 20 Ser residues, and all of the latter can be phosphorylated [19–22]. To examine the role of phosphorylation in the SRSF1-RNA interactions, we first dissected the contributions of the RRMs and the RS domain to RNA binding by producing different SRSF1 constructs and measuring their affinity to RNA. In a second step, we used enzymatic phosphorylation of recombinantly expressed RS domain as well as the incorporation of site-specific phosphoserine residues by automated fast-flow peptide synthesis (AFPS) to assess the RNA affinity as a function of the number and position of phosphate groups. We then compared their RNA affinity with those of recombinantly expressed phosphomimetic variants. Throughout, single-molecule FRET experiments enabled us to probe biomolecular interactions at very low protein and RNA concentrations, which minimizes many of the notorious complications arising from protein aggregation and (micro-)phase separation in systems such as SRSF1 [23, 24].

We thus recombinantly expressed full-length SRSF1 (containing the Y37S and Y72S substitutions to increase solubility without affecting RNA binding [16]); a construct containing only the N-terminal RRMs (here termed RRM12); and the RS domain (Figure 1a). Since neither of the RRMs interacts specifically with uracil, we designed ssRNA sequences of different lengths that contain only one target sequence for each RRM, CA for RRM1, and GGA for RRM2 [15, 16], flanked by stretches of uracil ($rU9^{CA/GGA}$, $rU33^{CA/GGA}$ and $rU59^{CA/GGA}$, see Figure 1b and Table S1). To distinguish sequence- and nucleobase-specific from purely electrostatically driven interactions, we furthermore included homopolymeric ssRNAs without these binding motifs (rA19, rC19, rG19, rU19 and rU33, see Table S1) in our measurements. We site-specifically labeled the 5' and 3' ends of the RNAs with acceptor and donor fluorophores, respectively, to study the protein-RNA interactions using single-molecule FRET. This experimental design allows us to obtain binding affinities at the same time as information on intramolecular distances and distance changes in the ssRNA conformational ensemble.

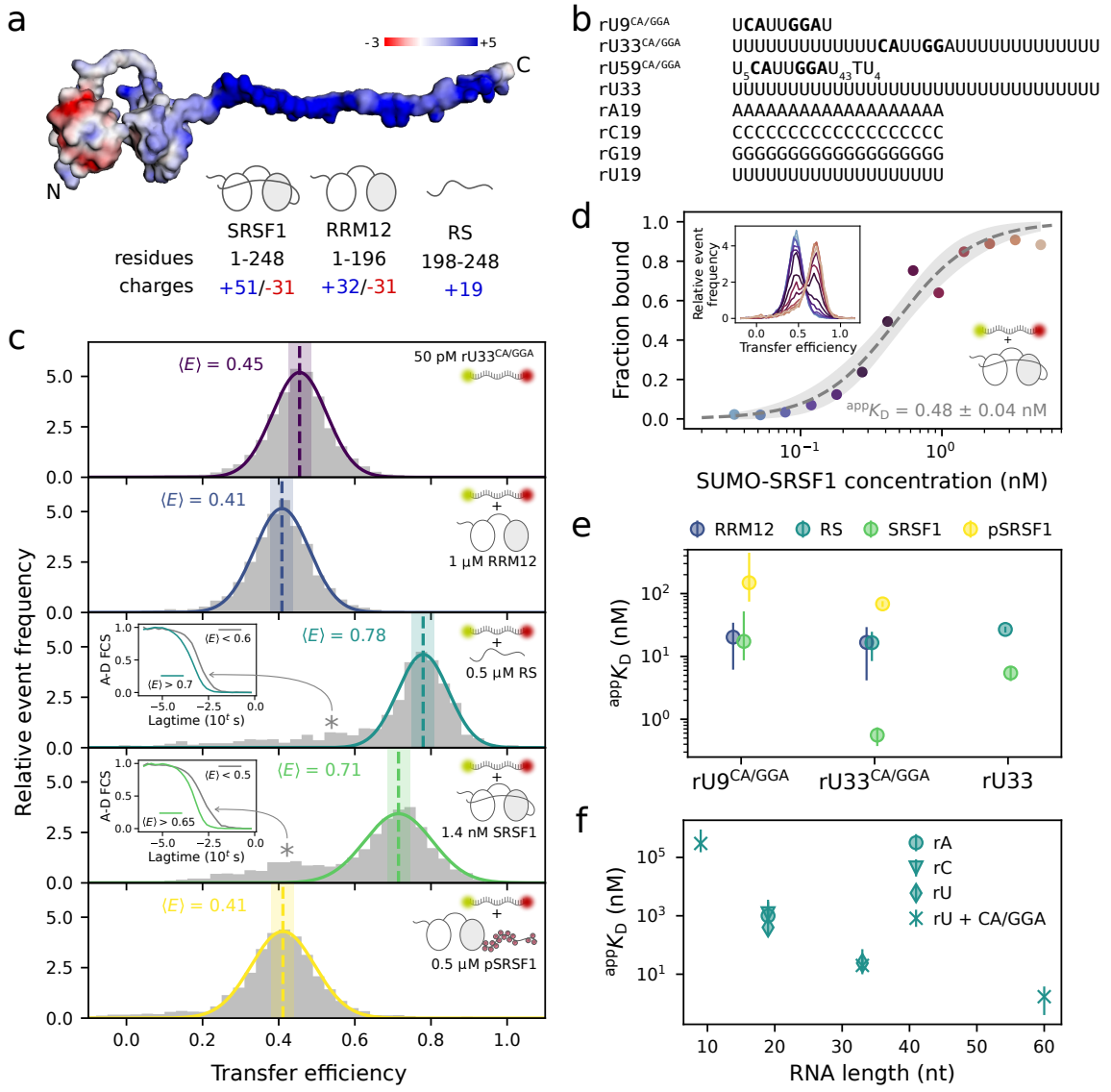

FIGURE 1. Quantifying SRSF1-RNA interactions with single-molecule spectroscopy. (a) Model of SRSF1 based on a structure generated by AlphaFold[25] (Q07955-F1), with the electrostatic surface potential colored in units of k_BT per unit charge[26]. For each protein construct indicated by the cartoons, the residue boundaries and the total number of positive and negatively charged amino acids are given. (b) Nucleotide sequences ($5' \rightarrow 3'$) of single-stranded RNAs used to probe protein-RNA interactions. The binding motifs for RRM1 (CA) and RRM2 (GGA) are highlighted in bold. Terminal and internal modifications used for labeling are omitted for clarity but are shown in Table S1. (c) Representative normalised transfer efficiency histograms showing how protein binding affects the compaction of RNA. The mean of the Gaussian peak fit is indicated by a dashed line, the shaded band represents an uncertainty of 0.03 [27]. Asterisks indicate populations arising from large oligomeric species or aggregates, as highlighted by subpopulation-specific FCS in the corresponding insets: the diffusion time of the 1:1 bound population (solid lines in respective color) is of the same timescale as that of the free RNA (see Figure S1), whereas the diffusion times of the large oligomers or aggregates are much longer (solid lines in gray). (d) Representative titration of donor/acceptor-labeled $\mathrm{rU33^{CA/GGA}}$ with SUMO-SRSF1. The inset shows an overlay of the normalised transfer efficiency histograms of individual measurements, color-coded according to protein concentration. The fraction bound is fitted with Equation (2) (gray dashed line with shaded area corresponding to 95% confidence intervals) to obtain the apparent equilibrium dissociation constant, ${}^{\mathrm{app}}K_{\mathrm{D}}$ (fit value and error of the fit shown). For more details on data analysis, see Figure S2 and Methods. (e) Comparison of ${}^{\mathrm{app}}K_{\mathrm{D}}$ for the different protein variants and RNAs. Binding of RRM12 to rU33 could not be detected at RRM12 concentrations of up to 5 µM. The onset of the interaction of the RS domain with $\mathrm{rU9^{CA/GGA}}$ could be observed, but only a rough estimate of the affinity could be obtained (Figure S4). Means represent values obtained for both SUMO-tagged and untagged proteins, error bars represent the standard deviation ($N > 1$ measurements) or an uncertainty of a factor of 2 ($N = 1$ measurements)[28]. See Table S2 for details of individual protein constructs. (f) ${}^{\mathrm{app}}K_{\mathrm{D}}$ of the untagged RS domain binding to ssRNA as a function of RNA length. The value for 9 nt is an estimate based on the binding response observed up to 30 µM RS domain (Figure S4). Error bars represent standard deviations ($N > 1$ measurements), an uncertainty of a factor of 2 ($N = 1$ measurements)[28], or a value derived from bootstrapping ($\mathrm{rU59^{CA/GGA}}$). See Table S2 and Methods for details. All affinities were measured at an ionic strength of 0.2 M.

RRMs and RS domain display different binding characteristics but similar affinities

To quantify the affinity of the different SRSF1 constructs to RNA and to elucidate their mode of interaction, we employ the changes in FRET efficiency that occur upon binding of the labeled RNA. Remarkably, RRM12 and the RS domain have very different effects on the RNA. Binding of the RRM12 construct to the rU33$^{\mathrm{CA/GGA}}$ RNA causes a shift in mean FRET efficiency from $\langle E \rangle \approx 0.46$ to $\langle E \rangle \approx 0.41$, corresponding to a slight extension of the RNA upon binding (Figure 1c). This observation suggests that binding of the two folded domains reduces the flexibility of the RNA segment containing the target sequence, resulting in an increase of the average end-to-end distance of the oligonucleotide chain. In contrast, binding of the RS domain causes a shift from $\langle E \rangle \approx 0.46$ to $\langle E \rangle \approx 0.78$, indicating a pronounced compaction of the RNA (Figure 1c). Similar to previous observations on other systems [29, 30], it is the flexibility of both the disordered RS domain and the ssRNA that enables this compaction upon mutual charge screening between the two oppositely charged polyelectrolytes. In the binding of full-length SRSF1, the combined effects of the RRMs and the RS domain cause a shift from $\langle E \rangle \approx 0.46$ to $\langle E \rangle \approx 0.71$ (Figure 1c). Presumably, binding of the RRMs constrains the conformational ensemble of RS domain-RNA interactions within the context of the full-length protein, leading to a slightly less pronounced overall compaction than for the isolated RS domain. Altogether, these results illustrate the different effects of SRSF1 and its component domains on its target RNA.

In a next step, the changes in FRET efficiency we observe can be employed to measure binding affinities. It is worth emphasizing that for a system such as SRSF1, which is prone to phase separation and aggregation even at relatively low concentrations [23, 24, 31], single-molecule spectroscopy provides a powerful way of identifying the presence of larger species in solution that interfere with the measurements. Moreover, subpopulation-specific FRET and fluorescence correlation spectroscopy (FCS) enable us to restrict the analysis to the species of interest (Figure S1), thus minimizing contributions from high-molecular-weight oligomers and aggregates that would interfere with the signal in ensemble measurements[32]. Based on these findings, we established an analysis that allowed us to to quantify RNA binding for all the constructs under study (Figure S2). Nevertheless, the low solubility of full-length SRSF1 required us to use a SUMO-tagged variant to obtain full titration curves (Figure 1d). However, a systematic comparison shows that the tag has little effect on RNA affinity (Figure S3).

To obtain information on the affinity and specificity of the binding of SRSF1 and its domains to RNA, we titrated different lengths and sequences of donor/acceptor-labeled ssRNAs with RRM12, the RS domain, and full-length SRSF1 (Figure 1e, Table S2). The apparent dissociation constant, $^{\mathrm{app}}K_{\mathrm{D}}$, of RRM12 to RNA containing the binding motifs was $20 \pm 10\,\mathrm{nM}$ for both rU9$^{\mathrm{CA/GGA}}$ and rU33$^{\mathrm{CA/GGA}}$ (Figure 1e). We could not detect binding of RRM12 to rU33, which lacks the binding motif, in the accessible protein concentration range of up to $5\,\mu\mathrm{M}$, suggesting that any non-sequence-specific binding is at least three orders of magnitude weaker than specific binding. In contrast, the RS domain lacks sequence specificity and binds RNAs with and without the RRM binding motifs with similarly high affinity of $17 \pm 8\,\mathrm{nM}$ (rU33$^{\mathrm{CA/GGA}}$) and $27 \pm 3\,\mathrm{nM}$ (rU33), respectively (Figure 1d). For the interaction between the RS domain and rU9$^{\mathrm{CA/GGA}}$, we can only estimate $^{\mathrm{app}}K_{\mathrm{D}}$ to be $\sim$300 µM, since the formation of large oligomers or aggregates at higher concentrations prevented measurements of a complete titration curve (Figure S4). Furthermore, the RS domain-RNA affinity increases by several orders of magnitude as the length of the RNA is increased from 9 to 60 nucleotides (Figure 1f), consistent with previous studies on protein-nucleic acid interactions that show higher affinity with greater length and net charge of the nucleic acid due to the larger number of possible binding configurations [33, 34]. Using a series of homopolymeric single-stranded RNAs, rA$_{19}$, rC$_{19}$, and rU$_{19}$ [35], we find that the affinity of the RS domain is similar for all nucleobases (see Figure S5 for more detail) – especially, when compared with the strong effect of RNA length (Figure 1e). It was not feasible to quantify the affinity to rG$_{19}$ owing to pronounced aggregation in the presence of RS domain, which occurred over the same range of protein concentrations where the other homopolymeric RNAs showed binding (Figure S5). Altogether, we have no indications that the RS domain-RNA interactions are specific for a particular nucleobase; rather, they are dominated by charge interactions.

In full-length SRSF1, the combined sequence-specific interactions of the folded RRMs for the binding motif and the electrostatic interactions of the disordered RS domain with RNA increase the total affinity to rU33$^{\mathrm{CA/GGA}}$ by almost two orders of magnitude compared to the RRMs alone, to $^{\mathrm{app}}K_{\mathrm{D}} = 0.6 \pm 0.2\,\mathrm{nM}$ (Figure 1e). On the much shorter rU9$^{\mathrm{CA/GGA}}$ RNA, full-length SRSF1 exhibits

a similar affinity as the RRM12 construct, suggesting that interactions of the RS domain are blocked by the footprint of the RRMs. Likewise, if the RRM binding motifs are removed, as in the rU33 RNA, SRSF1 exhibits an affinity closer to that of the isolated RS domain (Figure 1e). Together, these observations suggest that the binding of SRSF1 to RNA results both from sequence-specific interaction of the RRMs with the target sequences and from non-sequence-specific interaction of the disordered RS domain driven by its charges opposite to those of the phosphate groups on the RNA.

Enzymatic phosphorylation by SRPK1 abolishes RS domain-RNA binding

Depending on the cellular localization and function of SRSF1, the RS domain can be heavily phosphorylated [36], raising the question of how these modifications affect RNA affinity. If the interaction between the RS domain and RNA is predominantly charge-driven, introducing negative charges by phosphorylation should weaken binding. In the cell, SRSF1 is phosphorylated by two kinases: in a first step, the N-terminal serines of the RS domain are phosphorylated processively by SRPK1, starting from the centre of the RS domain [37, 38]. After transport into the nucleus, SRSF1 is fully phosphorylated by Clk1[20–22]. Here, we used a three-pronged approach for assessing how different degrees of phosphorylation affect RNA binding, by combining enzymatic phosphorylation, chemical synthesis, and phosphomimetics. Enzymatically, we can introduce a large number of phosphates; by chemical synthesis, we can introduce a few phosphates at specific locations; and phosphomimetic amino acid substitutions allow us to introduce acidic amino acids at arbitrary positions and assess their effect compared to phosphorylated residues.

Using SRPK1, we were able to obtain phosphorylated SRSF1 and RS domain with an average of 16 added phosphates (Figure S6), which we refer to as pSRSF1 and RS-16xpS, respectively. With a charge of -2 per phosphate at pH 8 based on the relevant phosphoserine side chain pK_a of ~ 5.6[39], phosphorylation thus inverts the net charge of SRSF1 and the RS domain, resulting in values of -12 and -13 for pSRSF1 and RS-16xpS, respectively. Owing to the repetitive nature of the RS domain, we were not able to identify the exact positions of the modifications, but given the processivity of SRPK1 [37, 38], the first 13 and final three serine residues are most likely to be phosphorylated. Transfer efficiency histograms of $rU33^{CA/GGA}$ RNA bound to pSRSF1 no longer indicate compaction but instead a slight expansion, similar to the behavior observed with RRM12 (Figure 1c). We did not observe binding of RS-16xpS to $rU33^{CA/GGA}$ RNA up to protein concentrations of 30 µM (Figure S7), indicating that SRPK1-mediated phosphorylation of the RS domain abolishes its affinity to RNA. However, we found a slightly but reproducibly lower affinity of pSRSF1 for RNAs containing the binding motif compared to RRM12 (Figure 1e). Since the affinities of pSRSF1 and RRM12 are similar for $rU9^{CA/GGA}$ and $rU33^{CA/GGA}$, electrostatic repulsion between RNA and the phosphorylated RS domain is unlikely to be the source of this decrease in affinity. Recently, Zhang and co-workers reported a similar trend and suggested that the phosphorylated RS domain interacts with arginine residues close to the RNA binding interfaces of the RRMs and in the inter-RRM linker, thereby reducing the overall affinity to RNA [40].

Titrating the net charge of the RS domain

Enzymatic phosphorylation shows that charge inversion of the RS domain prevents it from binding to RNA. Addressing the effect of phosphorylation on the RS domain affinity for RNA more quantitatively thus requires us to introduce a smaller and well-defined number of phosphate groups. Since enzymatic modification does not provide the necessary control, we leveraged automated fast-flow peptide synthesis (AFPS) to introduce site-specific phosphoserine residues [41–43] . A recently optimized method for the reliable incorporation of phosphorylated amino acids using AFPS [44] allowed us to synthesise RS domain variants containing one (RS-pS_{199}, RS-pS_{223}, RS-pS_{225}), two (RS-$pS_{223,225}$ termed RS-2×pS) or four phosphoserines (RS-$pS_{199,209,217,225}$ termed RS-4×pS, see Figure 2a and Table S3). Due to synthetic challenges of incorporating two phosphoserines in close proximity [45], we did not follow the SRPK1 phosphorylation pattern from the centre to the N-terminus of the RS domain but spaced them more widely in the RS-4×pS construct. Furthermore, as commonly observed with incorporating multiple phosphorylated residues, increasing side reactions led to an overall reduction in yield [45], preventing us from synthesising RS domain peptides with more than 4 phosphoserines.

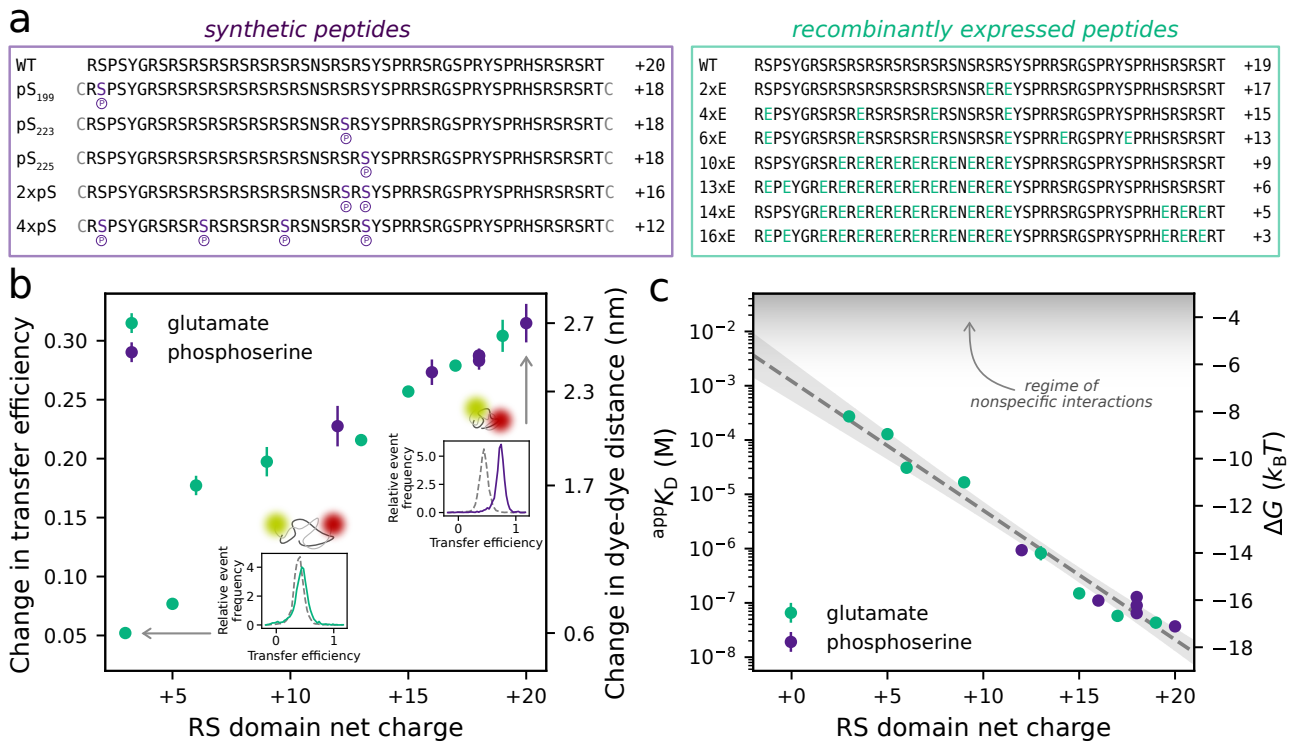

FIGURE 2. Changes in net charge fine-tune RS domain-RNA interactions. (a) Overview of peptide sequences and net charges. Altered amino acids are highlighted in purple (phosphoserine) and green (glutamate). Phosphopeptides were synthesized with terminal cysteines, which were found not to affect RNA-RS domain interactions (Figure S8). Note that synthetic peptides contain a C-terminal amide group (uncharged at pH 8) instead of a carboxyl group (negatively charged at pH 8). (b) Compaction of rU33$^{CA/GGA}$ RNA upon RS domain binding as a function of RS domain net charge (in units of the elementary charge e), plotted both in terms of the measured change in transfer efficiency (left axis) and as a change in donor-acceptor dye distance of the RNA (right axis). The data points show the average and standard deviation of at least triplicates; error bars are those of the change in transfer efficiency. Insets show normalised transfer efficiency histograms of unbound (gray dashed) and saturated RNA (purple/green) to illustrate the degree of compaction at either extreme of RS domain net charge. For representative histograms of RNA bound to each RS domain construct, see Figure S10. (c) Apparent equilibrium dissociation constant of the RS domain-rU33$^{CA/GGA}$ complex and corresponding free energy of binding as a function of RS domain net charge. The dashed line shows a fit of Equation (1) to the data, using $\varepsilon_r = 80$, $T = 295$ K, $q_{RNA} = -34$ e (taking into account the charges of both RNA and dyes) and an ionic strength of 0.2 M, yielding $\langle r \rangle = 2.08 \pm 0.02$ nm and $\Delta G_0 = -6.7 \pm 0.4$ k_BT. The shaded band around the fit line corresponds to 95% confidence intervals. The area shaded in gray at the top highlights the affinity regime where interactions can occur due to nonspecific stickiness [54]. All data shown represent the average and standard deviation of at least triplicates. In most cases, the error bars are smaller than the symbols.

We thus complemented our set of synthetic phosphopeptides with a series of bacterially expressed RS domain variants in which serines were replaced with glutamate residues. This 'phosphomimetic' approach is commonly used to imitate posttranslational phosphorylation in cell biology and biochemistry [see 46, and references therein] and gives us the opportunity to obtain peptides across a wide range of net charge (Figure 2a). Phosphomimetic variants can obviously not be expected to be fully equivalent to phosphorylated protein [47–53], but the availability of RS domain phosphopeptides from AFPS and phosphomimetic variants across the same range of net charge (+12 to +20) allows us to quantitatively compare their affinities to RNA and address the question of whether glutamate-based phosphomimetics are suitable for imitating the effect of phosphorylation on IDP-RNA binding.

To quantify how the affinity to RNA is affected by a change in the net charge of the RS domain, we titrated donor/acceptor-labeled rU33$^{CA/GGA}$ RNA with our thirteen different synthetic phosphopeptides and glutamate phosphomimetic RS domain variants. Since peptides bearing two and four phosphoserines were prone to aggregation, we performed all titrations in the presence of 1 M urea, an uncharged denaturant that is expected to have only minor effects on charge-dominated interactions. Indeed, compared to 0 M urea, the change in transfer efficiency between unbound and bound populations at 1 M urea differed by <10%, and values of $^{app}K_D$ changed only by up to a factor of ~2 (i.e.

$<1\,k_BT$ in binding free energy, Figure S9 and Table S2). From the resulting full titration curves for all peptides, we obtained the change in transfer efficiency upon binding and ${}^{app}K_D$ as a function of RS domain net charge (Figure 2b,c). The change in transfer efficiency shows that the end-to-end distance of the RNA decreases less and less upon binding as the positive net charge of the RS domain is reduced by the introduction of negatively charged residues (Figure 2b, Figure S10). A deviation from the trend is observed for RS-10×E and RS-13×E, possibly due to their block-copolymeric character. Concomitant with the decrease in compaction of the complex, the affinity between RNA and RS-domain decreases from ${}^{app}K_D \approx 40\,\text{nM}$ for the WT sequence to $270 \pm 30\,\mu\text{M}$ for RS-16×E. Notably, the free energy of binding inferred from ${}^{app}K_D$ scales roughly linearly with net charge, and the charge distribution across the sequence appears to only play a secondary role for the affinity. This contribution of charge location is best illustrated by the data for the three different single-phosphoserine peptides (Table S2), which show that affinity can vary slightly depending on the position of the phosphorylation site, making charge patterning one likely source of the residual deviations from the linear trend.

If we follow a very simple approach and assume the binding free energy, ΔG, to consist of a charge-independent contribution, ΔG_0, and an electrostatic term described by a screened Coulomb interaction, essentially a mean-field description of the dynamically distributed charges in the two disordered binding partners, we can approximate the charge dependence of ΔG as [2]

$$\Delta G \approx k_B T \ln({}^{app}K_D) \approx \Delta G_0 + \frac{q_{RS} q_{RNA} e^{-\langle r\rangle/\lambda_D}}{4\pi\epsilon_0\epsilon_r \langle r\rangle}, \tag{1}$$

where q_{RS} is the net charge of the RS domain variant; q_{RNA} is the net charge of the RNA, separated from the RS domain by an effective average distance $\langle r\rangle$; λ_D is the Debye screening length at our ionic strength of 0.2 M; ϵ_0 is the permittivity of vacuum; and ε_r is the dielectric constant of water. Fitting this linear dependence on q_{RS} to our data (Figure 2c), we obtain $\Delta G_0 = -6.7 \pm 0.4\ k_BT$ for the charge-independent contribution, and an average change in ΔG of $-0.55 \pm 0.03\ k_BT$ per single charge at near-physiological ionic strength. Extrapolated to a net charge of -13, the fit result suggests that an RS domain bearing 16 phosphorylations would exhibit an affinity for the $rU33^{CA/GGA}$ RNA in the molar range, corresponding to the loss of binding we observed experimentally. An RS domain with a net charge of zero would correspond to millimolar affinity, indicating that the non-electrostatic contribution to binding is small. Last but not least, the linear dependence of ΔG on q_{RS} and the slope we observe is independent of whether the charge is altered by phosphorylation or by introducing acidic residues, implying that the charge of two glutamate residues is required to mimic one phosphoserine. On the one hand, it is thus obvious that a single acidic amino acid is not appropriate for imitating a phosphate group, but on the other hand, introducing two glutamates, each bearing a charge of -1 at pH 8, is remarkably accurate in accounting for a phosphate group in the charge-dominated interaction between the RS domain and RNA.

DISCUSSION

SR proteins such as SRSF1 exemplify extreme cases of post-translational modification in which extensive phosphorylation inverts the net charge of regions highly enriched in positively charged residues. We find the electrostatic interactions of the disordered RS repeat domain to be the main driver of non-sequence-specific protein-RNA interactions, which can result in similar or even higher affinity than from the sequence-specific binding of folded RRMs. The binding affinity to RNA strongly depends on the phosphorylation state of the RS domain: it decreases as the positive net charge is reduced, and the variant heavily phosphorylated by SRPK1 exhibits no RNA affinity. We observe a remarkable linearity of the binding free energy on the net charge of the RS domain, and the change in free energy of $\sim 0.5\ k_BT$ per charge unit is surprisingly independent of whether the net charge is altered by phosphorylation or by introducing acidic amino acid residues. This finding implies that the use of glutamate as phosphomimetic residues, which is common practice in molecular biology [46], can be a valid approach for imitating the change in net charge for this type of protein-RNA interactions if two acidic residues are used to mimic the effect of one phosphate group.

Our findings highlight several practical considerations that can be applied to a variety of systems. First, mimicking one phosphorylation by two glutamates is a good rule of thumb for electrostatically driven interactions at a pH sufficiently far from the respective pK_a values of the phosphorylated and

phosphomimetic amino acids. Although phosphomimetics fail to capture the charge density and hydration properties of phosphorylated amino acids [55], they remain an attractive choice when site-specific phosphorylation cannot be achieved. However, whether or not such double-phosphomimics accurately imitate the phosphorylated protein will depend on the specific system and its interaction partners [47, 56]. Second, in cases where only a single-residue substitution is feasible for each phosphorylation site, the interpretation of the data has to account for effects from the difference in net charge. For example, depending on the context, phosphomimetic variants of SRSF1 only partially reproduce the effects of binding to interaction partners and localization of the phosphorylated WT[10, 57].

The roughly linear dependence of the binding free energy on the RS domain net charge is in line with previous results on electrostatic interactions involving IDPs [2, 3, 58], even though the systems under study differ in several respects: protein-protein *vs.* protein-nucleic acid interactions; interactions between one structured and one disordered binding partner *vs.* two disordered binding partners; differences in net charge, charge clustering and length of the IDPs; and the number of negative charges added or positive charges removed. Additionally, the measurement conditions in these studies differed in pH ($7 < \mathrm{pH} < 8$) and ionic strength ($0.06\mathrm{M} < \mathrm{IS} < 0.2\mathrm{M}$). Hence, it is remarkable that, despite these variations, the change in binding free energy per unit charge is consistently approximated by a linear dependence on the net charge of the ligand. Collectively, these findings indicate that, for electrostatically driven systems, phosphorylation, and charge content in general, play a central role in tuning binding affinity, with the overall net charge exerting a stronger influence than the specific positioning or local environment of individual charges. As such, similar principles of charge interactions are expected to be important for the majority of protein-RNA and protein-protein interactions that are regulated by SRSF1 phosphorylation, e.g. the inter- and intramolecular interactions between RS domain and RRMs [59–62], binding to other nuclear proteins involved in splicing regulation [36], and SRSF1 phase separation [23, 24, 40].

Splicing is a tightly regulated and dynamic process involving the controlled binding and dissociation of many proteins and ribonucleoprotein complexes, many of which compete for overlapping or neighboring RNA sequences [63]. SRSF1 is fully phosphorylated when localizing to active transcription sites, where it recognizes splicing enhancer sequences within nascent pre-mRNA transcripts [64, 65]. If SRSF1 were not fully phosphorylated, RS domains lacking or containing only a small number of phosphorylations could lead to non-sequence-specific binding of SRSF1 outside of enhancer motifs and non-productive compaction of RNA. Both effects could physically limit access of other splicing factors to their cognate binding motifs. That is, in the context of splice-site recognition and initial spliceosome recruitment, phosphorylation appears to function as a regulatory switch that reduces SRSF1-RNA interactions and promotes protein-protein interactions [14]. A quantitative understanding of the role of phosphorylation in RS domains and highly charged IDPs in general is thus essential for elucidating the resulting regulatory processes.

METHODS

Preparation and labeling of RNA

RNAs modified with a 5' di-thiol group and a 3' or internal amino group (Table S1) were purchased from Integrated DNA Technologies (IDT). RNAs were dissolved to ∼100 µM in 100 mM NaP pH 7 or RNAse-free water. TCEP was added to a final concentration of 5 to 10 mM to reduce the dithiol. After incubation at room temperature for ∼30 min, all RNAs but G19 were buffer exchanged into 100 mM NaP pH 7 by centrifugal filtration (Amicon Ultra-0.5 MWCO 3 kDa). G19 was buffer-exchanged into 20 mM NaP pH 7, 6.4 M GdnHCl. An approximately 10-fold excess of CF660R-maleimide (Sigma-Aldrich), dissolved in a total volume of 5 µL DMSO, was added to the RNA solution and incubated on ice overnight or at room temperature for 1 h. Subsequently, the pH was raised to 8 by addition of 0.5 M NaP pH 8 (all RNAs but G19) or 20 mM NaP pH 8.6, 6.4 M GdnHCl (G19), and a 10-fold excess of Cy3B-NHS (Cytiva or Lumiprobe) dissolved in 5 µL DMSO was added. The reaction was incubated further at 37 °C for 2 h. Unreacted dye was removed by buffer exchange using ZebaSpin desalting columns (ThermoFischer Scientific, MWCO 7 kDa), or by ethanol precipitation. The labeled RNAs were applied to a Reprosil Pur 200 C18-AQ column (4×250 mm, 5 µm, Dr. Maisch) equilibrated in 0.1 M acetic acid, 0.1 M triethylamine, 3.75% acetonitrile, and eluted with gradients of 0.25 to

1.5 %/min to achieve separation of labeled species for each RNA. Fractions containing donor/acceptor-labeled RNA were lyophilised and resuspended in RNase-free water, and stored at −80 °C.

Plasmids and molecular cloning

Sequences of full-length SRSF1 containing the solubilising mutations Y37S and Y72S in RRM1 [16], a construct containing only the RRMs including the same mutations (RRM12, amino acids 1-196 of SRSF1), and the RS domain (amino acids 198-248) were subcloned from a pET24b plasmid into a pSUMO plasmid (kind gift of B. B. Kragelund, University of Copenhagen, Denmark) using NEB Gibson Assembly Master Mix (NEB,E2611) according to manufacturers specifications (for primers, see Table S4). Glutamate variants were ordered as synthetic genes from Integrated DNA Technologies (IDT) and cloned into pSUMO, linearised using the pSUMO Fwd and Rev primers listed in Table S4 using the NEBuilder HiFi DNA Assembly Master Mix in a 5 µL volume, but otherwise according to the manufacturer's protocol. The plasmid bearing Ulp1 protease with a C-terminal His6-tag was a kind gift from the Kragelund Lab. GB1-His_6-SRPK1-His_6 in a pET24b plasmid was a kind gift of Guillaume Hautbergue (University of Sheffield). DNA sequences were confirmed by Sanger sequencing (Microsynth).

Preparation of recombinantly expressed proteins and peptides

For a detailed description of the expression and purification of each protein or peptide, please refer to the Supplementary Experimental Methods. Amino acid sequences are listed in Table S5. In brief, all proteins were expressed in *E. coli* BL21(DE3) derivatives, cell pellets were lysed mechanically and purified by immobilised metal ion affinity chromatography (IMAC). Ulp1 protease was dialysed into storage buffer containing 50% glycerol. SRSF1, RRM12 and RS domain variants were further purified by ion exchange (IEX) chromatography and, if necessary, size exclusion chromatography. The His_6-SUMO-tag was removed by cleaving with Ulp1 followed by further purification (reverse-IMAC or IEX chromatography). SRPK1 was further purified by IEX chromatography. All protein samples were flash-frozen in liquid N_2 and stored at −80 °C. Protein concentrations were determined by absorbance at 280 nm and converted to molar concentrations using the respective theoretical extinction coefficients. Protein masses were confirmed by electrospray ionisation mass spectrometry (ESI-MS) at the Functional Genomics Center Zurich.

In vitro phosphorylation of SRSF1 and the RS domain

To obtain phosphorylated SRSF1 (pSRSF1), phosphorylation reactions were carried out with 2.5 to 5 µM His_6-SUMO-SRSF1 or SRSF1 and 1.7 to 5 µM GB1-His_6-SRPK1-His_6 in 50 mM Tris-HCl pH 8, 10 mM $MgCl_2$, 0.5 mM TCEP, 1 mM ATP. Reactions were incubated at 37 °C for ∼3 h. The phosphorylated protein was then purified by IEX chromatography using a MonoS 5/50 GL (GE Healthcare). Prior to each purification run, the column was incubated in 1 mg/mL pepsin dissolved in 0.5 M NaCl, 0.1 M acetic acid for at least 1 h at room temperature before cleaning with 2 M NaCl and 1 M NaOH according to the manufacturers protocol. The phosphorylation reaction was diluted to a final NaCl concentration of 100 to 150 mM prior to loading onto the column equilibrated in 20 mM HEPES pH 8, 150 mM NaCl, 5% glycerol, 0.01% Tween, 0.5 mM TCEP. The column was then washed with 10 CV of the equilibration buffer. pSRSF1 was collected from the flow through and the wash. SRPK1 and partially phosphorylated or unphosphorylated SRSF1 were recovered from the column using a gradient of 0.15 to 1.5 M NaCl. Flow-through and wash fractions containing pSRSF1 were pooled and concentrated using Amicon Ultra 0.5ml centrifugal filters (10 kDa MWCO, cellulose, Merck Millipore). The concentrated protein was then buffer-exchanged into 20 mM HEPES pH 8, 200 mM NaCl, 0.5 mM TCEP using ZebaSpin Desalting columns (7 MWCO, Thermo Scientific), frozen in liquid nitrogen and stored at −80 °C. The total number of phosphorylations was determined by ESI or MALDI mass spectrometry at the Functional Genomics Center Zurich.

To obtain phosphorylated RS domain, phosphorylation reactions were carried out with equimolar RS domain and SRPK1 (usually 5 to 6 µM of each) in 50 mM TrisHCl pH 8, 10 mM $MgCl_2$, 2 mM

DTT. Reactions were started by the addition of two-fold excess of ATP per serine residue (effectively 40-fold excess over protein concentration) and incubated at 37 °C for 30 min before adding GdnHCl to a final concentration of 4 M to stop the reactions. The mixture was applied to a 250×4.6 mm, 5 µm ReproSil Gold 200 C18 column (Dr. Maisch) equilibrated with H_2O, 0.1% TFA, 5% methanol. Proteins were eluted using a gradient of 23 to 30 % methanol over 30 min. Fractions containing the phosphorylated RS domain were combined and lyophilized. The protein was resuspended in 20 mM HEPES pH 8, 150 mM NaCl and stored at −80 °C. To remove Fe^{3+} adducts, samples were dialysed against 20 mM HEPES pH 8, 150 mM NaCl containing 50 µg/ml Chelex (Bio-Rad Laboratories) for 1 hour at room temperature. The total number of phosphorylations was determined by ESI mass spectrometry at the Functional Genomics Center Zurich.

Peptide synthesis and purification

For a detailed description of peptide synthesis and purification, please refer to the Supplementary Experimental Methods. Amino acid sequences of the synthetic peptides are listed in Table S6. In brief, peptides were synthesised on NovaPEG Rink Amide resin using an automated-flow system built in the Hartrampf lab [66]. Crude peptides were purified by semi-preparative reverse-phase HPLC (RP-HPLC). Both crude and pure peptides were analyzed by analytical ultra-high performance liquid chromatography (UHPLC) and liquid chromatography with high-resolution electrospray ionization mass spectrometry (LC-MS). All phosphopeptides were synthesised with terminal cysteines for future studies. To ensure that their addition did not affect the interaction with RNA, we synthesised the unphosphorylated wild-type peptide with and without terminal cysteines and compared them, especially regarding their affinity to RNA, using single-molecule FRET measurements (see Figure S8). An overview of purities and yields is given in Table S3. Figures containing UHPLC and LCMS data are listed in Appendix A.

Single-molecule FRET of freely diffusing molecules

Measurements of freely diffusing molecules were performed on a MicroTime 200 (PicoQuant) or a custom-built confocal single-molecule fluorescence instrument. On both instruments, lasers were operated in pulsed mode with interleaved excitation [67] at a pulse repetition rate of 20 MHz. The MicroTime 200 was equipped with an Olympus UplanApo 60x/1.20 W objective. The donor was excited by light from a supercontinuum laser (SuperK EVO HP SC-04, NKT Phototnics) filtered to 532/3 nm (BrightLine HC, Semrock) at 50 µW, as measured at the back aperture of the objective. The acceptor was excited by light from a 640 nm diode laser (LDH-D-C-640, PicoQuant) at 55 µW, as measured at the back aperture. Fluorescence emitted by the sample was collected by the same objective lens and separated from backscattered light by a triple-band mirror (ZT405/530/630RPC, Chroma Technology). Scattered light was further removed by a long-pass filter (LP532, Chroma Technology), before the emitted photons passed a 100 µm pinhole. The emitted photons were separated into four detection channels with a polarizing beam splitter and two dichroic mirrors (635DCXR, Chroma Technology). Donor and acceptor emission was further filtered with ET585/65M band-pass and LP647RU long-pass filters (both Chroma Technology), respectively. Photons were detected by four single-photon avalanche diode detectors (SPCM-AQR-15, PerkinElmer Optoelectronics). The arrival times of single photons were recorded with a HydraHarp 400 counting module (PicoQuant) with a resolution of 16 ps. The custom-built instrument was operated with a different donor excitation laser (EXW-12 SuperK Extreme, NKT Photonics), as described previously [68], with the wavelength selected by a 520/15 nm band-pass filter (BrightLine HC, Semrock) at 45 µW, as measured at the back aperture of the objective. Acceptor excitation and all other filters were as described for the MicroTime 200.

Apparent dissociation constants were determined using single-molecule titrations of freely diffusing molecules. Unless otherwise stated, for all RNAs but $rU59^{CA/GGA}$, titrations were performed using 50 pM labeled RNA in 20 mM HEPES pH 8, 200 mM mM NaCl, 0.01% Tween-20, 0.2 mg/mL BSA or recombinant HSA (both NEB), 10 mM DTT, 0.1 U/µL RNAse inhibitor (Applied Biosystems). Titrations with the respective peptides/proteins were performed in untreated µ-slides (Ibidi) by starting at saturating conditions and then sequentially diluting the protein concentration with buffer containing

only donor/acceptor-labeled RNA until the data obtained from the last titration point(s) were indistinguishable from a sample containing only RNA. To keep the ionic strength constant over the whole titration, the amount of NaCl added to the measurement buffer was adjusted to compensate for the amount of NaCl introduced by the protein storage buffer.

We found that the number of photon bursts detected per second decreased with increasing protein concentration (Figure S2a): while barely noticeable for RRM12, it was pronounced for the RS domain and full-length SRSF1 at high protein concentrations. Concurrently, we observed a new population at lower transfer efficiencies (asterisk in Figure 1b) originating from slower-diffusing species, as evident in the overall acceptor-donor fluorescence correlations (AD-FCS) (Figure S2b) and by subpopulation-specific AD-FCS (insets in Figure 1b and Figure S1), in line with the propensity of SRSF1 to phase-separate and aggregate, even at low concentrations [23, 24]. Although the presence of higher-molecular weight species can complicate the interpretation of the titrations, the advantage of our single-molecule approach is that we can identify and distinguish these species spectroscopically and adapt our analysis accordingly (see Figure S2 and data analysis methods below).

For SRSF1 without solubilizing SUMO tag, the burst loss was so large that full titration curves could not be obtained. Since we found the solubility of SRSF1, RRM12, and the RS domain to increase with a solubilisation tag, SUMO-tagged and cleaved proteins were compared in standard measurement conditions using the $rU33^{CA/GGA}$ RNA, and affinities were found to be within error (Figure S3). Importantly, we could obtain titrations to saturation with SUMO-SRSF1 (Figure 1c).

We found that the addition of urea also improved data quality by reducing burst loss at high protein concentrations. Binding of SUMO-RRM12, and the synthetic and expressed RS domains to the $rU33^{CA/GGA}$ RNA was tested at different urea concentrations up to 1 M (Figure S9). For the RS domain peptides, we found that the effect of urea on the binding affinity to RNA is negligible, and hence all measurements used to determine the relationship between of RS domain net charge and its affinity to RNA were performed in the presence of 1 M urea. In contrast, urea clearly affects specific interactions, as observed with the RRM12 construct (either due to partial unfolding of the domains, or due to interfering with base recognition).

Data analysis of binding titrations

Data analysis was carried out using the Mathematica (Wolfram Research) package Fretica (`https://github.com/SchulerLab`) and custom-written Python scripts [69–73]. Recorded photon counts were corrected for background photons, acceptor direct excitation, channel cross-talk, differences in detector efficiencies, and quantum yields of the dyes, as previously described [74]. Signal from large aggregates or oligomers was removed by binning photons in 1 s bins and identifying bursts with intensities higher than three standard deviations above the mean. Fluorescence bursts from individual molecules were identified by dual-channel burst search [75]. Successive photons separated by inter-photon times of less than 150 µs were combined and defined as a burst if the corrected total number of photons detected was more than 60 [76]. To select the molecules with active donor and acceptor, only bursts with stochiometry ratio, S, within the interval $0.2 < S < 0.7$ were retained.

Ratiometric transfer efficiencies were obtained for each burst from $E = n_A/(n_A+n_D)$, where n_D and n_A are the counts of donor and acceptor photons, respectively, corrected for direct excitation, channel crosstalk, and differences in fluorescence quantum yields and detection efficiencies. The transfer efficiencies of all bursts were binned into histograms. Fluorescence correlations of acceptor and donor photons were calculated as described previously [77]. Subpopulation-specific FCS was performed on selected bursts within a given range of transfer efficiency of the FRET population of interest.

For all data acquired in native buffer conditions to compare interactions between different RNAs and protein constructs, transfer efficiency histograms were globally fitted with Gaussian peak functions for the free and bound populations, respectively, keeping the mean transfer efficiency and the width of the respective Gaussian functions as shared parameters. The peaks were then integrated to extract the fractions of the bound and unbound populations[29, 30]. To obtain the midpoint of the binding titrations, data of the fraction bound, θ, vs. protein concentration, c_{prot}, were fitted with a binding isotherm [29, 78]

$$\theta = \frac{{c_{prot}}^n}{{c_{prot}}^n + {}^{app}{K_D}^n}, \tag{2}$$

where ${}^{\mathrm{app}}K_{\mathrm{D}}$ is the apparent dissociation constant corresponding to the midpoint, and n is treated as an adjustable parameter. For consistency, affinity data as a function of urea concentration (Figure S9) were all based on fitting the fraction bound.

In the experiment series with RS domain peptides of different net charge, the RS-RNA binding equilibrium varies from being in slow exchange for peptides of high positive net charge (e.g. WT peptides and single phosphorylation variants) to being in fast exchange for peptides approaching neutral net charge (e.g. 14xE and 16xE). Hence, we computed the mean transfer efficiency over all selected bursts with active donor and acceptor dye for each protein concentration, instead of fitting independent peaks, to be consistent across all data sets. Note that the presence of urea ensured that oligomerisation was usually below the detection limit. Data of mean transfer efficiency, $\bar{E}$, vs. protein concentration were fitted with

$$\bar{E} = \bar{E}_{min} + (\bar{E}_{max} - \bar{E}_{min}) \frac{{c_{\mathrm{prot}}}^n}{{c_{\mathrm{prot}}}^n + {}^{\mathrm{app}}{K_{\mathrm{D}}}^n}, \tag{3}$$

where $\bar{E}_{max}$ and $\bar{E}_{min}$ are the mean transfer efficiencies of the fully bound and unbound states, respectively. Where possible, uncertainties in ${}^{\mathrm{app}}K_{\mathrm{D}}^n$ are given as standard deviations estimated from multiple measurements. Otherwise, we assume an uncertainty of a factor of two as an estimate[28], based on the typical variability we have observed among binding measurements.

To calculate the net charge of RNAs and proteins, we use the following values (all in units of elementary charge): $q = -1$ per RNA nucleotide (due to backbone phosphate), $q = -1$ for CF660R, $q = 0$ for Cy3B, $q = +1$ for arginine, $q = -1$ for glutamate, $q = -2$ for phosphoserine, $q = -1$ for the C-terminal carboxyl group, $q = 0$ for the C-terminal amide group, and $q = +1$ for the N-terminal amino group. We note that the N-terminal amine might be partially deprotonated at pH 8. However, this would only introduce an offset for all peptides and would thus not change the outcome of the data analysis or the conclusions.

Quantifying the change in RNA compaction upon binding

We use the change in transfer efficiency upon binding to estimate the corresponding change in average end-to-end distance of the RNA. To minimize variability due to burst loss and poor photon statistics at high protein concentrations, we averaged several saturated transfer efficiency histograms (usually the two highest protein concentrations) for the fully bound state. Every titration data set contains a measurement of free, unbound donor/acceptor-labeled RNA (without protein). We fitted both the saturated and the RNA-only transfer efficiency histograms with Gaussian peak functions to obtain the mean transfer efficiency of the bound (E_b) and unbound population (E_u), respectively. Multiple measurements of the difference in transfer efficiency, $\Delta E_{b\to u} = E_b - E_u$, were then averaged for each RS domain peptide (see Figure 2b).

For converting the transfer efficiency change into a change in average end-to-end distance, we approximate the distribution of the inter-dye distance, r, by a Gaussian chain with the probability density function

$$P(r) = 4\pi r^2 \left[\frac{3}{2\pi\langle r^2\rangle}\right]^{\frac{3}{2}} e^{\frac{-3}{2}\frac{r^2}{\langle r^2\rangle}}. \tag{4}$$

From the measured mean values of E_b and E_u, we obtain the respective root-mean-square distances, $\langle r_b^2\rangle^{1/2}$ and $\langle r_u^2\rangle^{1/2}$, by numerically solving

$$\langle E\rangle = \int_o^\infty \epsilon(r) P(r) dr \tag{5}$$

with respect to $\langle r^2\rangle$ [79], where $\epsilon(r)$ is the transfer efficiency

$$\epsilon(r) = \frac{R_0^6}{R^6 + r^6}, \tag{6}$$

with R_0 being the Förster radius ($R_0 = 6.0$ nm for Cy3B-CF660R [80]). Values obtained from several titrations were averaged for each RS domain peptide.

As an estimate of the systematic uncertainty dominated by instrument calibration, we report an uncertainty of ~ 0.03 on single mean transfer efficiency values [27, 80, 81]. Note, however, that the statistical uncertainty of the measurements taken on the same instrument on the same day are much lower (typically $\leq$0.01 [80]). The measured differences in transfer efficiency are thus more accurate than individual transfer efficiencies and less dependent on systematic errors, and hence we quote one standard deviation based on multiple measurements as an uncertainty for the difference in transfer efficiency between bound and unbound RNA.

Single-molecule FRET of surface-immobilised molecules

Due to the much larger net charge of the 60-nucleotide RNA than of the RS domain, we could observe two RS domain molecules binding to one rU59$^{\mathrm{CA/GGA}}$ molecule. However, as both binding events were within one order of magnitude in terms of affinity, we chose to perform surface experiments to best resolve single- and double-bound states and to obtain an accurate affinity for the 1:1 interaction.

The MicroTime 200 described above was used with a 532 nm continuous wave laser (LaserBoxx LBX-532-50-COL-PP, Oxxius) instead of the white light source. The laser power for acceptor excitation that was used to scan the surface and locate immobilized molecules was set to 5 µW (Sepia PLD 828 Multichannel Picosecond Diode Laser Driver, PicoQuant). Slides for surface measurements were prepared using PEGylated, biotinylated fused silica cover slides (MicroSurfaces, Inc) and a hybridization chamber (Secure Seal Hybridization Chambers SA8R-2.5, Grace Bio-Labs). Before assembling the two parts, the cover slide was sonicated for 15 min in MilliQ water containing 0.1% Tween [82]. The slides were then washed with MilliQ water and dried with compressed air, before binding them to the adhesive hybridization chambers.

First, 0.2 mg/mL neutravidin in 50 mM HEPES pH 8 and 128 mM NaCl was added to the measurement chamber and incubated for 10 - 15 minutes. Afterwards, the chamber was washed three times with buffer (50 mM HEPES pH 8, 128 mM NaCl) before 10 pM donor/acceptor-labeled rU59$^{\mathrm{CA/GGA}}$, diluted in the same buffer, was added. After approximately one minute, the chamber was washed again with buffer to remove unbound RNA. Measurements were performed with different protein concentrations in 50 mM HEPES pH 8, 128 mM NaCl, 0.01 % Tween, 0.2 mg/ml BSA, 5 mM ascorbic acid, 5 mM methyl viologen, 5 mM protocatechuic acid (PCA) and 10 nM protocatechuate-3,4-dioxygenase (PCD). Due to the additional ions introduced by the redox and oxygen scavenger systems, the NaCl concentration was reduced to match the overall ionic strength to that used in the free-diffusion experiments. The experiment was performed under argon atmosphere to reduce dye bleaching. The donor dye was excited with a laser power of 1.4 µW, and trajectories were recorded until both dyes bleached, followed by an additional 0.5 seconds to be used for background correction during data analysis.

Fluorescence time traces were selected for analysis based on the following criteria: (i) presence of anti-correlation between donor and acceptor signals coming from all three states (unbound, single-bound, double-bound); (ii) stable total photon count rates prior to bleaching; and (iii) $\geq$ 0.5 s of background signal available post-bleaching. For building transfer efficiency histograms, trajectories were binned in 50 ms bins. Counts of donor and acceptor photons where corrected for background and other effects analogous to the correction of burst counts described above. For the kinetic analysis, uncorrected trajectories were used with 20 ms binning and analyzed by hidden-Markov modelling as described previously [83]. Individual labeled and immobilized RNA molecules were observed in three states: (1) free, (2) with one RS domain, or (3) with two RS domains bound. The rate coefficients of the corresponding kinetic system

$$1 \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}^{\mathrm{obs}}}{\rightleftharpoons}} 2 \underset{k_{\mathrm{off},2}}{\overset{k_{\mathrm{on},2}^{\mathrm{obs}}}{\rightleftharpoons}} 3 \tag{7}$$

were obtained using a maximum-likelihood analysis of the trajectories, errors were determined using bootstrapping. The equilibrium dissociation constant for a single RS domain bound to RNA was determined using

$$K_{\mathrm{D}} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}} = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}^{\mathrm{obs}} \cdot {c_{\mathrm{prot}}}^{-1}}, \tag{8}$$

where c_{prot} is the concentration of the RS domain in solution. The error of the equilibrium dissociation constant was obtained by error propagation using an estimated error of 20% in protein concentration and the standard deviation of $k_{off}^{obs} \cdot (k_{on}^{obs})^{-1}$ obtained from bootstrapping.

REFERENCES

[1] Alberts, B. *Molecular biology of the cell*, 7th ed.; W. W. Norton & Company: New York, 2022.

[2] Borg, M.; Mittag, T.; Pawson, T.; Tyers, M.; Forman-Kay, J. D.; Chan, H. S. Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proceedings of the National Academy of Sciences (USA)* **2007**, *104*, 9650–5.

[3] Lee, C. W.; Ferreon, J. C.; Ferreon, A. C.; Arai, M.; Wright, P. E. Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation. *Proceedings of the National Academy of Sciences USA* **2010**, *107*, 19290–5.

[4] Bah, A.; Forman-Kay, J. D. Modulation of intrinsically disordered protein function by post-translational modifications. *Journal of Biological Chemistry* **2016**, *291*, 6696–705.

[5] Frege, T.; Uversky, V. N. Intrinsically disordered proteins in the nucleus of human cells. *Biochemistry and Biophysics Reports* **2015**, *1*, 33–51.

[6] Modic, M.; Adamek, M.; Ule, J. The impact of IDR phosphorylation on the RNA binding profiles of proteins. *Trends in Genetics.* **2024**, *40*, 580–586.

[7] Long, J. C.; Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal* **2009**, *417*, 15–27.

[8] Slišković, I.; Eich, H.; Müller-McNicoll, M. Exploring the multifunctionality of SR proteins. *Biochemical Society Transactions* **2021**, *50*, 187–198.

[9] Kumar, K.; Sinha, S. K.; Maity, U.; Kirti, P. B.; Kumar, K. R. R. Insights into established and emerging roles of SR protein family in plants and animals. *WIREs RNA* **2023**, *14*, e1763.

[10] Cho, S.; Hoang, A.; Sinha, R.; Zhong, X. Y.; Fu, X. D.; Krainer, A. R.; Ghosh, G. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences USA* **2011**, *108*, 8233–8238.

[11] Manley, J. L.; Krainer, A. R. A rational nomenclature for serine/arginine-rich protein splicing factors (SR proteins). *Genes & Development* **2010**, *24*, 1073–1074.

[12] Labourier, E.; Rossi, F.; Gallouzi, I.-E.; Allemand, E.; Divita, G.; Tazi, J. 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 Research* **1998**, *26*, 2955–2962.

[13] Haynes, C.; Iakoucheva, L. M. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic Acids Research* **2006**, *34*, 305–312.

[14] Xiang, S.; Gapsys, V.; Kim, H. Y.; Bessonov, S.; Hsiao, H. H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Lührmann, R.; Groot, B. D.; Zweckstetter, M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure* **2013**, *21*, 2162–2174.

[15] Cléry, A.; Sinha, R.; Anczuków, O.; Corrionero, A.; Moursy, A.; Daubner, G. M.; Valcárcel, J.; Krainer, A. R.; Allain, F. H. Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition. *Proceedings of the National Academy of Sciences USA* **2013**, *110*.

[16] Cléry, A.; Krepl, M.; Nguyen, C. K.; Moursy, A.; Jorjani, H.; Katsantoni, M.; Okoniewski, M.; Mittal, N.; Zavolan, M.; Sponer, J.; Allain, F. H. Structure of SRSF1 RRM1 bound to RNA reveals an unexpected bimodal mode of interaction and explains its involvement in SMN1 exon7 splicing. *Nature Communications* **2021**, *12*.

[17] Tackel, R.; Manley, J. L. The human splicing factors ASF/SF2 and SC35 possess distinct, functionally significant RNA binding specificities. *The EMBO Journal* **1995**, *14*, 3540–3551.

[18] Sanford, J. R.; Coutinho, P.; Hackett, J. A.; Wang, X.; Ranahan, W.; Caceres, J. F. Identification of nuclear and cytoplasmic mRNA targets for the shuttling protein SF2/ASF. *PLoS ONE* **2008**, *3*.

[19] Velazquez-Dones, A.; Hagopian, J. C.; Ma, C. T.; Zhong, X. Y.; Zhou, H.; Ghosh, G.; Fu, X. D.; Adams, J. A. Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *Journal of Biological Chemistry* **2005**, *280*, 41761–41768.

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

[21] Aubol, B. E.; Plocinik, R. M.; Hagopian, J. C.; Ma, C. T.; McGlone, M. L.; Bandyopadhyay, R.; Fu, X. D.; Adams, J. A. Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *Journal of Molecular Biology* **2013**, *425*, 2894–2909.

[22] Keshwani, M. M.; Aubol, B. E.; Fattet, L.; Ma, C. T.; Qiu, J.; Jennings, P. A.; Fu, X. D.; Adams, J. A. Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochemical Journal* **2015**, *466*, 311–322.

[23] Fargason, T.; De Silva, N. I. U.; Powell, E.; Zhang, Z.; Paul, T.; Shariq, J.; Zaharias, S.; Zhang, J. Peptides that Mimic RS repeats modulate phase separation of SRSF1, revealing a reliance on combined stacking and electrostatic interactions. *eLife* **2023**, *12*, e84412.

[24] Shinn, M. K.; Tomares, D. T.; Liu, V.; Pant, A.; Qiu, Y.; Vitalis, A.; Song, Y. J.; Ayala, Y.; Ruff, K. M.; Strout, G. W.; Lew, M. D.; Prasanth, K. V.; Pappu, R. V. Nuclear speckle proteins form intrinsic and MALAT1-dependent microphases. *Cell* **2026**, *189*, 832–852.

[25] Abramson, J. et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500.

[26] Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Science* **2018**, *27*, 112–128.

[27] Hellenkamp, B. et al. Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nature Methods* **2018**, *15*, 669–676.

[28] Chowdhury, A.; Borgia, A.; Ghosh, S.; Sottini, A.; Mitra, S.; Eapen, R. S.; Borgia, M. B.; Yang, T.; Galvanetto, N.; Ivanović, M. T.; Łukijańczuk, P.; Zhu, R.; Nettels, D.; Kundagrami, A.; Schuler, B. Driving forces of the complex formation between highly charged disordered proteins. *Proceedings of the National Academy of Sciences USA* **2023**, *120*, e2304036120.

[29] Holmstrom, E. D.; Liu, Z.; Nettels, D.; Best, R. B.; Schuler, B. Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. *Nature Communications* **2019**, *10*, 2453.

[30] Borgia, A.; Borgia, M. B.; Bugge, K.; Kissling, V. M.; Heidarsson, P. O.; Fernandes, C. B.; Sottini, A.; Soranno, A.; Buholzer, K. J.; Nettels, D.; Kragelund, B. B.; Best, R. B.; Schuler, B. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **2018**, *555*, 61–66.

[31] Kundinger, S. R.; Dammer, E. B.; Yin, L.; Hurst, C.; Shapley, S.; Ping, L.; Khoshnevis, S.; Ghalei, H.; Duong, D. M.; Seyfried, N. T. Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization. *Journal of Biological Chemistry* **2021**, *297*, 101306.

[32] Hillger, F.; Nettels, D.; Dorsch, S.; Schuler, B. Detection and analysis of protein aggregation with confocal single molecule fluorescence spectroscopy. *Journal of Fluorescence* **2007**, *17*, 759–765.

[33] Zhang, W.; Ni, H.; Capp, M. W.; Anderson, C. F.; Lohman, T. M.; Record, M. T. The Importance of Coulombic End Effects: Experimental Characterization of the Effects of Oligonucleotide Flanking Charges on the Strength and Salt Dependence of Oligocation (L8+) Binding to Single-Stranded DNA Oligomers. *Biophysical Journal* **1999**, *76*, 1008–1017.

[34] Cubuk, J.; Alston, J. J.; Incicco, J. J.; Holehouse, A. S.; Hall, K. B.; Stuchell-Brereton, M. D.; Soranno, A. The disordered N-terminal tail of SARS-CoV-2 Nucleocapsid protein forms a dynamic complex with RNA. *Nucleic Acids Research* **2023**, *52*, 2609–2624.

[35] Nüesch, M. F.; Pietrek, L.; Holmstrom, E. D.; Nettels, D.; von Roten, V.; Kronenberg-Tenga, R.; Medalia, O.; Hummer, G.; Schuler, B. Nanosecond chain dynamics of single-stranded nucleic acids. *Nature Communications* **2024**, *15*, 6010.

[36] Wegener, M.; Müller-McNicoll, M. *Advances in Experimental Medicine and Biology*; Springer, 2019; Vol. 1203; pp 83–112.

[37] Ma, C. T.; Velazquez-Dones, A.; Hagopian, J. C.; Ghosh, G.; Fu, X. D.; Adams, J. A. Ordered Multi-site Phosphorylation of the Splicing Factor ASF/SF2 By SRPK1. *Journal of Molecular Biology* **2008**, *376*, 55–68.

[38] Ma, C. T.; Hagopian, J. C.; Ghosh, G.; Fu, X. D.; Adams, J. A. Regiospecific Phosphorylation Control of the SR Protein ASF/SF2 by SRPK1. *Journal of Molecular Biology* **2009**, *390*, 618–634.

[39] Xie, Y.; Jiang, Y.; Ben-Amotz, D. Detection of amino acid and peptide phosphate protonation using Raman spectroscopy. *Analytical Biochemistry* **2005**, *343*, 223–230.

[40] Fargason, T.; Powell, E.; De Silva, N. I. U.; Paul, T.; Zhang, Z.; Prevelige, P.; Zhang, J. Controlled by disorder: Phosphorylation modulates SRSF1 domain availability for spliceosome assembly. *Protein Science* **2025**, *34*, e70070.

[41] Mijalis, A. J.; Thomas, D. A.; Simon, M. D.; Adamo, A.; Beaumont, R.; Jensen, K. F.; Pentelute, B. L. A fully automated flow-based approach for accelerated peptide synthesis. *Nature Chemical Biology 2017 13:5* **2017**, *13*, 464–466.

[42] Hartrampf, N. et al. Synthesis of proteins by automated flow chemistry. *Science* **2020**, *368*, 980–987.

[43] Conibear, A. C. Deciphering protein post-translational modifications using chemical biology tools. *Nature Reviews Chemistry 2020 4:12* **2020**, *4*, 674–695.

[44] Williams, E. T.; Schiefelbein, K.; Schuster, M.; Ahmed, I. M. M.; De Vries, M.; Beveridge, R.; Zerbe, O.; Hartrampf, N. Rapid flow-based synthesis of post-translationally modified peptides and proteins: a case study on MYC's transactivation domain. *Chem. Sci.* **2024**, *15*, 8756–8765.

[45] Bilbrough, T.; Piemontese, E.; Seitz, O. Dissecting the role of protein phosphorylation: a chemical biology toolbox. *Chem. Soc. Rev.* **2022**, *51*, 5691–5730.

[46] Bressler, S. G.; Grunhaus, D.; Hurevich, M.; Friedler, A. Methods for studying the effects of phosphorylation patterns in proteins. *Biochemical Society Transactions* **2026**, *54*, 20250137.

[47] Kozeleková, A.; Náplavová, A.; Brom, T.; Gašparik, N.; Šimek, J.; Houser, J.; Hritz, J. Phosphorylated and phosphomimicking variants may differ – a case study of 14-3-3 protein. *Frontiers in Chemistry* **2022**,

Volume 10 - 2022.

[48] Le, S.; Yu, M.; Yan, J. Phosphorylation Reduces the Mechanical Stability of the α-Catenin/ β-Catenin Complex. *Angewandte Chemie International Edition* **2019**, *58*, 18663–18669.

[49] Paleologou, K. E.; Schmid, A. W.; Rospigliosi, C. C.; Kim, H.-Y.; Lamberto, G. R.; Fredenburg, R. A.; Lansbury, P. T.; Fernandez, C. O.; Eliezer, D.; Zweckstetter, M.; Lashuel, H. A. Phosphorylation at Ser-129 but not the phosphomimics S129E/D inhibits the fibrillation of α-synuclein. *Journal of Biological Chemistry* **2008**, *283*, 16895–16905.

[50] Roberts-Galbraith, R. H.; Ohi, M. D.; Ballif, B. A.; Chen, J.-S.; McLeod, I.; McDonald, W. H.; Gygi, S. P.; Yates, J. R.; Gould, K. L. Dephosphorylation of F-BAR protein Cdc15 modulates its conformation and stimulates its scaffolding activity at the cell division site. *Molecular Cell* **2010**, *39*, 86–99.

[51] Pérez-Mejías, G.; Velázquez-Cruz, A.; Guerra-Castellano, A.; Baños-Jaime, B.; Díaz-Quintana, A.; González-Arzola, K.; Ángel De la Rosa, M.; Díaz-Moreno, I. Exploring protein phosphorylation by combining computational approaches and biochemical methods. *Computational and Structural Biotechnology Journal* **2020**, *18*, 1852–1863.

[52] Somale, D.; Di Nardo, G.; di Blasio, L.; Puliafito, A.; Vara-Messler, M.; Chiaverina, G.; Palmiero, M.; Monica, V.; Gilardi, G.; Primo, L.; Gagliardi, P. A. Activation of RSK by phosphomimetic substitution in the activation loop is prevented by structural constraints. *Scientific Reports* **2020**, *10*, 591.

[53] Kliche, J.; Garvanska, D. H.; Simonetti, L.; Badgujar, D.; Dobritzsch, D.; Nilsson, J.; Davey, N. E.; Ivarsson, Y. Large-scale phosphomimetic screening identifies phospho-modulated motif-based protein interactions. *Molecular Systems Biology* **2023**, *19*, e11164.

[54] von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic cluster formation determines viscosity and diffusion in dense protein solutions. *Proceedings of the National Academy of Sciences* **2019**, *116*, 9843–9852.

[55] Mandell, D. J.; Chorny, I.; Groban, E. S.; Wong, S. E.; Levine, E.; Rapp, C. S.; Jacobson, M. P. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *Journal of the American Chemical Society* **2007**, *129*, 820–827, PMID: 17243818.

[56] Strickfaden, S. C.; Winters, M. J.; Ben-Ari, G.; Lamson, R. E.; Tyers, M.; Pryciak, P. M. M. A mechanism for cell-cycle regulation of MAP kinase signaling in a yeast differentiation pathway. *Cell* **2007**, *128*, 519–531.

[57] Cazalla, D.; Zhu, J.; Manche, L.; Huber, E.; Krainer, A. R.; Cáceres, J. F. Nuclear export and retention signals in the RS Domain of SR proteins. *Molecular and Cellular Biology* **2002**, *22*, 6871–6882.

[58] Bugge, K.; Sottini, A.; Ivanović, M. T.; Buus, F. S.; Saar, D.; Fernandes, C. B.; Kocher, F.; Martinsen, J. H.; Schuler, B.; Best, R. B.; Kragelund, B. B. Role of charges in a dynamic disordered complex between an IDP and a folded domain. *Nature Communications 2025 16:1* **2025**, *16*, 3242.

[59] Serrano, P.; Aubol, B. E.; Keshwani, M. M.; Forli, S.; Ma, C. T.; Dutta, S. K.; Geralt, M.; Wüthrich, K.; Adams, J. A. Directional Phosphorylation and Nuclear Transport of the Splicing Factor SRSF1 is Regulated by an RNA Recognition Motif. *Journal of Molecular Biology* **2016**, *428*, 2430–2445.

[60] Aubol, B. E.; Hailey, K. L.; Fattet, L.; Jennings, P. A.; Adams, J. A. Redirecting SR Protein Nuclear Trafficking through an Allosteric Platform. *Journal of Molecular Biology* **2017**, *429*, 2178–2191.

[61] Aubol, B. E.; Serrano, P.; Fattet, L.; Wüthrich, K.; Adams, J. A. Molecular interactions connecting the function of the serine-arginine-rich protein SRSF1 to protein phosphatase 1. *Journal of Biological Chemistry* **2019**, *293*, 16751–16760.

[62] Long, Y.; Sou, W. H.; Yung, K. W. Y.; Liu, H.; Wan, S. W. C.; Li, Q.; Zeng, C.; Law, C. O. K.; Chan, G. H. C.; Lau, T. C. K.; Ngo, J. C. K. Distinct mechanisms govern the phosphorylation of different SR protein splicing factors. *Journal of Biological Chemistry* **2019**, *294*, 1312–1327.

[63] Martínez-Lumbreras, S.; Morguet, C.; Sattler, M. Dynamic interactions drive early spliceosome assembly. *Current Opinion in Structural Biology* **2024**, *88*, 102907.

[64] Misteli, T.; Cáceres, J. F.; Spector, D. L. The dynamics of a pre-mRNA splicing factor in living cells. *Nature 1997 387:6632* **1997**, *387*, 523–527.

[65] Misteli, T.; Cáceres, J. F.; Clement, J. Q.; Krainer, A. R.; Wilkinson, M. F.; Spector, D. L. Serine phosphorylation of SR Proteins is required for their recruitment to sites of transcription in vivo. *The Journal of Cell Biology* **1998**, *143*, 297–307.

[66] Gates; Hartrampf Flow-based SPPS for protein synthesis: A perspective. *Pept. Sci.* **2020**, *112*, e24198.

[67] Müller, B. K.; Zaychikov, E.; Bräuchle, C.; Lamb, D. C. Pulsed Interleaved Excitation. *Biophysical Journal* **2005**, *89*, 3508–3522.

[68] König, I.; Zarrine-Afsar, A.; Aznauryan, M.; Soranno, A.; Wunderlich, B.; Dingfelder, F.; Stüber, J. C.; Plückthun, A.; Nettels, D.; Schuler, B. Single-molecule spectroscopy of protein conformational dynamics in live eukaryotic cells. *Nature Methods* **2015**, *12*, 773–779.

[69] Millman, K. J.; Aivazis, M. Python for scientists and engineers. *Computing in Science & Engineering* **2011**, *13*, 9–12.

[70] Oliphant, T. E. Python for scientific computing. *Computing in Science & Engineering* **2007**, *9*, 10–20.

[71] Walt, S. v. d.; Colbert, S. C.; Varoquaux, G. The NumPy array: a Structure for efficient numerical Computation. *Computing in Science & Engineering* **2011**, *13*, 22–30.

[72] Hunter, J. D. Matplotlib: a 2D graphics environment. *Computing In Science & Engineering* **2007**, *9*,

90–95.
[73] Virtanen, P. et al. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods* **2020**, *17*, 261–272.
[74] Schuler, B.; Müller-Späth, S.; Soranno, A.; Nettels, D. Application of confocal single-molecule FRET to intrinsically disordered proteins. *Intrinsically Disordered Protein Analysis: Volume 2, Methods and Experimental Tools* **2012**, 21–45.
[75] Nir, E.; Michalet, X.; Hamadani, K. M.; Laurence, T. A.; Neuhauser, D.; Kovchegov, Y.; Weiss, S. Shot-noise limited single-molecule FRET histograms: Comparison between theory and experiments. *Journal of Physical Chemistry B. Chem. B* **2006**, *110*, 22103–22124.
[76] Eggeling, C.; Fries, J. R.; Brand, L.; Günther, R.; Seidel, C. A. Monitoring conformational dynamics of a single molecule by selective fluorescence spectroscopy. *Proceedings of the National Academy of Sciences USA* **1998**, *95*, 1556–1561.
[77] Laurence, T. A.; Fore, S.; Huser, T. Fast, flexible algorithm for calculating photon correlations. *Optics Letters* **2006**, *31*, 829–831.
[78] Cantor, C.; Schimmel, P. *Biophysical Chemistry Part III: The Behavior of Biological Macromolecules.*, 1st ed.; W.H. Freeman and Company: New York, 1980; pp 859 – 866.
[79] Schuler, B.; Soranno, A.; Hofmann, H.; Nettels, D. Single-molecule FRET spectroscopy and the polymer physics of unfolded and intrinsically disordered proteins. *Annual Review of Biophysics* **2016**, *45*, 207–231.
[80] Klose, D.; Holla, A.; Gmeiner, C.; Nettels, D.; Ritsch, I.; Bross, N.; Yulikov, M.; Allain, F. H.; Schuler, B.; Jeschke, G. Resolving distance variations by single-molecule FRET and EPR spectroscopy using rotamer libraries. *Biophysical Journal* **2021**, *120*, 4842–4858.
[81] Hoffmann, A.; Kane, A.; Nettels, D.; Hertzog, D. E.; Baumgärtel, P.; Lengefeld, J.; Reichardó, G.; Horsley, D. A.; Seckler, R.; Bakajin, O.; Schuler, B. Mapping protein collapse with single-molecule flourescence and kinetic synchrotron radiation circular dichroism spectroscopy. *Proceedings of the National Academy of Sciences USA* **2007**, *104*, 105–110.
[82] Pan, H.; Xia, Y.; Qin, M.; Cao, Y.; Wang, W. A simple procedure to improve the surface passivation for single molecule fluorescence studies. *Physical Biology* **2015**, *12*.
[83] Zosel, F.; Mercadante, D.; Nettels, D.; Schuler, B. A proline switch explains kinetic heterogeneity in a coupled folding and binding reaction. *Nature Communications 2018 9:1* **2018**, *9*, 3332–.

ACKNOWLEDGEMENTS

We thank Guillaume Hautbergue for providing the pET24b-GB1-His_6-SRPK1-His_6 plasmid, Cristina Kim Xuan Nguyen for providing the pET24b-GB1-His_6-TEV-SRSF1-YS plasmid, Nicholas Krebs for cloning the different SRSF1 constructs, Ella Knüpling for providing assistance with SRPK1 production, and Erik Holmstrom and Andrea Soranno for insightful discussion. Furthermore, the authors gratefully acknowledge the Functional Genomics Center Zurich (FGCZ) of the University of Zurich and ETH Zurich, and in particular Dr. Serge Chesnov, for the support on mass spectrometry analysis. Financial support for this project was provided by the Swiss National Science Foundation: CRSII5_205922 and CRSII5_205922/2 (FA, BS), 10006187 (BS), 200021_200,865 (NH) and 10000072 (NH). Additionally, MS gratefully acknowledges the support provided through a FEBS LTF.

AUTHOR CONTRIBUTIONS

MS, BS, HB, NH designed research. MS, SH, LF, NK and MN prepared samples. HB synthesised the peptides. MS, LF, SH, AP and HB acquired the data. MS, AP, HB analyzed the data. Technical support and analytical tools were provided by AC and DN. BS, NH and FA supervised the project. MS and BS wrote the manuscript with input from all authors.

DECLARATION OF CONFLICT OF INTEREST

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information

The Supporting Information is available online, and contains the supplementary figures, tables and methods as well as the peptide synthesis data.

Materials and correspondence

Any correspondence and requests for materials should be addressed to Marie Synakewicz, Nina Hartrampf or Benjamin Schuler.

Supplementary Information

Phosphorylation tunes electrostatically driven protein-RNA interactions

Marie Synakewicz,[1,*] Adhvitha Premanand,[2,1] Héloïse Bürgisser,[2] Sarah Habeler,[1] Lucia R. Franchini,[1] Noémie Kociolek,[3] Mark Nüesch,[1] Antoine Cléry,[3] Daniel Nettels,[1] Frédéric Allain,[3] Nina Hartrampf,[2,*] and Benjamin Schuler[1,4,*]

[1] *Department of Biochemistry, University of Zurich, Switzerland*
[2] *Department of Chemistry, University of Zurich, Switzerland*
[3] *Institute of Biochemistry, ETH Zurich, Switzerland*
[4] *Department of Physics, University of Zurich, Switzerland*

[*] To whom correspondence should be addressed: m.synakewicz@bioc.uzh.ch, nina.hartrampf@chem.uzh.ch, schuler@bioc.uzh.ch

CONTENTS

I. SUPPLEMENTARY FIGURES

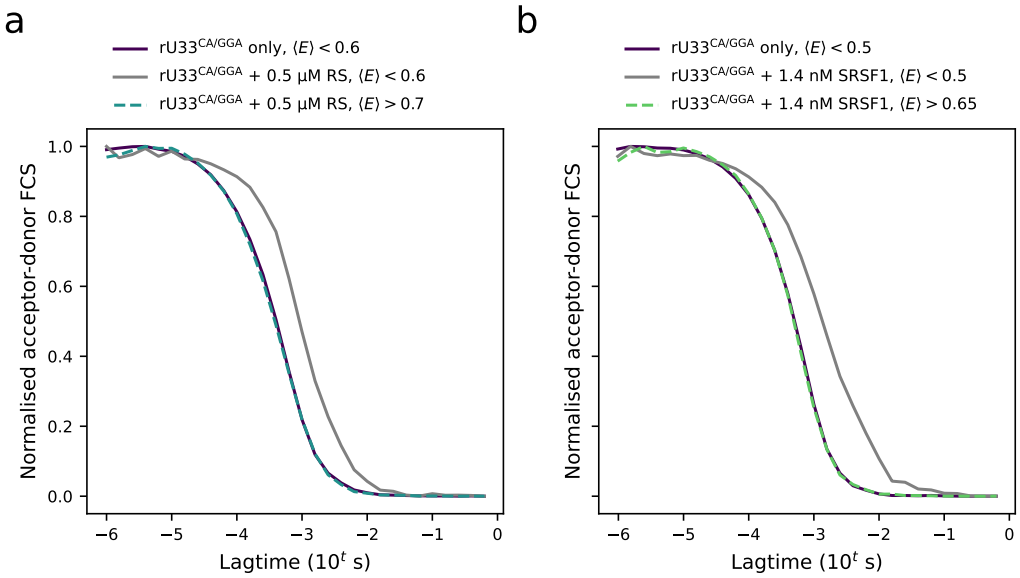

FIGURE S1. Using subpopulation-specific FCS to distinguish between monomeric and oligomeric species. Shown are the same curves displayed as insets for RS domain (a) and SRSF1 (b) histograms in Figure 1c, here overlaid with data of the respective RNA-only measurements (purple curves). At saturation, the data obtained for transfer efficiency ranges $E < 0.6$ (RS domain) and $E < 0.5$ (SRSF1) are shifted to larger lag times, indicating the presence of slower-diffusing molecules (gray curves). In contrast, molecules with transfer efficiency $E > 0.7$ (RS domain, dashed blue curve) and $E > 0.65$ (SRSF1, dashed green curve), corresponding to protein bound to RNA, diffuse on a similar timescale as the unbound RNA. Together, these data show that at saturating protein concentrations, particles with low apparent transfer efficiency are distinct from the unbound RNA and RNA bound to one protein molecule.

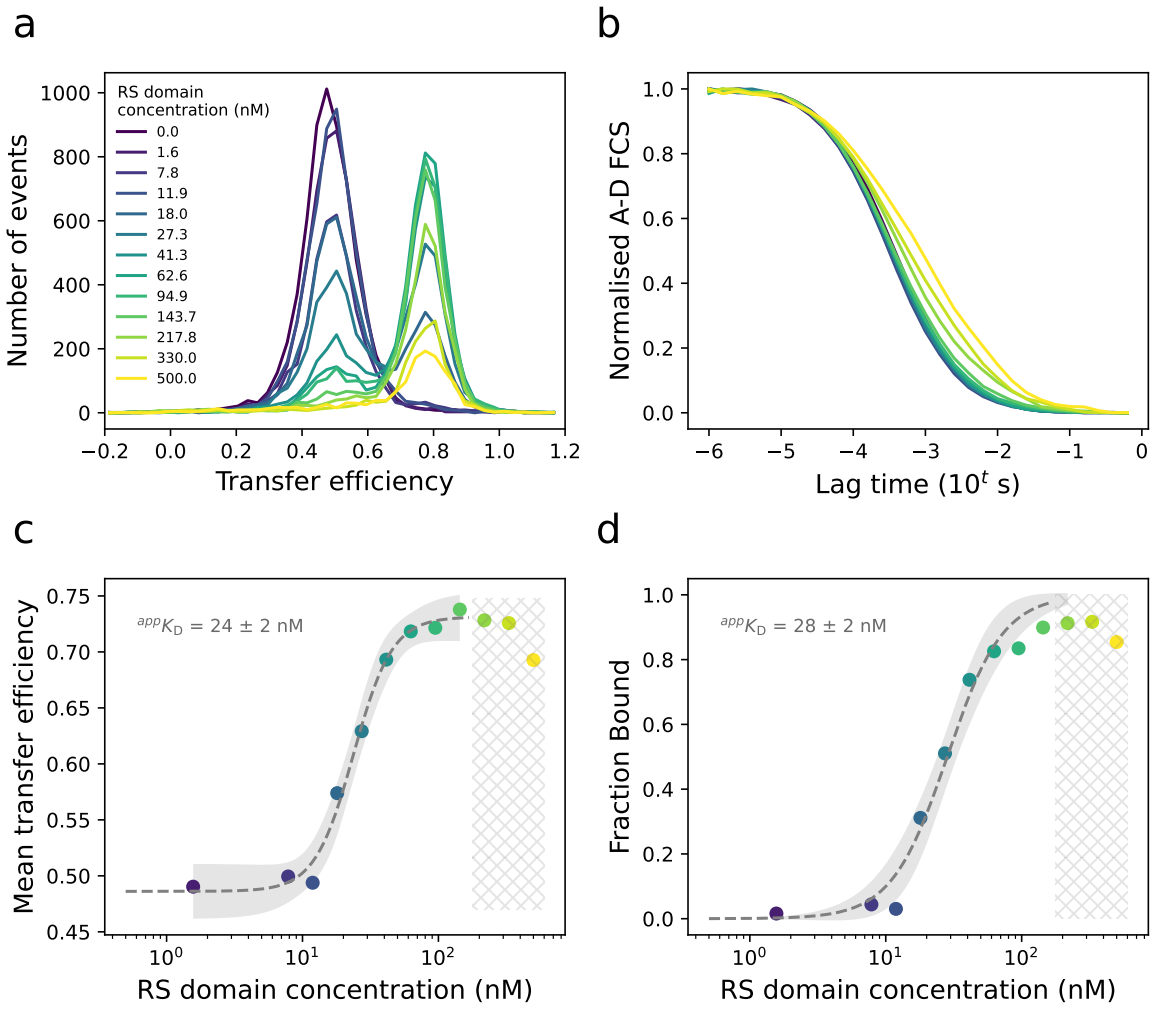

FIGURE S2. Illustration of the data analysis workflow for identifying and removing the contribution of aggregates in binding isotherms, using the rU33$^{CA/GGA}$ RNA and the RS domain as a representative example. (a) Overlay of transfer efficiency histograms of donor/acceptor-labeled RNA at different protein concentrations (see legend) showing a reduced number of bursts at high protein concentrations. (b) Overlay of normalised acceptor-donor (AD) fluorescence correlation curves for each protein concentration. At protein concentrations above ∼150 nM, the signal clearly contains slower-diffusing species in addition to the 1:1 RNA-RS domain complex. (c) At each protein concentration, the mean transfer efficiency is obtained from all selected bursts. Owing to the appearance of higher-molecular weight species with lower apparent transfer efficiencies, the mean transfer efficiency decreases at high protein concentrations. Hence, data containing substantial contributions from slower-diffusing species, as determined by FCS (hatched in gray), are excluded from the fit of the binding isotherm. This method of determining the affinity is only reliable if saturation is achieved to good approximation before burst loss becomes significant, as the top baseline is otherwise not representative of the transfer efficiency of the fully bound RNA. The gray dashed line and shaded area correspond to a fit of Equation (3) to the data and the 95% confidence band. Fit value and fit error for the apparent affinity are indicated. (d) Alternatively, the fraction of bound RNA at a given protein concentration can be obtained by globally fitting the histograms shown in (a) with two Gaussian peak functions, one for the bound and the unbound population, respectively. Since the transfer efficiencies of slower-diffusing species overlap with those of the unbound population, the apparent fraction bound will never be equal to 1 – even at saturation. Hence, data containing contributions from slower-diffusing species, as determined by FCS (b), are excluded from the fit of a binding isotherm. Since fitting isotherms to the fraction bound does not rely on fit parameters for the unbound and saturation baselines, it is still possible to obtain a reasonable estimate for the midpoint even in the presence of significant burst loss. The gray dashed line and shaded area correspond to a fit of Equation (2) to the data and the 95% confidence band. Fit value and fit error for the apparent affinity are indicated.

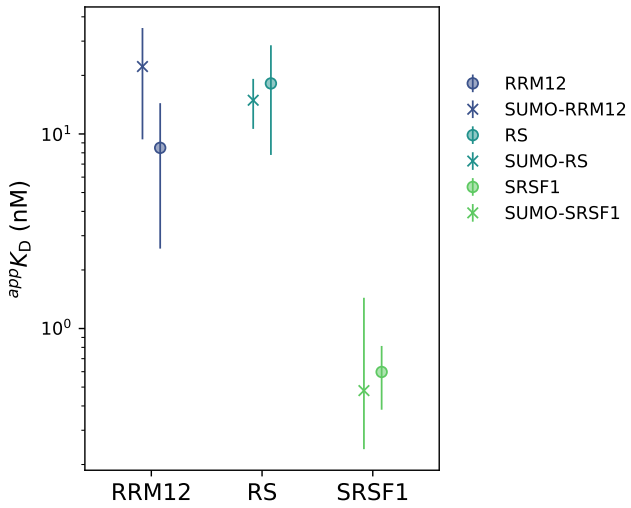

FIGURE S3. Presence of the SUMO-tag has no measurable effect on affinity. Comparison of affinities of different protein variants (see legend) to rU33$^{CA/GGA}$ RNA obtained with and without SUMO tag. For details on affinities and errors see Table S2. On average, all affinities of SUMO-tagged and un-tagged constructs are within error.

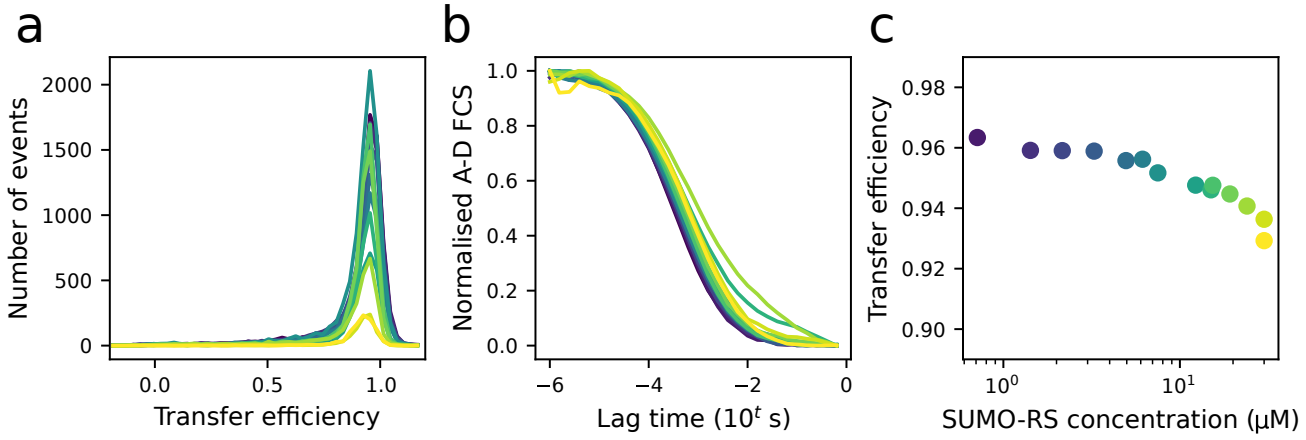

FIGURE S4. Obtaining an affinity for the interaction of the RS domain to the rU9$^{CA/GGA}$ RNA. (a) Transfer efficiency histograms of a titration of SUMO-RS into rU9$^{CA/GGA}$ show a substantial reduction in the number of bursts at the micromolar protein concentrations that are required for these measurements due to the reduced affinity of the RS domain for short ssRNAs . The SUMO-tagged construct was chosen to be able to increase the maximum protein concentration that could be used compared to the untagged RS domain. Colours corresponds to protein concentrations between 0 µM (purple) and 30 µM (yellow). Above 30 µM, burst loss was too severe to obtain meaningful histograms. (c) Fluorescence correlation curves indicate oligomerisation at high protein concentration as a likely source of the reduction in bursts. (b) Although the mean transfer efficiency (obtained by fitting Gaussian peak functions to the respective histograms in (a)) shows the onset of a binding response, it cannot be fitted meaningfully to obtain an apparent affinity. Hence, by estimating that 30 µM SUMO-RS corresponds to 5 to 20 % bound RNA, we infer an apparent affinity in the range of $150 < {}^{app}K_D < 600$ µM.

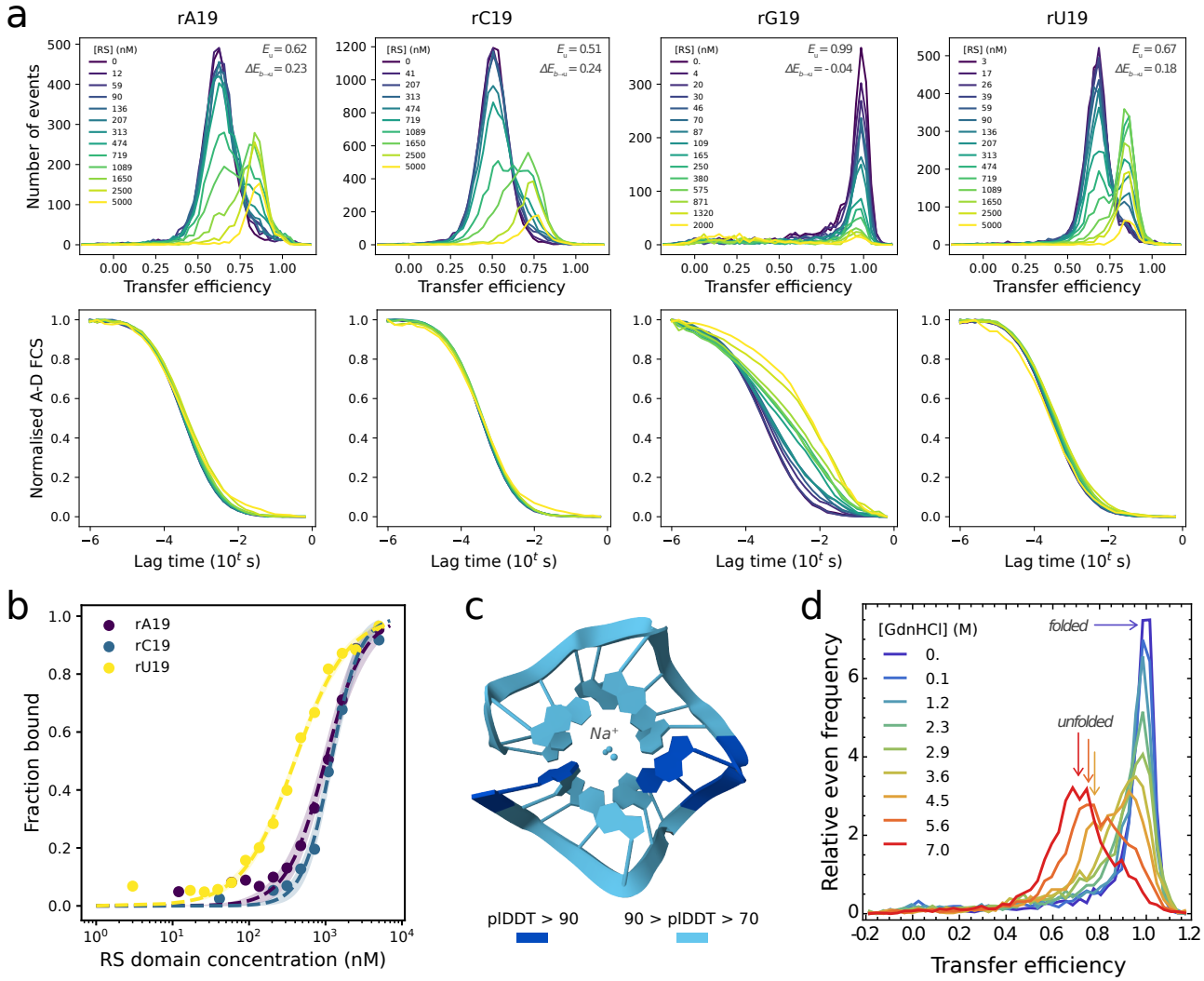

FIGURE S5. RS domain interactions with homopolymeric single-stranded RNA (ssRNA) containing different nucleobases. (a) Transfer efficiency histograms (top) and acceptor-donor (AD) fluorescence correlation curves (bottom) of titrations of donor/acceptor-labeled 19 bp ssRNAs with RS domain. Note that data obtained for rG19 differ from the other ssRNAs, likely due to it folding into a G-quadruplex-like structure (see (c) and (d)). All RNAs except rG19 are relatively expanded, with transfer efficiencies between 0.51 and 0.67, and they show compaction upon RS domain binding. In contrast, rG19 exhibits a very high transfer efficiency near 1, and addition of the RS domain results in only a small shift to lower transfer efficiency ($\Delta E \sim 0.04$) and pronounced aggregation/oligomerisation, as evident from the FCS measurements shown below. It has previously been proposed that the RS domain can unfold G-quadruplexes [1]. We do not observe indications for G-quadruplex unfolding, and based on our observations of ssRNA-RS domain interactions and previous findings regarding nucleic acid interactions with positively charged proteins [2], we expect binding of the RS domain to favour RNA compaction (as observed for rA19, rC19, and rU19) and structure formation. (b) Binding isotherms obtained from the transfer efficiency histograms shown in (a), including fits to Equation (2) and 95% confidence bands. The apparent affinities are listed in Table S2. Due to extensive aggregation and only a minor shift in transfer efficiency upon the addition of RS domain, it was not possible to obtain a meaningful binding isotherm for rG19. Although preferential interactions with arginine have been suggested for guanosine and cytosine [1, 3], we suggest that the minor differences in affinity observed ($\Delta G < 2\ k_B T$) could arise from differences in base stacking within ssRNAs (e.g. $\Delta G \sim 1\,k_B T$ for 2-3 stacked bases [4, 5]). (c) The model of the rG19 structure and 3 Na^+ ions generated with AlphaFold3 [6] resembles a G-quadruplex. However, due to the absence of non-G bases in loops, the topology appears strained and less ordered than usual for G-quadruplexes. Based on this model, the predicted distance between the terminal fluorophore attachment sites would be ~2 nm, resulting in a theoretical transfer efficiency of $E \sim 1$ for the dye-pair Cy3B-CF660R ($R_0 = 6$ nm), in accord with the high transfer efficiency we observe (a). (d) The rG19 structure requires high concentrations of guanidine hydrochloride (GdnHCl) to be unfolded, highlighting the high stability of this structure. Denaturation of 50 pM donor/acceptor-labeled rG19 was performed in 10 mM HEPES pH 8, 0.01 % Tween-20, 5 mM DTT. Experiments were performed in untreated µ-slides (Ibidi), starting at 7 M GdnHCl (ThermoScientific) and then sequentially diluting the denaturant concentration.

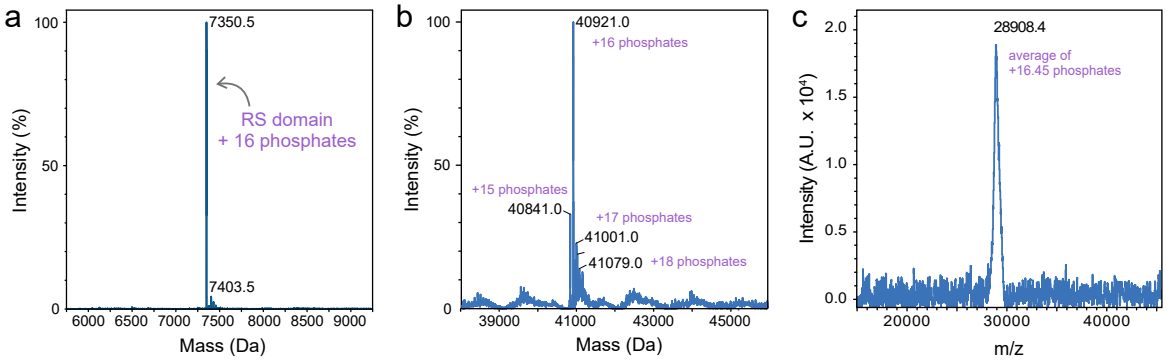

FIGURE S6. Mass spectrometry (MS) data of phosphorylated proteins. (a) Electrospray ionisation (ESI) MS of RS domain phosphorylated by SRPK1 corresponding to a species with 16 added phosphates (*cf.* calculated mass for unphosphorylated RS domain: 6071.59 Da). (b) ESI-MS of SUMO-SRSF1 phosphorylated by SRPK1 (*cf.* calculated mass for unphosphorylated SUMO-SRSF1 with N-terminal Met cleaved: 39 641.80 Da). Annotations highlight a major species of 16 added phosphates and several minor species with 15, 17 and 18 added phosphates. (c) MALDI-MS of SRSF1 phosphorylated by SRPK1 yielding on average of 16 phosphates added (*cf.* calculated mass for unphosphorylated SRSF1: 27 592.38 Da). Here, the apparent non-integer number of added phosphates can be due to lower accuracy and signal broadening typical of MALDI and/or sample heterogeneity. For all calculations, we assumed a mass increase of 79.98 Da per phosphorylation. Mass spectroscopy was performed by the Functional Genomics Center Zurich.

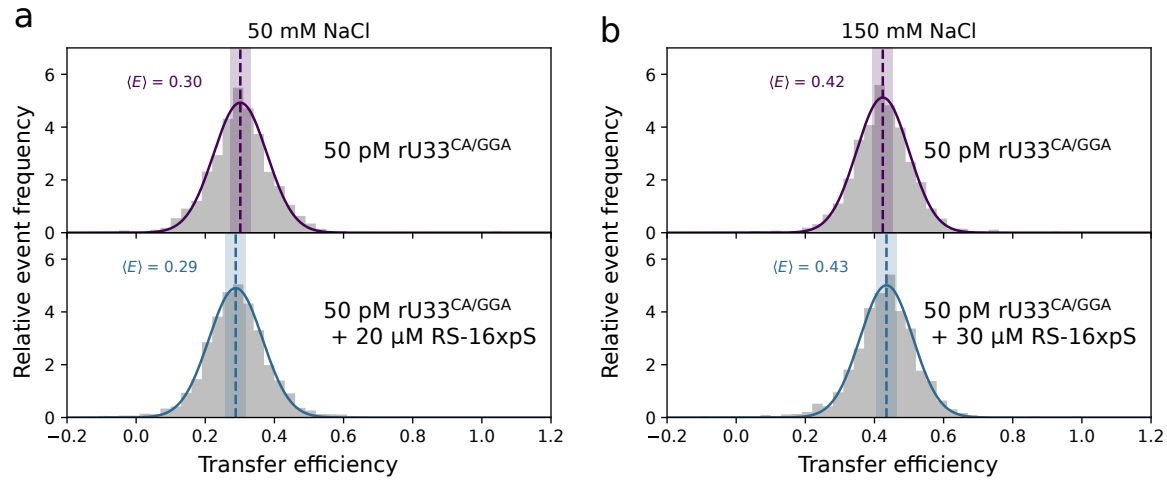

FIGURE S7. Probing possible interactions between RNA and the phosphorylated RS domain. Measurements performed using donor/acceptor-labeled $rU33^{CA/GGA}$ in standard measurement buffer but containing either (a) 50 mM NaCl or (b) 150 mM NaCl. Shown are normalised transfer efficiency histograms of RNA alone (top) and RNA with enzymatically phosphorylated RS domain (bottom). Even at reduced ionic strength, we cannot detect a significant change in RNA end-to-end distance upon the addition of phosphorylated RS domain, indicating that there is no binding at these protein concentrations.

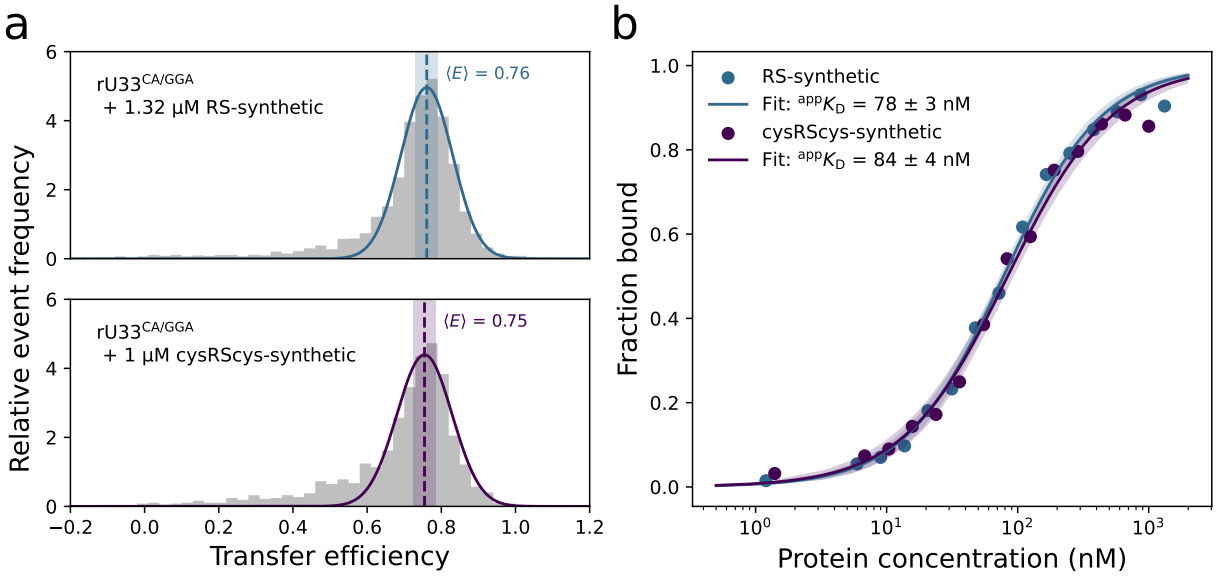

FIGURE S8. Cysteine residues have a negligible effect on the RNA affinity of synthetic WT RS domain peptides. (a) Representative normalised transfer efficiency histograms of donor/acceptor-labeled rU33$^{CA/GGA}$ RNA saturated with synthesised RS domain variants measured in 20 mM HEPES pH 8, 300 mM NaCl, 0.01 % Tween-20, 0.2 mg/mL BSA, 10 mM DTT, 0.1 U/µL RNAse inhibitor. The mean of the fit with a Gaussian peak function is indicated by a dashed line, the shaded band represents an uncertainty of 0.03 [7]. (b) Corresponding titration curves for each peptide showing the fraction bound as a function of protein concentration. Apparent affinities resulting from the fit to Equation (2) (given as best-fit value and fit error) are shown in the legend. Shaded areas around the respective fits indicate 95% confidence bands.

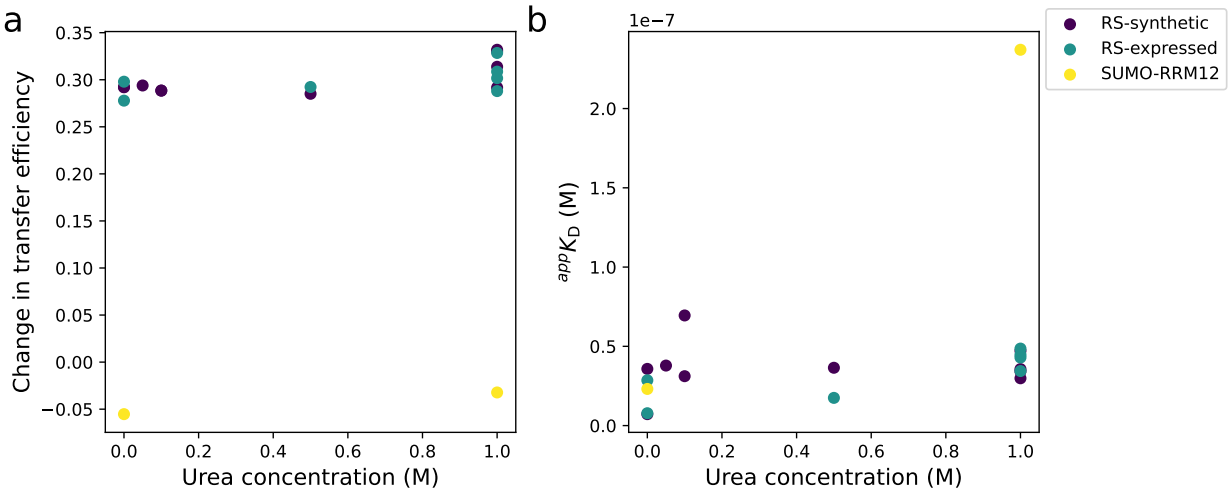

FIGURE S9. Assessing the impact of urea on RNA-protein interactions. Donor/acceptor-labeled rU33$^{CA/GGA}$ RNA was titrated with synthetically produced and with expressed wild-type RS domain at different urea concentrations to obtain binding curves. For comparison, we also include data for SUMO-RRM12 at 0 M and 1 M urea. The data shown are values obtained for individual titrations. (a) The change in the mean transfer efficiency between bound and unbound RNA as a function of urea concentration. The change in transfer efficiency upon protein binding remains similar with increasing urea concentration ($<$ 10% change on average between 0 M and 1 M urea). (b) The apparent affinity between RS domain and RNA changes only moderately with increasing urea concentration (∼2-fold change on average between 0 M and 1 M urea). The scatter of the data at $<$0.2 M urea indicates that experimental variability (e.g. introduced by aggregation and burst loss) contributes substantially to the uncertainty of the apparent affinity obtained under (near-)native conditions. Compared to the RRM12-RNA affinity, which displays a decrease by almost one order of magnitude between 0 M and 1 M urea, the urea dependence observed for RS domain-RNA interactions is small.

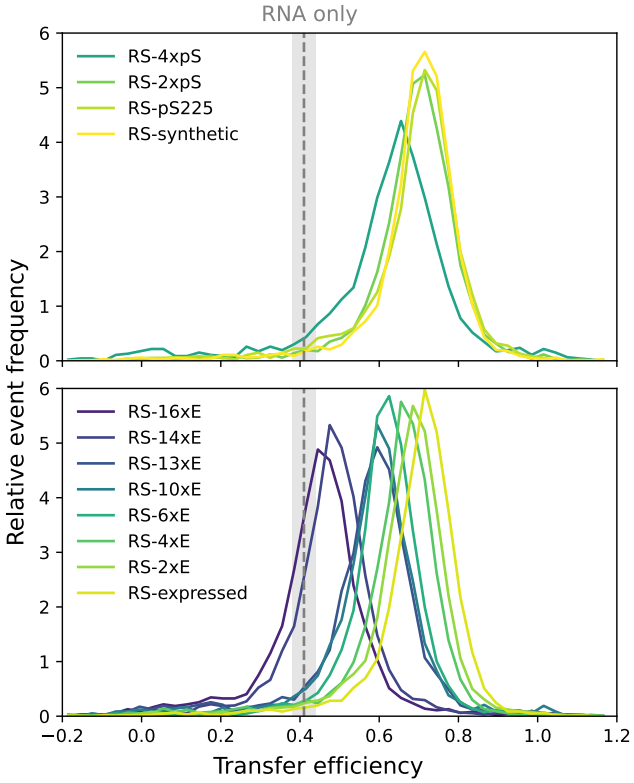

FIGURE S10. Representative normalised transfer efficiency histograms of donor/acceptor-labeled rU33$^{\mathrm{CA/GGA}}$ RNA saturated with different RS domain variants, one for each net charge, separated into phosphorylated (top) and phosphomimetic (bottom) peptides, measured in 1 M urea. The transfer efficiency of the unbound RNA is indicated with a gray dashed line and a shaded band representing an uncertainty of 0.03 [7]. For clarity, the histogram of only one of the peptides with a single phosphoserine is shown.

II. SUPPLEMENTARY TABLES

TABLE S1. Synthetic single-stranded RNAs with 5'- and 3'-end, and internal modifications for site-specific labeling. Bases binding specifically to the RRMs of SRSF1 are highlighted in bold. Modification nomenclature: 5ThioMC6-D – 5'-end modification bearing a di-thiol after a 6-carbon linker; 3AmMO – 3'-end modification bearing an amino group; iAmMC6T – internal modification of a thymine base with a 6-carbon linker and an amino group; 3Bio - 3'-end biotinylation.

Identifier	Sequence
rU9$^{CA/GGA}$	/5ThioMC6-D/U**CA**UU**GGA**U/3AmMO/
rU33$^{CA/GGA}$	/5ThioMC6-D/UUUUUUUUUUUUU**CA**UU**GGA**UUUUUUUUUUUUU/3AmMO/
rU59$^{CA/GGA}$	/5ThioMC6-D/UUUUU**CA**UU**GGA**UUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUUU/iAmMC6T/UUUU/3Bio
rU33	/5ThioMC6-D/UUUUUUUUUUUUUUUUUUUUUUUUUUUUUU/3AmMO/
rA19	/5ThioMC6-D/AAAAAAAAAAAAAAAAAAA/3AmMO/
rC19	/5ThioMC6-D/CCCCCCCCCCCCCCCCCCC/3AmMO/
rG19	/5ThioMC6-D/GGGGGGGGGGGGGGGGGGG/3AmMO/
rU19	/5ThioMC6-D/UUUUUUUUUUUUUUUUUUU/3AmMO/

TABLE S2. List of affinities obtained for each construct. Errors shown are standard deviations for values with $N > 1$ measurements, and a factor of two for those values with only one measurement available.

	Protein/peptide	RNA	Number of measurements	$^{app}K_D$ / M
native buffer	SUMO-RRM12	rU9$^{CA/GGA}$	2	$(2 \pm 1) \times 10^{-8}$
		rU33$^{CA/GGA}$	3	$(2 \pm 1) \times 10^{-8}$
		rU33	1	no binding detected
	RRM12	rU33$^{CA/GGA}$	2	$(9 \pm 6) \times 10^{-9}$
		rU33	1	no binding detected
	SUMO-RS	rU9$^{CA/GGA}$	1	$> 3 \times 10^{-5}$
		rU33$^{CA/GGA}$	2	$(1.5 \pm 0.4) \times 10^{-8}$
		rU33	1	$3^{+3}_{-1.5} \times 10^{-8}$
	RS	rU33$^{CA/GGA}$	2	$(2 \pm 1) \times 10^{-8}$
		rU33	1	$(2.4 \pm 0.1) \times 10^{-8}$
		rA19	1	$1^{+1}_{-0.5} \times 10^{-6}$
		rC19	1	$1.2^{+1.2}_{-0.6} \times 10^{-6}$
		rG19	1	could not be determined[a]
		rU19	1	$4^{+4}_{-2} \times 10^{-7}$
		rU59$^{CA/GGA}$	n.a.	$(1.7 \pm 0.4) \times 10^{-9}$ [b]
	RS-synthetic	rU33$^{CA/GGA}$	2	$(2 \pm 1) \times 10^{-8}$
	SUMO-SRSF1	rU9$^{CA/GGA}$	1	$(1.7 \pm 0.1) \times 10^{-8}$
		rU33$^{CA/GGA}$	1	$(5^{+5}_{-2.5}) \times 10^{-10}$
		rU33	2	$(5 \pm 2) \times 10^{-9}$
	SRSF1	rU33$^{CA/GGA}$	2	$(6 \pm 2) \times 10^{-10}$
		rU33	1	$(6^{+6}_{-3}) \times 10^{-9}$
	SUMO-pSRSF1	rU9$^{CA/GGA}$	1	$(7^{+7}_{-3.5}) \times 10^{-8}$
		rU33$^{CA/GGA}$	4	$(7.0 \pm 0.8) \times 10^{-8}$
	pSRSF1	rU33$^{CA/GGA}$	1	$(6.5^{+6.5}_{-3.3}) \times 10^{-8}$
1 M urea buffer	SUMO-RRM12	rU33$^{CA/GGA}$	1	$(2.4^{+2.4}_{-1.2}) \times 10^{-7}$
	RS	rU33$^{CA/GGA}$	4	$(4.3 \pm 0.5) \times 10^{-8}$
	RS2×E	rU33$^{CA/GGA}$	3	$(5.8 \pm 0.5) \times 10^{-8}$
	RS4×E	rU33$^{CA/GGA}$	3	$(1.5 \pm 0.2) \times 10^{-7}$
	RS6×E	rU33$^{CA/GGA}$	5	$(8 \pm 2) \times 10^{-7}$
	RS10×E	rU33$^{CA/GGA}$	4	$(1.7 \pm 0.3) \times 10^{-5}$
	RS13×E	rU33$^{CA/GGA}$	4	$(3.08 \pm 0.08) \times 10^{-5}$
	RS14×E	rU33$^{CA/GGA}$	3	$(1.3 \pm 0.3) \times 10^{-4}$
	RS16×E	rU33$^{CA/GGA}$	3	$(2.7 \pm 0.3) \times 10^{-4}$
	RS-synthetic	rU33$^{CA/GGA}$	4	$(3.7 \pm 0.7) \times 10^{-8}$
	RS-pS_{199}	rU33$^{CA/GGA}$	3	$(6 \pm 1) \times 10^{-8}$
	RS-pS_{223}	rU33$^{CA/GGA}$	4	$(9.0 \pm 0.8) \times 10^{-8}$
	RS-pS_{225}	rU33$^{CA/GGA}$	4	$(1.3 \pm 0.2) \times 10^{-7}$
	RS-2×pS	rU33$^{CA/GGA}$	4	$(1.1 \pm 0.1) \times 10^{-7}$
	RS-4×pS	rU33$^{CA/GGA}$	5	$(9 \pm 1) \times 10^{-7}$

[a] Due to aggregation and burst loss, see Figure S5 for details.
[b] Value for the first binding event obtained from surface experiments, for error propagation, see Methods section.

TABLE S3. Overview of RS domain peptide synthesis data. Purities are based on peak area of absorbances at 214 nm in UHPLC traces. Reported yields of crude peptides are relative to resin loading. Purities of all peptides after purification were >95 %. Molecular weights are calculated from monoisotopic masses of individual residues. See Supplementary Experimental Methods and Appendix A for more details. MW – molecular weight, n.d. – not determined.

Peptide	Crude peptide		Pure peptide			
	mass (mg)	purity (%)	mass (mg)	yield (%)	MW (Da)	
					calculated	observed
RS-synthetic	23	72	0.5	1.3	6067.1817	6067.1768
cysRScys-synthetic	8.4	71	n.d.	n.d.	6273.2001	6273.1808
RS-pS_{199}	8.4	49	0.79	4.5	6353.1701	6354.1582
RS-pS_{223}	6.2	51	0.9	5.3	6353.1701	6353.1682
RS-pS_{225}	12.6	32	0.6	1.6	6353.1701	6353.0933
RS-2×pS	7.4	24	0.6	3.1	6433.1327	6433.1041
RS-4×pS	57	28	1	1.8	6593.0654	6593.0495

TABLE S4. Sequences of DNA oligonucleotides used for molecular biology.

Name	**DNA sequence (5' → 3')**
SRSF1 Fwd / RRM12 Fwd	ACAGAGGACAGATTGGTGGGATGAGCGGTGGTGGTGTTATTCGTGGTCC
SRSF1 Rev / RS Rev	TTCGGATCCGGTCTTTATCAGGTGCGACTACGGCTGCGACTATGGC
RRM12 Rev	TTCGGATCCGGTCTTTATCAACCATCAACTTTCACACGAATATAGGCGGTTTCA CCTTC
RS Fwd	ACAGAGAACAGATTGGTGGGCGCAGTCCGAGCTATGGTCGTAGTCGTAGTC
pSumo Fwd	GAGCTCCGTCGACAAGCTTGCG
pSumo Rev	CCCACCAATCTGTTCTCTGTGAGCC

TABLE S5. Amino acid sequences of recombinantly expressed proteins and peptides. The Ulp1 cleavage site is indicated by "/". According to mass spectrometry, N-terminal methionines in parentheses were usually cleaved during expression in *E. coli*.

Name	**Amino acid sequence**
SUMO-SRSF1	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/MSGGGV IRGPAGNNDCRIYVGNLPPDIRTKDIEDVFSKYGAIRDIDLKNRRGGPPFAFVEFEDP RDAEDAVSGRDGYDYDGYRLRVEFPRSGRGTGRGGGGGGGGGAPRGRYGPPSRRS ENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVVEFVRKEDMTYAVR KLDNTKFRSHEGETAYIRVKVDGPRSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSR GSPRYSPRHSRSRSRT
SUMO-RRM12	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/MSGGGV IRGPAGNNDCRIYVGNLPPDIRTKDIEDVFSKYGAIRDIDLKNRRGGPPFAFVEFEDP RDAEDAVSGRDGYDYDGYRLRVEFPRSGRGTGRGGGGGGGGGAPRGRYGPPSRRS ENRVVVSGLPPSGSWQDLKDHMREAGDVCYADVYRDGTGVVEFVRKEDMTYAVR KLDNTKFRSHEGETAYIRVKVDG
SUMO-RS (RS-expressed)	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/RSPSYGR SRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRT
RS-2×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/RSPSYGR SRSRSRSRSRSRSRSNSREREYSPRRSRGSPRYSPRHSRSRSRT
RS-4×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/REPSYG RSRSRERSRSRSRERSNSRSREYSPRRSRGSPRYSPRHSRSRSRT
RS-6×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/REPSYG RSRSRERSRSRSRERSNSRSREYSPRRERGSPRYEPRHSRSRSRT
RS-10×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/RSPSYGR SRERERERERERERENEREREYSPRRSRGSPRYSPRHSRSRSRT
RS-13×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/REPEYG RERERERERERERERENEREREYSPRRSRGSPRYSPRHSRSRSRT
RS-14×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/RSPSYGR ERERERERERERERENEREREYSPRRSRGSPRYSPRHERERERT
RS-16×E	(M)GHHHHHHGSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRL MEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG/REPEYG RERERERERERERERENEREREYSPRRSRGSPRYSPRHERERERT
Ulp1	MLVPELNEKDDDQVQKALASRENTQLMNRDNIEITVRDFKTLAPRRWLNDTIIEFFM KYIEKSTPNTVAFNSFFYTNLSERGYQGVRRWMKRKKTQIDKLDKIFTPINLNQSHW ALGIIDLKKKTIGYVDSLSNGPNAMSFAILTDLQKYVMEESKHTIGEDFDLIHLDCPQQ PNGYDCGIYVCMNTLYGSADAPLDFDYKDAIRMRRFIAHLILTDALKLEHHHHHH
SRPK1	MQYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTV TEGSHHHHHHMERKVLALQARKKRTKAKKDKAQRKSETQHRGSAPHSESDLPEQEE EILGSDDDEQEDPNDYCKGGYHLVKIGDLFNGRYHVIRKLGWGHFSTVWLSWDIQG KKFVAMKVVKSAEHYTETALDEIRLLKSVRNSDPNDPNREMVVQLLDDFKISGVNGT HICMVFEVLGHHLLKWIIKSNYQGLPLPCVKKIIQQVLQGLDYLHTKCRIIHTDIKPEN ILLSVNEQYIRRLAAEATEWQRSGAPPPSGSAVSTAPQPKPADKMSKNKKKKLKKKQ KRQAELLEKRMQEIEEMEKESGPGQKRPNKQEESESPVERPLKENPPNKMTQEKLE ESSTIGQDQTLMERDTEGGAAEINCNGVIEVINYTQNSNNETLRHKEDLHNANDCDV QNLNQESSFLSSQNGDSSTSQETDSCTPITSEVSDTMVCQSSSTVGQSFSEQHISQLQES IRAEIPCEDEQEQEHNGPLDNKGKSTAGNFLVNPLEPKNAEKLKVKIADLGNACWVH KHFTEDIQTRQYRSLEVLIGSGYNTPADIWSTACMAFELATGDYLFEPHSGEEYTRD EDHIALIIELLGKVPRKLIVAGKYSKEFFTKKGDLKHITKLKPWGLFEVLVEKYEWSQ EEAAGFTDFLLPMLELIPEKRATAAECLRHPWLNSVEHHHHHH

TABLE S6. Amino acid sequences of synthetic peptides. Note that the C-terminal functional group is an amide ($CONH_2$).

Name	**Amino acid sequence**
RS-synthetic	RSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRT
cysRScys-synthetic	CRSPSYGRSRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRTC
RS-pS_{199}	CRpSPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRTC
RS-pS_{223}	CRSPSYGRSRSRSRSRSRSRSRSNSRpSRSYSPRRSRGSPRYSPRHSRSRSRTC
RS-pS_{225}	CRSPSYGRSRSRSRSRSRSRSRSNSRSRpSYSPRRSRGSPRYSPRHSRSRSRTC
RS-2×pS	CRSPSYGRSRSRSRSRSRSRSRSNSRpSRpSYSPRRSRGSPRYSPRHSRSRSRTC
RS-4×pS	CRpSPSYGRSRSRpSRSRSRSRpSRSNSRSRpSYSPRRSRGSPRYSPRHSRSRSRTC

III. SUPPLEMENTARY EXPERIMENTAL METHODS

A. Preparation of recombinantly expressed proteins and peptides

1. SRSF1

The plasmid bearing His_6-SUMO-SRSF1 was transformed into chemically competent *E. Coli* BL21(DE3)-C41 cells (Lucigen). Cultures were grown in 2xYT supplemented with 50 µg/mL kanamycin at 37 °C until an OD600 of ~ 1.5 was reached, and protein expression was induced using 0.5 mM IPTG at 20 °C overnight. Cells were harvested, resuspended in lysis buffer (20 mM sodium phosphate (NaP) pH 7.5, 800 mM NaCl, 5 % glycerol, 0.01 % Tween-20, 2 mM DTT, 150 mM L-arginine, 150 mM L-glutamate) supplemented with 10 mM MgCl2, 10 U/mL benzonase (Merck) and 20 µg/mL RNase A (NEB), and lysed by passing the suspension three times through a high pressure homogeniser (HPL6, Maximator) cooled to 4 °C at 15,000-20,000 psi. The lysate was clarified by centrifugation at 35 000 ×g for 30 min at 4 °C and applied to a HisTrap Excel column (Cytiva, 5 ml per 1 l cell culture) equilibrated in lysis buffer. The column was washed with 15 column volumes (CVs) of lysis buffer, followed by 7 CVs of high-salt wash buffer (20 mM NaP pH 7.5, 3 M NaCl, 5 % glycerol, 0.01 % Tween-20, 2 mM DTT, 100 mM L-arginine, 100 mM L-glutamate) and 7 CVs of lysis buffer with 15 mM imidazole, before elution using 20 mM NaP pH 7.5, 800 mM NaCl, 5 % glycerol, 0.01 % Tween-20, 2 mM DTT, 150 mM L-arginine, 150 mM L-glutamate, 500 mM imidazole. Fractions containing protein were pooled and diluted at least 8-fold using 20 mM HEPES pH 8.0, 10 % glycerol, 0.01 % Tween-20, 0.5 mM TCEP and 5 M NaCl until the sample was clear (0.6 to 0.8 M ionic strength). Nucleic acid and protein contaminants were removed by ion exchange chromatography (IEX). After filtering (0.22 µm syringe filter, PES membrane, Techno Plastic Products AG), the diluted protein solution was applied to a MonoS 5/50 GL (Cytiva) equilibrated in 20 mM HEPES pH 8.0, 600 mM NaCl, 10 % glycerol, 0.01 % Tween-20, 0.5 mM TCEP. The column was then washed with 10 CV of the equilibration condition and a gradient of 0.6 to 1.5 M NaCl was used to elute His_6-SUMO-SRSF1. Fractions with absorbance ratios of $A260/A280 < 0.7$ were pooled, concentrated by centrifugation (Vivaspin 6, 10 kDa MWCO, PES, Cytiva, and Amicon Ultra 0.5, 10 kDa MWCO, cellulose, Merck Millipore), flash-frozen in liquid N_2 and stored at −80 °C. To obtain untagged SRSF1, His_6-SUMO-SRSF1 was diluted in of 20 mM HEPES pH 8, 150 mM NaCl, 5 % glycerol, 0.5 mM TCEP, 0.01 % Tween-20 and incubated with 100 µg ULP1 for 2 h at 30 °C. Cleaved constructs were again purified by IEX, as described above.

2. RRM12

The plasmid bearing His_6-SUMO-RRM12 was transformed into *E. Coli* BL21(DE3)-C41 cells. An overnight culture of this transformation was used to inoculate 2xYT medium containing 50 µg/mL kanamycin, which was then grown at 37 °C until an OD600 of ~ 0.8 was reached. The culture was left to cool to 20 °C and induced with 1 mM IPTG overnight. Cells were harvested at 4000 ×g and the pellet was resuspended in lysis buffer (20 mM NaP pH 7.4, 800 mM NaCl, 10 mM imidazole, 1 % TritonX) with subsequent flash-freezing. After thawing, the cells were lysed using a Maximator high-pressure homogenizer HPL6 (Maximator) at 17 to 21 MPa for 3 iterations. The lysate was clarified by centrifugation for 25 min at 4 °C and 40 000 ×g. The soluble protein fraction was loaded onto ∼20 mL Ni Sepharose Excel resin (GE Healthcare) equilibrated in wash buffer (20 mM NaP pH 7.4, 800 mM NaCl, 20 mM imidazole, 150 mM L-Arg, 150 mM L-Glu, 6 % v/v glycerol, 0.005 % v/v Tween20, 5 mM DTT). The resin-supernatant mixture was incubated for 30 min at 4 °C on a shaker. After incubation and subsequent gravity-assisted elution of the flow through, the column was washed with 10 column volumes (CV) wash buffer. H_6-SUMO-RRM12 was eluted with 8 CV elution buffer (20 mM NaP pH 7.4, 800 mM NaCl, 500 mM imidazole, 100 mM L-Arg, 100 mM L-Glu, 6 % v/v glycerol, 0.005 % v/v Tween20, 5 mM DTT). IMAC elutions were combined and dialysed against dialysis buffer (20 mM KP_i pH 7.4, 25 mM KCl, 10 % v/v glycerol, 0.01 % v/v Tween20) overnight at 4 °C. A HiPrep Q FF 16/10 column (GE Healthcare) was equilibrated in 20 mM KP_i pH 7.4, 10 % v/v glycerol, 2 mM DTT before loading the dialysed SUMO-RRM12. After washing, the protein was eluted with a gradient of 0 to 0.44 M KCl over 20 CV. His_6-SUMO-RRM12 was incubated with Ulp1 (10 µg Ulp1 per 5 µg of H_6-

SUMO-RRM12) in cleavage buffer (20 mM HEPES pH 7, 200 mM NaCl, 2 mM DTT, 0.5 % v/v glycerol) overnight at 4 °C. The cleaved RRM12 was purified by reverse IMAC using 20 mL Ni Sepharose Excel resin (Cytavia) pre-equilibrated with wash buffer (20 mM NaP pH 7.4, 800 mM NaCl, 142 mM L-Arg, 142 mM L-Glu, 6 % v/v glycerol, 0.005 % v/v Tween20, 5 mM DTT). The resin was incubated with the cleavage reaction mixture for 40 min at 4 °C on a shaker. After collecting the flow through, the resin was washed with 3×4 CV wash buffer with subsequent elution of the His_6-SUMO tag, uncleaved His_6-SUMO-RRM12, and Ulp1 using 3×4 CV elution buffer (20 mM NaP pH 7.4, 800 mM NaCl, 500 mM imidazole, 142 mM L-Arg, 142 mM L-Glu, 6 % v/v glycerol, 0.005 % v/v Tween20, 5 mM DTT). The collected flow-through and wash were pooled and concentrated by centrifugal filtering (Vivaspin 20, 10 kDa MWCO), and dialysed into 20 mM HEPES pH 8.0, 400 mM NaCl, 5 mM DTT. Proteins were flash-frozen in liquid N_2 and stored at −80 °C.

3. RS domain

Chemically competent *E.coli* BL21(DE3)-C41 cells (Lucigen) were transformed with pET-His_6-SUMO-RS and grown in 2xYT medium with 50 µg/mL kanamycin at 37 °C. Protein expression was induced by the addition of 0.5 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) at an OD600 of ∼0.7, followed by overnight incubation at 20 °C. Purification of the RS domain was carried out using the same IMAC protocol as the one used for His_6-SUMO-SRSF1. Fractions containing protein were diluted to an approximate ionic strength of 0.3 M and further purified by IEX chromatography using a MonoS 5/50 GL column (GE Healthcare), equilibrated in 20 mM HEPES pH 8, 10 % Glycerol, 0.5 mM TCEP and 300 mM NaCl. After washing with the equilibration conditions, a gradient from 0.3 to 1.5 M NaCl was used to elute His_6-SUMO-RS. Fractions containing protein were analyzed by SDS-PAGE and subsequently, those of sufficient purity were pooled, flash-frozen in liquid N_2 and stored at −80 °C. To obtain untagged RS domain, 1 mL of His_6-SUMO-RS was diluted in 15 mL of 20 mM HEPES pH 8, 150 mM NaCl, 5 % glycerol, 0.5 mM TCEP, 0.01 % Tween-20 and incubated with 100 µg ULP1 for 2 h at 30 °C. Cleaved constructs were again purified by IEX, as described above.

4. RS domain glutamate variants

Plasmids were transformed into chemically competent *E. coli* BL21 (DE3) cells (Thermo Scientific, EC0114), and protein expression and purification was carried out similar to that of the RS domain with following modifications. Protein expression was induced at OD600∼1. All IMAC buffers were prepared without Tween-20, and the lysis buffer did not contain Benzonase, RNAse and $MgCl_2$. After loading clarified cell lysates of constructs containing 2 to 6 glutamates, the HisTrap Excel column was washed with 10 CV of lysis buffer, 5 CV of high-salt wash buffer, and 5 CV of 3 % elution buffer (lysis buffer with 15 mM imidazole). Columns loaded with constructs containing ≥10 glutamates were washed using 15 CV lysis buffer and 5 CV of 3 % elution buffer. After elution, fractions containing protein were combined and diluted ∼0.3 M ionic strength using H_2O (peptides with 2 to 6 glutamates), or buffer exchanged into 20 mM NaP pH 7.5, 50 mM Glu, 50 mM Arg, 5 % glycerol, 2 mM DTT using a HiPrep 26/10 Desalting column (Pharmacia) (peptides with ≥10 glutamates). If phase separation occurred the construct was either diluted further, or 5 M NaCl was added gradually until condensates dissolved. SUMO-tag cleavage was achieved by the addition of 100 µg Ulp1 and incubation at 30 °C for ≥1 h. Precipitate was removed by filtration (0.22 µm syringe filter, PES membrane, Techno Plastic Products AG) and the cleavage reaction was loaded onto a MonoS 5/50 GL column (GE Healthcare) equilibrated in 20 mM HEPES pH 8, 5 % glycerol, 0.5 mM TCEP containing 300 mM NaCl (RS domain with 2 glutamates), 150 mM NaCl (peptides with 4 and 6 glutamates), or 50 mM (peptides with ≥10 glutamates). After a 10 CV wash using the respective equilibration conditions, elution gradients (20 CV in length) were adjusted to start at the equilibration salt concentration and end at 1.5 M NaCl (2 to 6 glutamates) or 1.2 M NaCl (≥10 glutamates). Fractions containing protein of sufficient purity as determined by SDS-PAGE were combined and concentrated by centrifugal filtration (Vivaspin 6, 3 kDa MWCO, PES, Cytiva, and Amicon Ultra 0.5 3 kDa MWCO, cellulose, Merck Millipore). All proteins but the variant containing 2 glutamates were further purified by size exclusion chromatography using a Superdex 75 Increase 10/300

GL column (Cytiva) equilibrated in 20 mM HEPES pH 8 and 1 M NaCl (4 and 6 glutamates), 0.75 M NaCl (10 and 13 glutamates) or 0.5 M NaCl (14 and 16 glutamates). Constructs containing 14 and 16 glutamates were purified twice by SEC to remove a closely eluting contaminant. Fractions containing pure protein were pooled, concentrated to maximally ∼0.9 mM (construct dependent), flash-frozen in liquid N_2 and stored at −80 °C. If necessary, proteins were further concentrated after thawing, directly prior to sample preparation for single-molecule experiments.

5. SRPK1

GB1-His_6-SRPK1-His_6 was transformed into *E. coli* BL21-CodonPlus(DE3)-RIL cells and incubated in LB media supplemented with 50 µg/mL kanamycin and 100 µg/mL chloramphenicol overnight at 37 °C. The overnight culture was used to inoculate 6 L of media with antibiotics. At an OD600 ∼ 0.6 protein expression was induced by the addition of 0.5 mM IPTG and incubation for 4 h at 37 °C. Cells were harvested by centrifugation at 4000 ×g, 4 °C and resuspended in 10 mL lysis buffer (20 mM NaP pH 7.4, 800 mM NaCl, cOMPLETE EDTA-free protease inhibitor cocktail per 1 g cell pellet). Cells were lysed using a microfluidizer (M-110L, Microfluidics) using three passes at 15000 psi. The lysate was clarified by centrifugation at 17 000 ×g, 4 °C. The supernatant was applied to 6 ml Ni-NTA sepharose CL-6B (Qiagen) equilibrated in lysis buffer. The matrix was then washed with 2 CV of 3 M NaCl, followed by 2 CV of wash buffer (20 mM NaP pH 7.4, 800 mM NaCl, 20 mM imidazole, 3 mM DTT, 150 mM L-arginine, 150 mM L-glutamate). Protein was eluted with 4 to 5 CV of elution buffer (20 mM NaP pH 7.4, 800 mM NaCl, 500 mM imidazole, 3 mM DTT, 150 mM L-arginine, 150 mM L-glutamate). The protein was dialysed against 2 L of 20 mM NaP pH 7.4, 800 mM NaCl, 3 mM DTT, 150 mM L-arginine, 150 mM L-glutamate using a 3.5 kDa MWCO regenerated cellulose membrane (Spectra/Por®,Spectrumlabs), incubating at 4 °C overnight. For phosphorylation reactions of proteins that would later be used in single-molecule experiments, the protein was further purified by IEX using a MonoS 5/50 GL column (GE Healthcare) equilibrated in 20 mM HEPES pH 8, 50 mM NaCl, 0.5 mM TCEP. The protein was eluted with a gradient of 0.05 to 1 M NaCl over 15 CV. Fractions containing protein were pooled, flash-frozen in liquid N_2 and stored at −80 °C.

6. Ulp1

The plasmid containing Ulp1-His_6 was transformed into *E. coli* BL21(DE3) and stored as a glycerol stock at −80 °C. For expression, 100 mL of 2xYT supplemented with 50 µg kanamycin were inoculated and incubated at 37 °C overnight. The overnight culture was diluted into 2 L of 2xYT and incubated until $OD600 \sim 0.6$. Protein expression was induced with 1 mM IPTG for ∼4 h at 30 °C. After harvesting at 4000 ×g, 4 °C, cell pellets were resuspended in wash buffer (50 mM Tris-HCl pH 8.0, 350 mM NaCl, 10 mM imidazole, 2 mM DTT, 2% glycerol) and lysed mechanically sing a high pressure homogenizer (HPL6, Maximator). The lysate was clarified by centrifugation at 35 000 ×g, 4 celsius and the supernatant was applied to a 5 mL HisTrap Excel column (GE Healthcare). The column was washed with 10 CV wash buffer, 5 CV wash buffer containing 1 M NaCl and again 10 CV of normal wash buffer. The protein was eluted using 50 mM Tris-HCl pH 8.0, 350 mM NaCl, 400 mM imidazole, 2 mM DTT, 2% glycerol. Fractions containing sufficiently pure protein were pooled and buffer exchanged into 50 mM Tris-HCl pH 8.0, 400 mM NaCl, 2 mM DTT, 2% glycerol using a HiPrep 26/10 Desalting column (Pharmacia). Fractions containing protein were combined and concentrated to ∼4 mg/mL by centrifugation (Vivaspin 20, 10 kDa MWCO, PES). The protein was diluted to ∼2 mg/mL by the addition of 100% glycerol, flash-frozen in liquid N_2 and stored at −80 °C.

B. Peptide synthesis and purification

1. Reagents and solvents

Fmoc- and side chain-protected L-amino acids (Fmoc-Arg(Pbf)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Pro-OH, Fmoc-Ser(*t*-Bu)-OH, Fmoc-Thr(*t*-Bu)-OH, Fmoc-Tyr(*t*-Bu)-OH) were purchased from the Novabiochem-line from Sigma-Aldrich Canada Ltd. Fmoc-L-Ser(PO(OBzl)OH)-OH was purchased from Iris Biotech GmbH. HATU (O-(7-azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate) and PyAOP (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate) were purchased from Advanced ChemTech CreoSalus. N,N-diisopropylethylamine (iPr_2NEt, DIPEA, 99.5 %), trifluoroacetic acid (TFA, for HPLC, ≥99.0 %), triisopropylsilane (TIPS, 98 %), 3,6-dioxa-1,8-octane-dithiol (DODT, 95 %), acetonitrile (MeCN, for HPLC gradient grade, ≥99.9 %) were purchased from Sigma-Aldrich. N,N-Dimethylformamide (DMF) was purchased from the Supelco-line of Sigma-Aldrich Canada Ltd. Dichloromethane (DCM, ≥99.8 %) was purchased from Fisher Scientific Ltd. Diethyl ether was purchased from Honeywell Riedel-de Haën. NovaPEG Rink Amide resin (loading 0.20 mmol/g) was purchased from the Novabiochem-line of Sigma-Aldrich Canada Ltd.

2. Automated fast-flow peptide synthesis (AFPS)

Peptides were synthesized on an automated-flow system built in the Hartrampf lab that is similar to a previously published AFPS system [8]. The following settings were used for peptide synthesis: a flow rate of 20 mL/min for coupling and deprotection steps, the reactor was kept at 90 °C, the heating loop at 90 °C or 30 °C (specified below). In each synthetic cycle the resin is pre-washed at 90 °C for 60 s at 40 mL/min. During the coupling step, three HPLC pumps are used: a 50 mL/min pump head pumps the activating agent, a second 50 mL/min pump head pumps the amino acid, and a 5.0 mL/min pump head pumps DIPEA (*neat*). The 50 mL/min pump head delivered 0.398 679 mL of liquid per pump stroke, the 5.0 mL/min pump head pumps 0.039 239 mL of liquid per pump stroke.

All peptides were prepared by AFPS and were synthesized on commercially available Novabiochem® NovaPEG Rink Amide resin (0.20 mmol/g, exact amounts of resin used for each peptide are listed in Section Section III B 7) and standard Fmoc/*t*-Bu protected amino acids (0.40 M in DMF, 0.20 M final concentration) were coupled using HATU (0.38 M in DMF, 0.19 M final concentration) or PyAOP (0.38 M in DMF, 0.19 M final concentration) with DIPEA (delivered *neat*, approx. 0.27 M final concentration) at a total flow rate of 20 mL/min. Amino acids G, P, S, phosphoserine (pS), and Y were coupled using HATU. Amino acids C, H, N, R and T were coupled using PyAOP. For amino acids G, P, S, and Y, a total volume of 6.4 mL of the "coupling solution" (i.e. containing amino acid at 0.20 M, HATU or PyAOP at 0.19 M and DIPEA at 0.27 M final concentration in DMF) was applied for each coupling. For amino acids C, H, N, R, S and T, a total of 10.4 mL of "coupling solution" was applied for each coupling. For pS, a total of 2.4 mL "coupling solution" was applied for each coupling unless stated otherwise. Removal of the N^{α}-Fmoc group was achieved using 20 % piperidine with 1 % formic acid in DMF (total volume 6.4 mL) at a flow rate of 20 mL/min. The deprotection solution was pre-heated at 30 °C for C, H and pS, and 90 °C for all other Fmoc-protected amino acids. Between each coupling and deprotection step, the resin was washed with DMF (32 mL) at a flow rate of 40 mL/min, preheated to 30 °C for amino acids C, H and pS, and at 90 °C for all others. Total synthesis time to afford resin-bound RS domain peptides was approximately 2.5 h. After completion of the peptide sequence, resins were manually washed with DCM (3 × 5 mL) and dried under reduced pressure.

3. Peptide cleavage and deprotection

Exact amounts of resin taken forward for cleavage are listed in Section Section III B 7. Peptides were cleaved using a solution of TFA/TIPS/DODT/H_2O (94:1:2.5:2.5, v/v/v/v, 1 to 3 mL) for 2 h at 23 °C with gentle mixing. TFA was removed by evaporation under a light stream of N_2, and peptides were precipitated and isolated by centrifugation from ice-cold diethyl ether using 2 × 15 mL. The resulting

peptide pellets were briefly dried under a light stream of N_2, dissolved in an aqueous solution containing 10 to 50 % MeCN and 0.1 % TFA, and lyophilized. Crude peptides were analyzed for purity by LCMS and UHPLC.

4. Analytical Ultra-High Performance Liquid Chromatography (UHPLC)

Filtered peptide solutions were diluted in 500 µL of 10 to 50 % acetonitrile (MeCN) in water with 0.1 % TFA to a final concentration of approximately 0.5 mg/mL. To achieve the best possible resolution for each peptide, the samples were applied to either an Agilent ZORBAX RRHD Eclipse Plus C8 column (2.1 mm×50 mm, 1.8 µm particle size, hereafter referred to as column A), an Agilent ZORBAX RRHD 300Å StableBond C18 column (2.1 mm×100 mm, 1.8 µm particle size, hereafter referred to as column B), or an Agilent ZORBAX StableBond 300 C18 column (2.1 mm×150 mm, 5 µm particle size, hereafter referred to as column C), connected to an UHPLC system of the Agilent 1290 Infinity II Series controlled by Agilent OpenLab CDS and ChemStation software. The column was kept at 40 °C, the flow rate was 0.80 mL/min, and peptide absorbance was detected at 214 nm. A binary solvent system consisting of Solvent A (5 % MeCN in 95 % water with 0.1 % TFA) and Solvent B (95 % MeCN containing 5 % water and 0.1 % TFA) was used. After sample application, the column was washed with 100 % Solvent A for 1.5 min before peptides were eluted with a linear gradient corresponding to 5 to 95 % MeCN or 0 to 100 % MeCN over varying lengths of time (exact gradients are listed in the corresponding figure legends in Section Section III B 7). At the end of the gradient, the column was first washed in 100 % Solvent B at 0.3 mL/min for 2 min before re-equilibration in 100 % Solvent A for 5 min. Using the ChemStation software, purities of crude and purified peptides were determined by integration of all UV signals at 214 nm (integration intervals are listed in the corresponding figure legends in Section Section III B 7). Purities are given in percent of the product peak relative to all other integrated absorbances.

5. Liquid Chromatography with High-Resolution Electrospray Ionization Mass Spectrometry (LCMS)

To determine peptide mass and purity by LCMS, filtered peptide solutions were diluted in either 10 to 50 % acetonitrile (MeCN) in water with 0.1 % TFA (60 to 500 µL), or in guanidine hydrochloride (6.0 M) to a final concentration of approximately 0.1 mg/mL. The samples were analyzed on an Acquity UHPLC (Waters, Milford, USA) connected to an Acquity eλ diode array detector and a Synapt G2HR-ESI-QTOF-MS (Waters, Milford, USA). For standard analysis of all peptides, samples were applied to an Acquity BEH C8 HPLC column (2.1 mm×100 mm, 1.7 µm particle size, Waters) kept at 30 °C at a flow rate of 0.4 mL/min. UV absorbances were monitored at 190 to 300 nm at 1.2 nm resolution and 20/s. The binary solvent system used for LC-MS consisted of water containing 0.02 % formic acid and 0.04 % TFA (Solvent A) and MeCN containing 0.04 % formic acid and 0.02 % TFA (Solvent B). All samples were analyzed using one of the following gradients:

- **LC-MS Gradient A:** isocratic step at 10 % Solvent B for 3 min, linear gradient of 10 to 70 % Solvent B over 9 min, isocratic step at 70 % Solvent B for 1 min
- **LC-MS Gradient B:** isocratic step at 3 % Solvent B for 3 min, linear gradient of 3 to 95 % Solvent B over 9 min, isocratic step at 95 % Solvent B for 1 min

The exact gradient used is specified in the corresponding figure legend for each peptide in Section Section III B 7. The mass spectrometer was set to positive ionization mode with a capillary voltage of 3.0 kV, sampling cone set to 40 V, extraction cone set to 4 V, N_2 cone gas at 4 L/h, N_2 desolvation gas at 800 L/min, and a source temperature set to 120 °C. The mass analyzer in resolution mode was used to resolve masses in a range of 150 to 3000 m/z with a scan rate of 1 Hz, mass calibration set to <2 ppm within 50 to 2500 m/z using an aqueous solution of 5 mM $NaHCO_2$. The lock masses used were caffeine (m/z = 195.088, 0.7 ng/mL) and leucine-enkephalin (m/z = 556.2771, 2 ng/mL). All mass spectra shown are deconvolved masses calculated from raw m/z values using Mestrelab Research S.L.© MestReNova v. 14.1 Mnova MS Suite. Purity based on LC-MS was determined by calculating the area under the curve of the desired product peak as a percentage of the integral of all peaks within 2 to 10 min of the absorbance chromatogram at 214 nm.

6. Semi-Preparative Reverse-Phase High Performance Liquid Chromatography (RP-HPLC)

Semi-preparative RP-HPLC was performed on a Shimadzu prominence HPLC system (Shimadzu Corp., Japan) with a CBM-40 system controller module, an FRC-10A fraction collector, two LC-20AR pumps, and an SPD-40 UV/VIS detector, using either an Agilent Zorbax Eclipse XDB-C8 column (9.4 mm×250 mm, 5 µm particle size) or an Agilent Zorbax Eclipse XDB-C18 Semi-Preparative column (9.4 mm×250 mm, 5 µm particle size), kept at 23 °C, with a flow rate of 4 mL/min or 3.5 mL/min, respectively. A binary solvent system was used, wherein Solvent A was H_2O containing 0.1 % TFA, and Solvent B was MeCN containing 0.1 % TFA. Purifications were executed using the gradients specified for each peptide in Section Section III B 7. Fractions that contained peptide of the correct m/z and high purity, as determined by LC-MS, were combined and lyophilized to afford RS domain peptides as white amorphous solids.

7. Peptide synthesis of RS domain variants - peptide specific details

a. Synthetic RS domain The peptide corresponding to the RS domain of SRSF1 (residues 198-248) was synthesized on NovaPEG Rink Amide resin (144 mg, 29 µmol) using the standard AFPS protocol. Cleavage of the peptidyl-resin (94.2 mg, approx. 8.3 µmol) afforded the crude peptide as a colorless solid (23 mg, 79 % purity by LCMS, 72 % purity by UHPLC, Figures S11 and S12). The purification of the crude material (18 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C8 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 4 mL/min and a gradient of 0.5 %B/min over 42.5 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S13) were combined and lyophilized to afford RS domain peptide as a white amorphous solid (0.50 mg, 1.3 % yield based on resin loading, >95 % purity by UHPLC, Figure S14).

b. cysRScys The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines was synthesized on NovaPEG Rink Amide resin (156.6 mg, 31 µmol) using the standard AFPS protocol. Cleavage of the peptidyl-resin (50.3 mg, approx. 4.7 µmol) afforded the crude peptide as a colorless solid (8.4 mg, 52 % purity by LCMS, 71 % purity by UHPLC, Figures S15 and S16). The purification of the crude material was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C8 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 4 mL/min and a gradient of 1 %B/min over 62 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S17) were combined and lyophilized to afford cysRScys domain peptide as a white amorphous solid (>95 % purity by UHPLC, Figure S18). The yield was not determined.

c. RS-pS$_{199}$ The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines and bearing a phosphorylation at Ser199 was synthesized on NovaPEG Rink Amide resin (151.1 mg, 31 µmol) using the standard AFPS protocol. Fmoc-Ser(PO(OBzl)OH)-OH was incorporated at position Ser199. Cleavage of the peptidyl-resin (33 mg, approx. 2.9 µmol) afforded the crude peptide as a colorless solid (8.4 mg, 89 % purity by LCMS, 49 % purity by UHPLC, Figures S19 and S20). The purification of the crude material (8.4 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C8 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 4 mL/min and an initial isocratic gradient of 1 %B for 5 min followed by a gradient of 0.5 %B/min over 62 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S21) were combined and lyophilized to afford RS-pS$_{199}$ as a white amorphous solid (0.79 mg, 4.5 % yield based on resin loading, >95 % purity by UHPLC, Figure S22).

d. RS-pS$_{223}$ The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines bearing a phosphorylation at Ser223 was synthesized on NovaPEG Rink Amide resin (148 mg, 30 µmol) using the standard AFPS protocol. Fmoc-Ser(PO(OBzl)OH)-OH was incorporated at position Ser223. Cleavage of the peptidyl-resin (64 mg, approx. 5.4 µmol) afforded the crude peptide as a colorless solid (6.2 mg, 59 % purity by LCMS, 51 % purity by UHPLC, Figures S23 and S24). The purification of the crude material (3.1 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C8 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 4 mL/min and an initial isocratic gradient of 1 %B for 5 min followed by a gradient of 0.5 %B/min over 60 min starting at

1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S25) were combined and lyophilized to afford RS-pS_{223} as a white amorphous solid (0.90 mg, 5.3 % yield based on resin loading, 95 % purity by UHPLC, Figure S26).

e. RS-pS_{225} The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines bearing a phosphorylation at Ser225 was synthesized on NovaPEG Rink Amide resin (162.8 mg, 33 µmol) using the standard AFPS protocol. Fmoc-Ser(PO(OBzl)OH)-OH was incorporated at position Ser225. Cleavage of the peptidyl-resin (approx. 70 mg, approx. 5.4 µmol) afforded the crude peptide as a colorless solid (12.7 mg, mass determined by LCMS, 32 % purity by UHPLC, Figures S27 and S28). The purification of the crude material (12.6 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C8 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 4 mL/min and an initial isocratic gradient of 1 %B for 5 min followed by a gradient of 0.5 %B/min over 60 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S29) were combined and lyophilized to afford RS-pS_{225} as a white amorphous solid (0.6 mg, $\sim$ 2 % yield based on resin loading, >95 % purity by UHPLC, Figure S30).

f. RS-2×pS The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines bearing phosphorylations at Ser223 and Ser225 was synthesized on NovaPEG Rink Amide resin (150 mg, 30 µmol) using the standard AFPS protocol. Fmoc-Ser(PO(OBzl)OH)-OH was incorporated at positions Ser223 and Ser225. Cleavage of the peptidyl-resin (32 mg, approx. 2.9 µmol) afforded the crude peptide as a colorless solid (7.5 mg, 24 % purity by LCMS, 24 % purity by UHPLC, Figures S31 and S32). The purification of the crude material (7.4 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C18 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 3.5 mL/min and an initial isocratic gradient of 1 %B for 5 min followed by a gradient of 1.40 %B/min over 43 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S33) were combined and lyophilized to afford RS-2×pS as a white amorphous solid ($\sim$ 0.6 mg, $\sim$ 3 % yield based on resin loading, >95 % purity by UHPLC, Figure S34).

g. RS-4×pS The peptide corresponding to the RS domain of SRSF1 (residues 198-248) with terminal cysteines bearing phosphorylations at Ser199, Ser207, Ser217 and Ser225 was synthesized on commercially available NovaPEG Rink Amide resin (165.8 mg, 33 µmol) using the standard AFPS protocol. Since crude purity of the 2xpS peptide was already reduced three-fold compared to unphosphorylated peptides, Fmoc-Ser(PO(OBzl)OH)-OH was incorporated at positions Ser199, Ser207, Ser217 and Ser225 with 10.4 mL of "coupling solution". Cleavage of the peptidyl-resin (228 mg, approx. 1.7 µmol) afforded the crude peptide as a colorless solid (57 mg, mass determined by LCMS, 28 % purity by UHPLC, Figures S35 and S36). The purification of the crude material (29 mg) was carried out by semi-preparative RP-HPLC using an Agilent Eclipse C18 column (5 µm, 9.4 mm × 250 mm) at 60 °C with a flow rate of 3.5 mL/min and an initial isocratic gradient of 1 %B for 5 min followed by a gradient of 0.60 %B/min over 42.5 min starting at 1 %B. Fractions identified with the correct m/z and high purity by LCMS analysis (Figure S37) were combined and lyophilized to afford RS-4×pS as a white amorphous solid (1.005 mg, 1.8 % yield based on resin loading, >95 % purity by UHPLC, Figure S38).

IV. REFERENCES

[1] De Silva, N. I. U.; Lehman, N.; Fargason, T.; Paul, T.; Zhang, Z.; Zhang, J. Unearthing a novel function of SRSF1 in binding and unfolding of RNA G-quadruplexes. *Nucleic Acids Research* **2024**, *52*, 4676–4690.

[2] Holmstrom, E. D.; Liu, Z.; Nettels, D.; Best, R. B.; Schuler, B. Disordered RNA chaperones can enhance nucleic acid folding via local charge screening. *Nature Communications* **2019**, *10*, 2453.

[3] Chavali, S. S.; Cavender, C. E.; Mathews, D. H.; Wedekind, J. E. Arginine Forks Are a Widespread Motif to Recognize Phosphate Backbones and Guanine Nucleobases in the RNA Major Groove. *Journal of the American Chemical Society* **2020**, *142*, 19835–19839.

[4] Viader-Godoy, X.; Manosas, M.; Ritort, F. Stacking correlation length in single-stranded DNA. *Nucleic Acids Research* **2024**, *52*, 13243–13254.

[5] McIntosh, D. B.; Duggan, G.; Gouil, Q.; Saleh, O. A. Sequence-Dependent Elasticity and Electrostatics of Single-Stranded DNA: Signatures of Base-Stacking. *Biophysical Journal* **2014**, *106*, 659–666.

[6] Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **2024**, *630*, 493–500.

[7] Hellenkamp, B. *et al.* Precision and accuracy of single-molecule FRET measurements—a multi-laboratory benchmark study. *Nature Methods* **2018**, *15*, 669–676.

[8] Hartrampf, N. *et al.* Synthesis of proteins by automated flow chemistry. *Science* **2020**, *368*, 980–987.

Appendix A: Peptide synthesis data

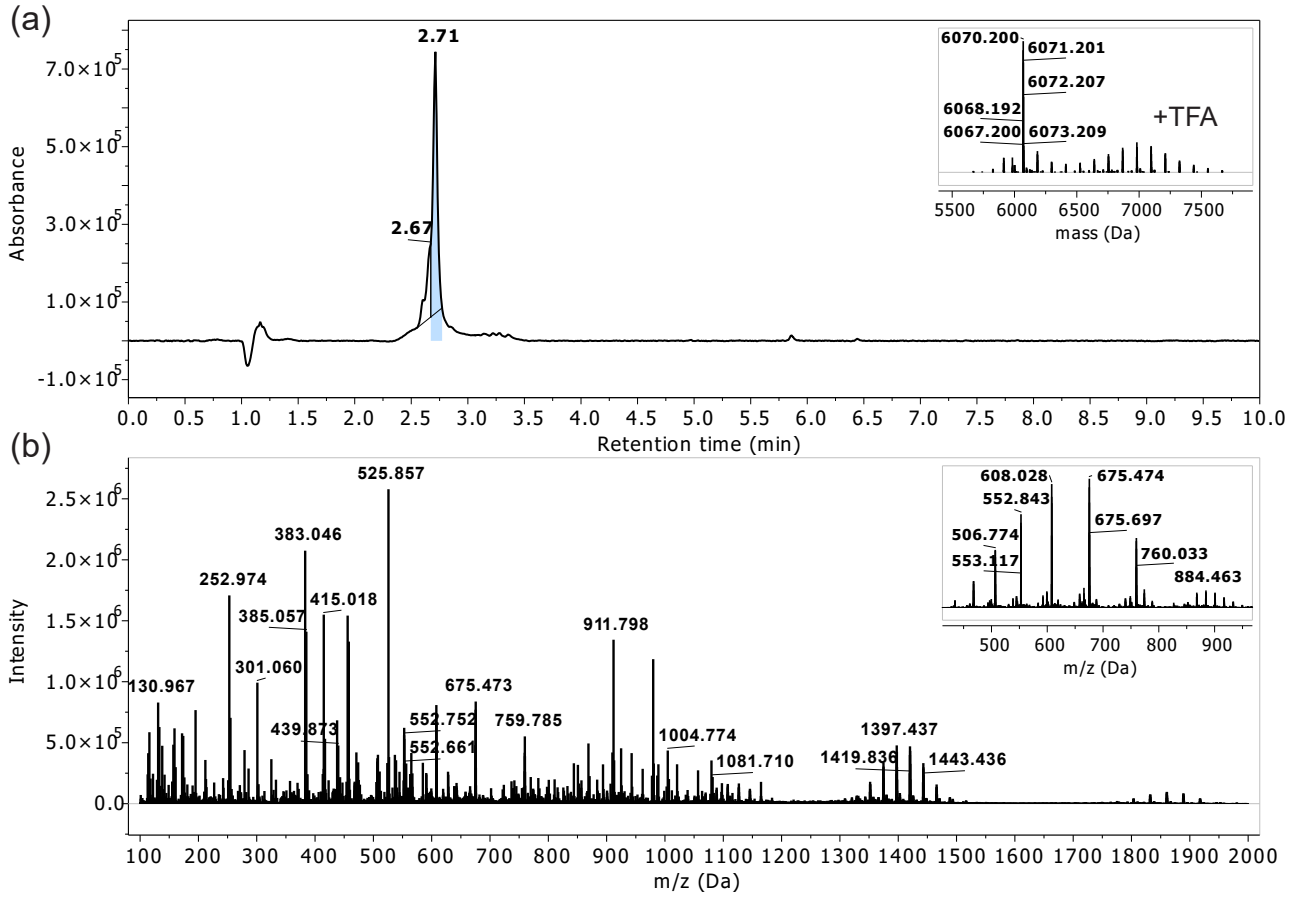

FIGURE S11. LCMS profile of crude RS-synthetic. (a) Absorbance chromatogram of RS at $\lambda = 214\,\text{nm}$ with a retention time (Rt) of 2.71 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.71 min). Monoisotopic mass (ESI+) calculated for $C_{239}H_{412}N_{112}O_{76}$ is 6067.1817 Da; found 6067.1885 Da. LCMS Gradient A. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

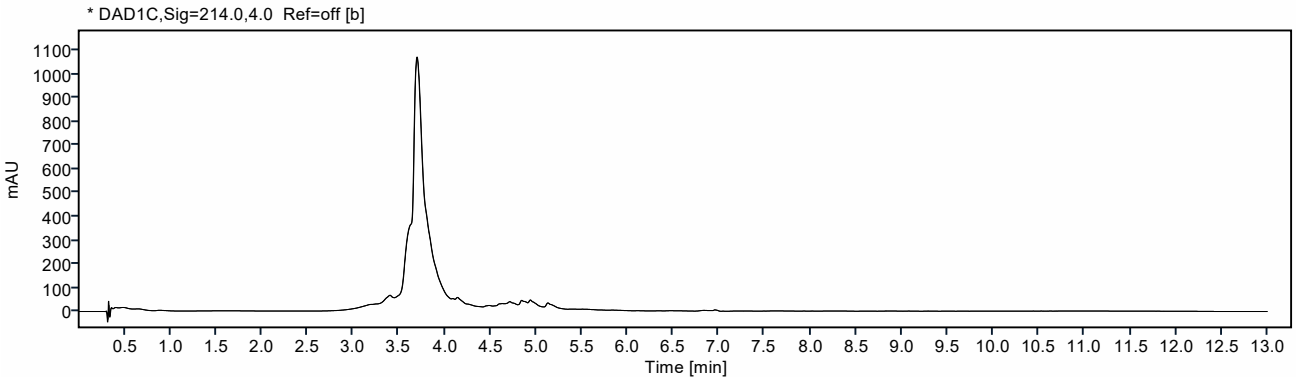

FIGURE S12. UHPLC profile of crude RS-synthetic using column A and a gradient of 5 to 95 % MeCN over 10 min at 40 °C. The peptide eluted at 3.70 min. Integration of the absorbance at $\lambda = 214\,\text{nm}$ between 1 to 10 min yields 72 % purity.

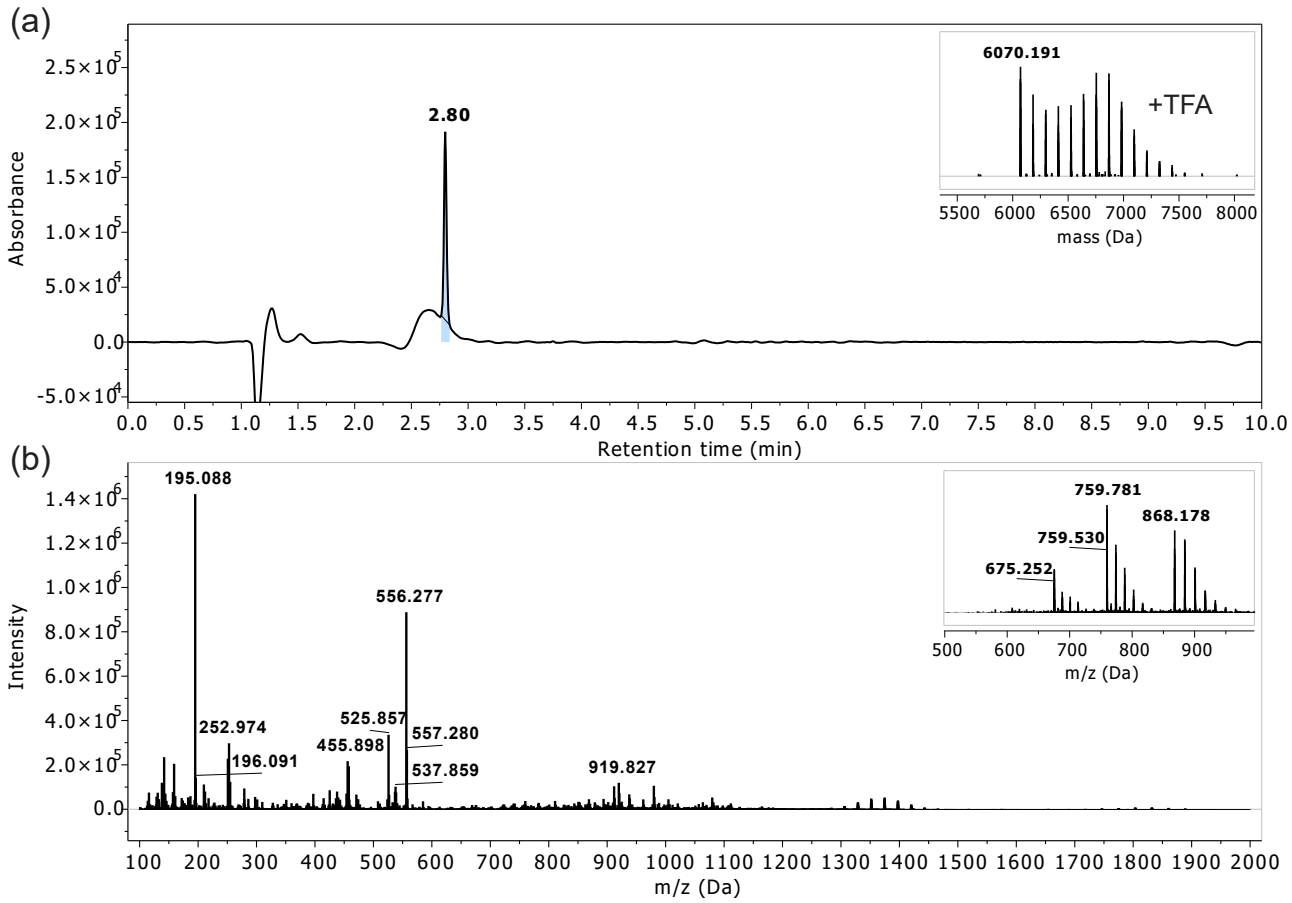

FIGURE S13. LCMS profile of pure RS-synthetic. (a) Absorbance chromatogram of pure RS at $\lambda = 214$ nm with an Rt of 2.80 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra of RT 2.80 min). Monoisotopic mass (ESI+) calculated for $C_{239}H_{412}N_{112}O_{76}$ 6067.1817 Da, found 6067.1768 Da. LCMS Gradient A. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

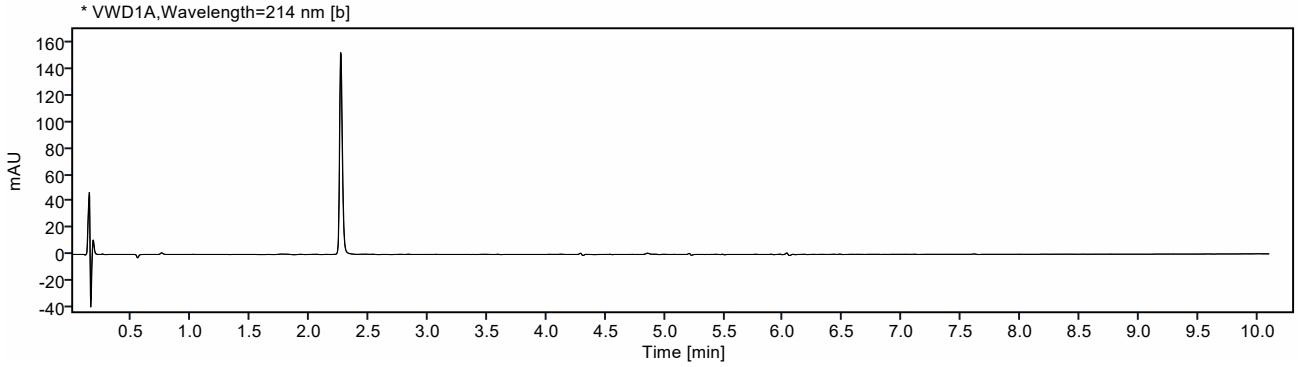

FIGURE S14. UHPLC profile of pure RS suing column B and a gradient of 0 to 100 % MeCN over 10 min at 40 °C. The peptide eluted at 2.36 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 10 min yields 99 % purity.

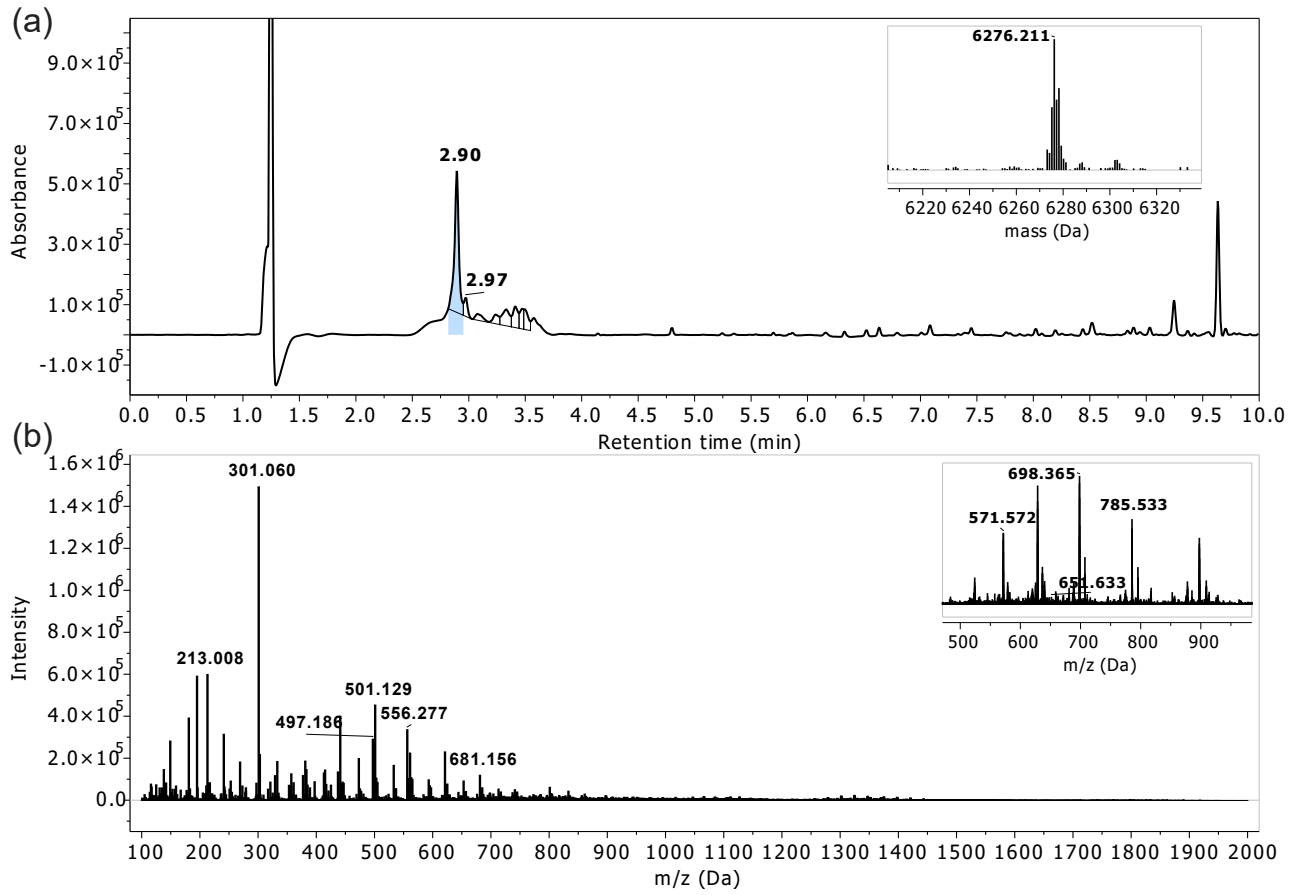

FIGURE S15. LCMS profile of crude cysRScys. (a) Absorbance chromatogram of cysRScys at $\lambda = 214$ nm with an Rt of 2.90 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.90 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{422}N_{114}O_{78}S_2$ is 6273.2001 Da; found 6273.1780 Da. LCMS Gradient B.

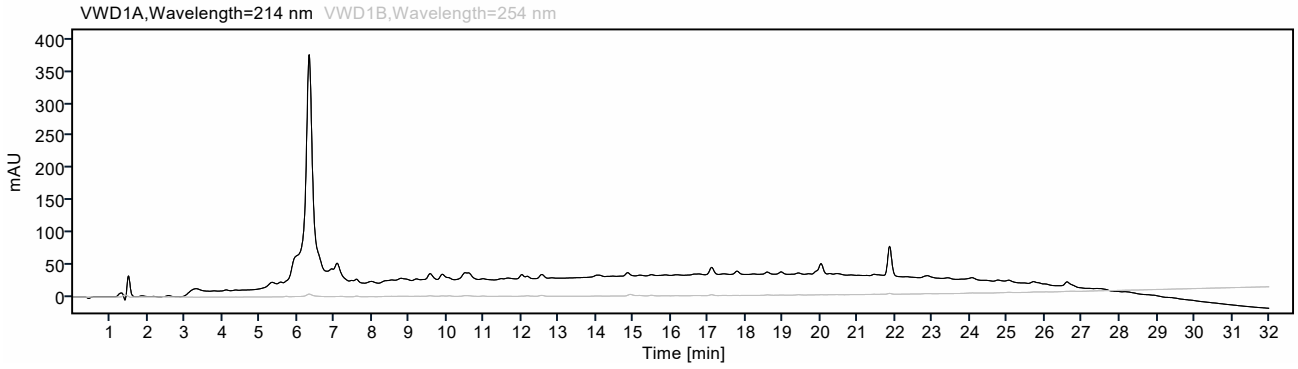

FIGURE S16. UHPLC profile of crude cysRScys using column C and a gradient of 0 to 100 % MeCN over 30 min at 40 °C. The peptide eluted at 6.34 min. Integration of the absorbance at $\lambda = 214$ nm between 2 to 30 min yields 71 % purity.

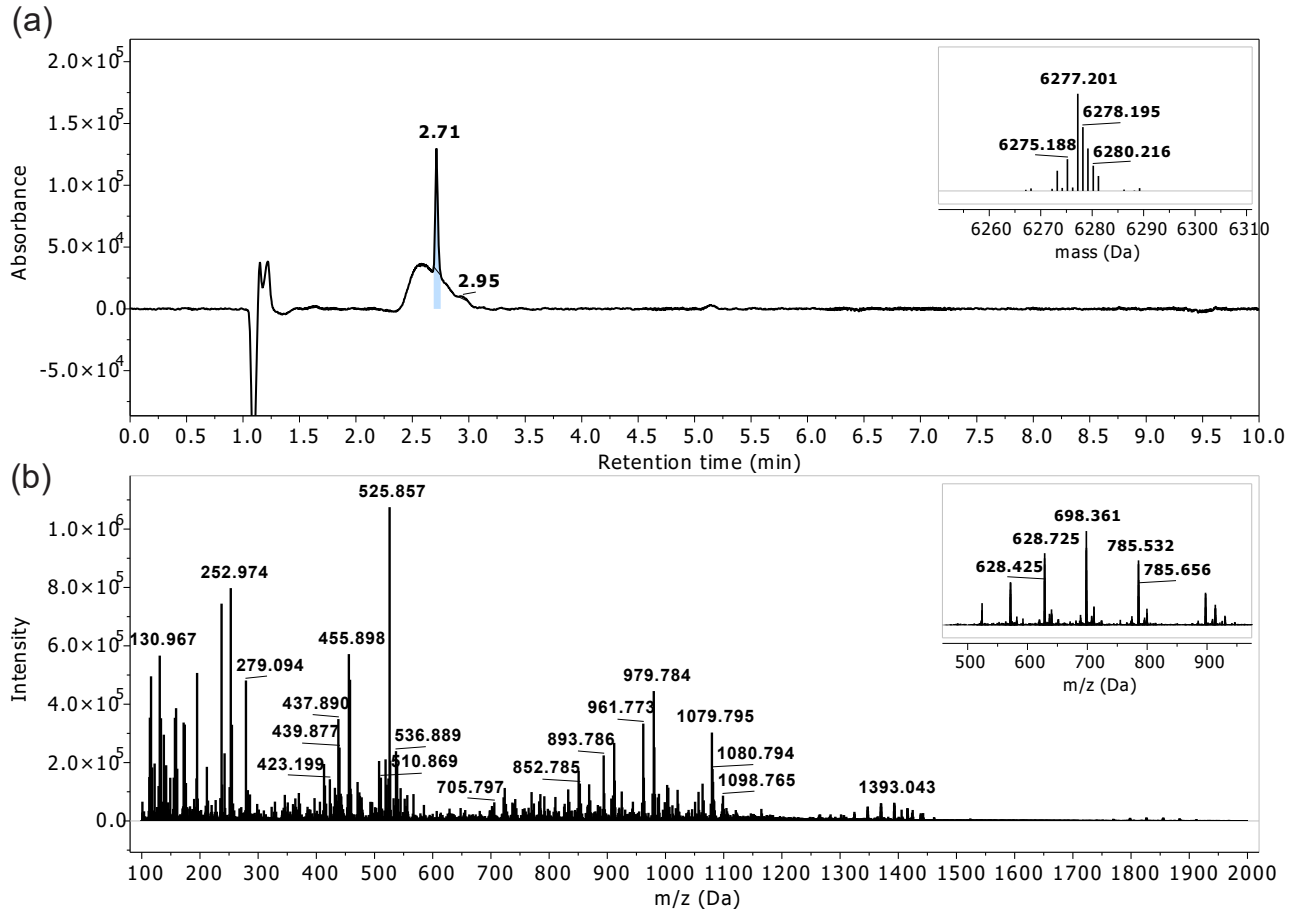

FIGURE S17. LCMS profile of pure cysRScys. (a) Absorbance chromatogram of cysRScys at $\lambda = 214\,\text{nm}$ with an Rt of 2.71 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.71 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{422}N_{114}O_{78}S_2$ is 6273.2001 Da; found 6273.1808 Da. LCMS Gradient B. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

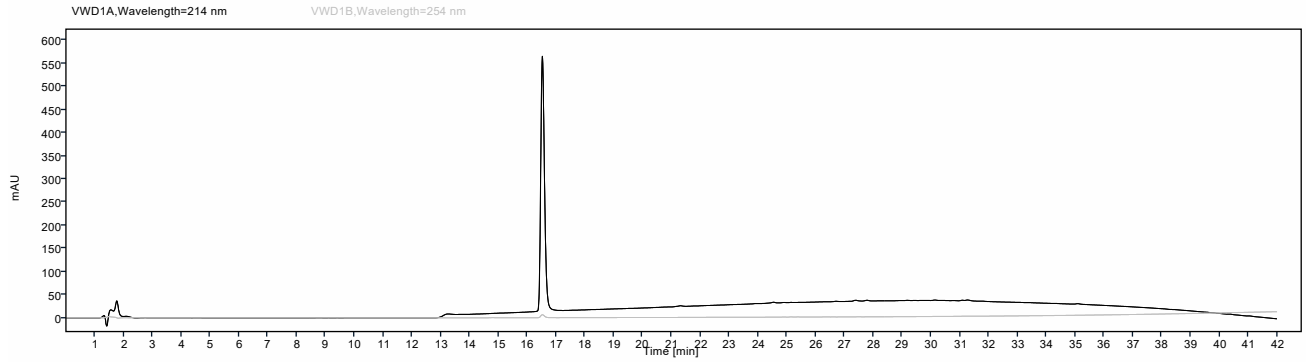

FIGURE S18. UHPLC profile of pure cysRScys using column C and a gradient of 0 to 100 % MeCN over 40 min at 40 °C. The peptide eluted at 16.52 min. Integration of the absorbance at $\lambda = 214\,\text{nm}$ between 3 to 40 min yields 99 % purity.

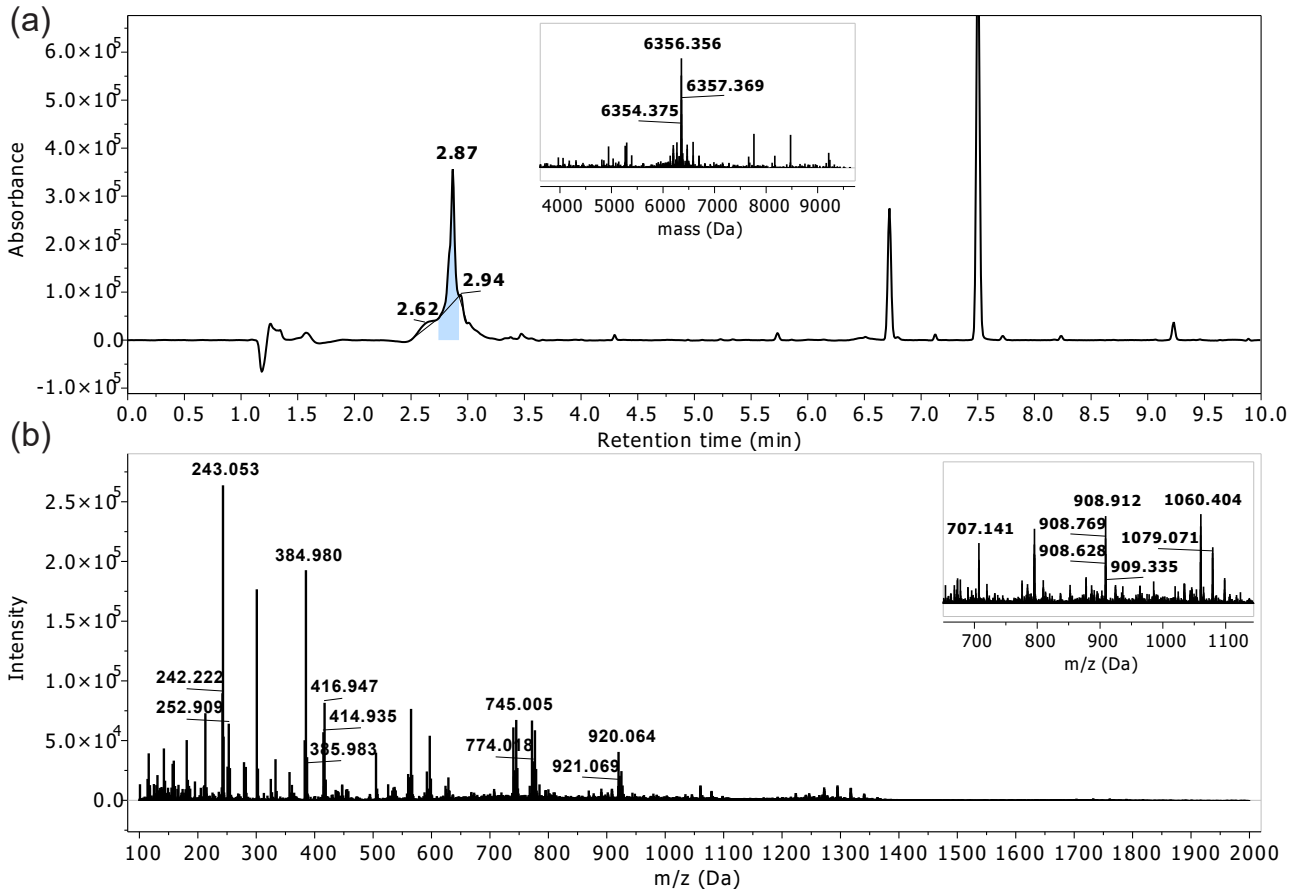

FIGURE S19. LCMS profile of crude RS-pS199. (a) Absorbance chromatogram of RS-pS199 at $\lambda = 214$ nm with an Rt of 2.87 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.87 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6353.0729 Da. LCMS Gradient A.

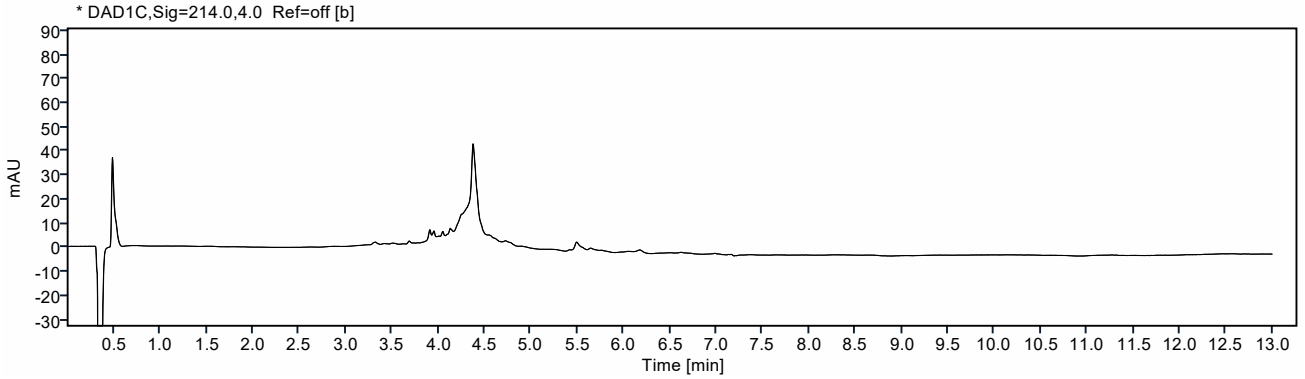

FIGURE S20. UHPLC profile of crude RS-pS199 using an column A and a gradient of 5 to 95 % MeCN over 10 min at 40 °C. The peptide eluted at 4.38 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 10 min yields 49 % purity.

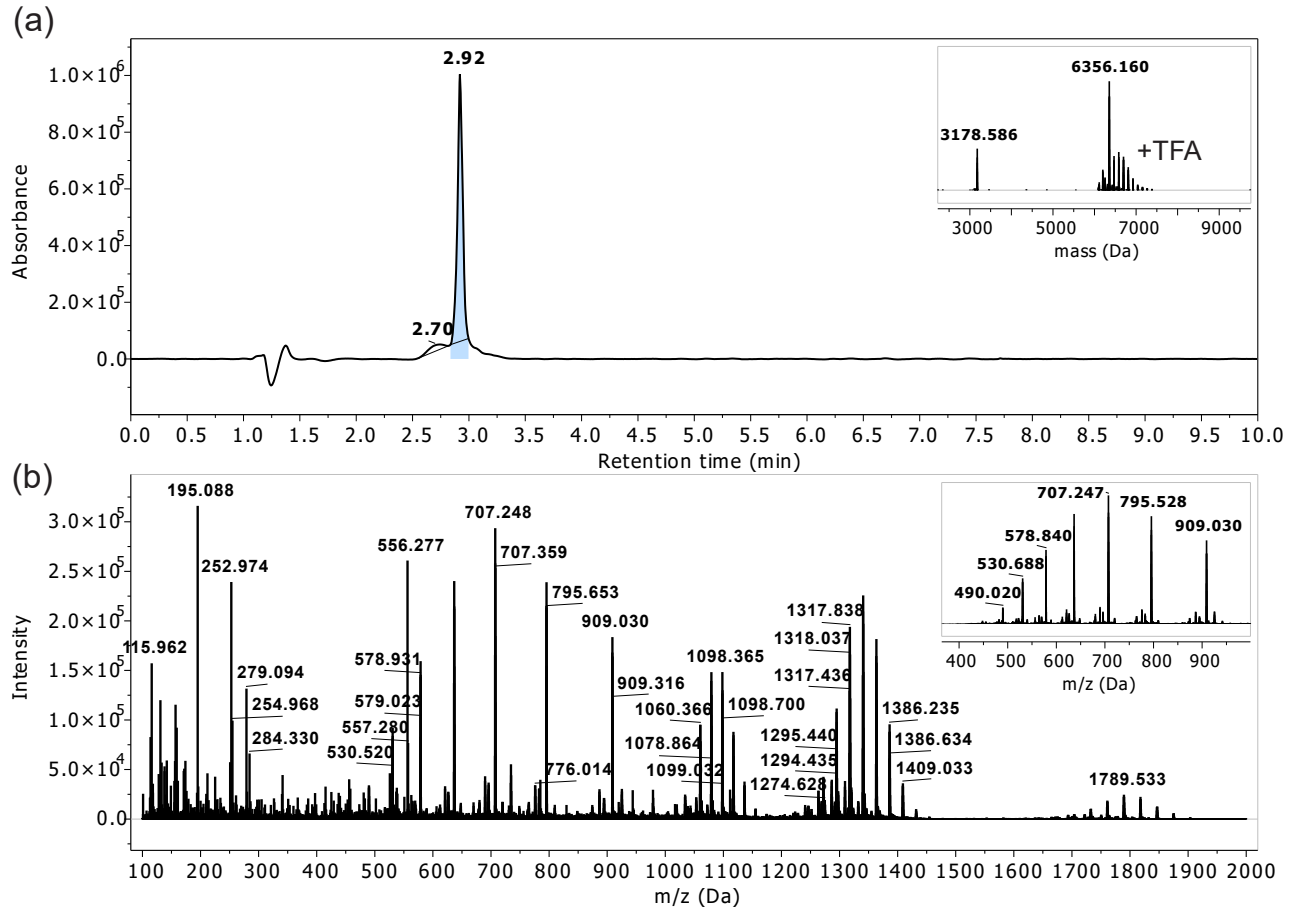

FIGURE S21. LCMS profile of pure RS-pS199. (a) Absorbance chromatogram of pure RS-pS199 at $\lambda = 214\,\text{nm}$ with an Rt of 2.92 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.92 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6354.1582 Da. LCMS Gradient A. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

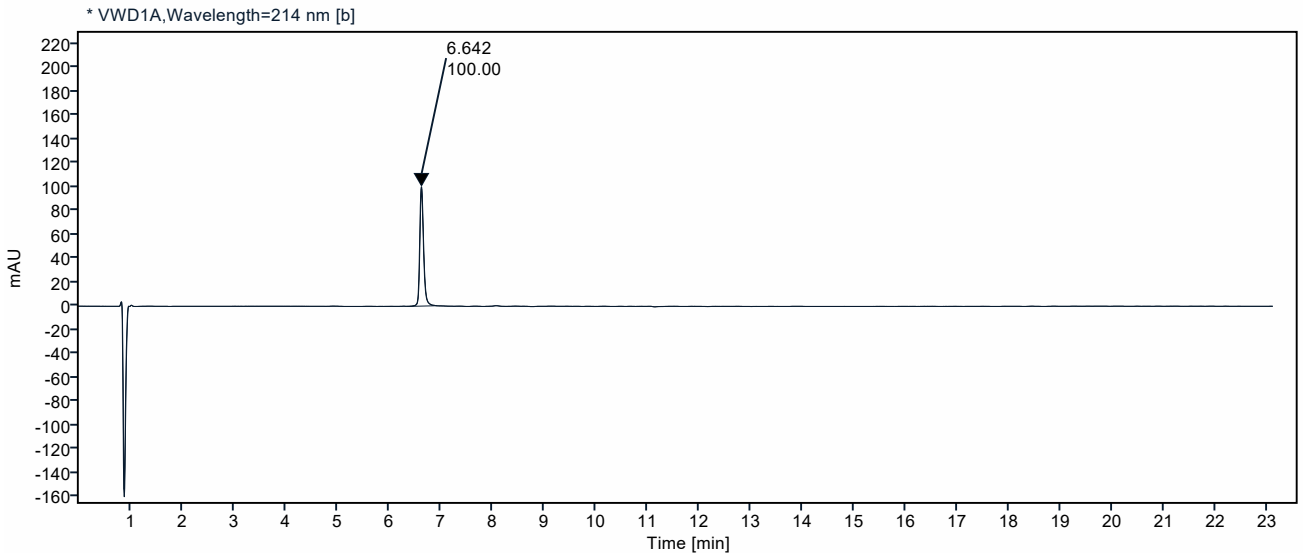

FIGURE S22. UHPLC profile of pure RS-pS199 using column C and a gradient of 0 to 100 % MeCN over 20 min at 40 °C. The peptide eluted at 6.642 min. Integration of the absorbance at $\lambda = 214\,\text{nm}$ between 2 to 20 min yields 100 % purity.

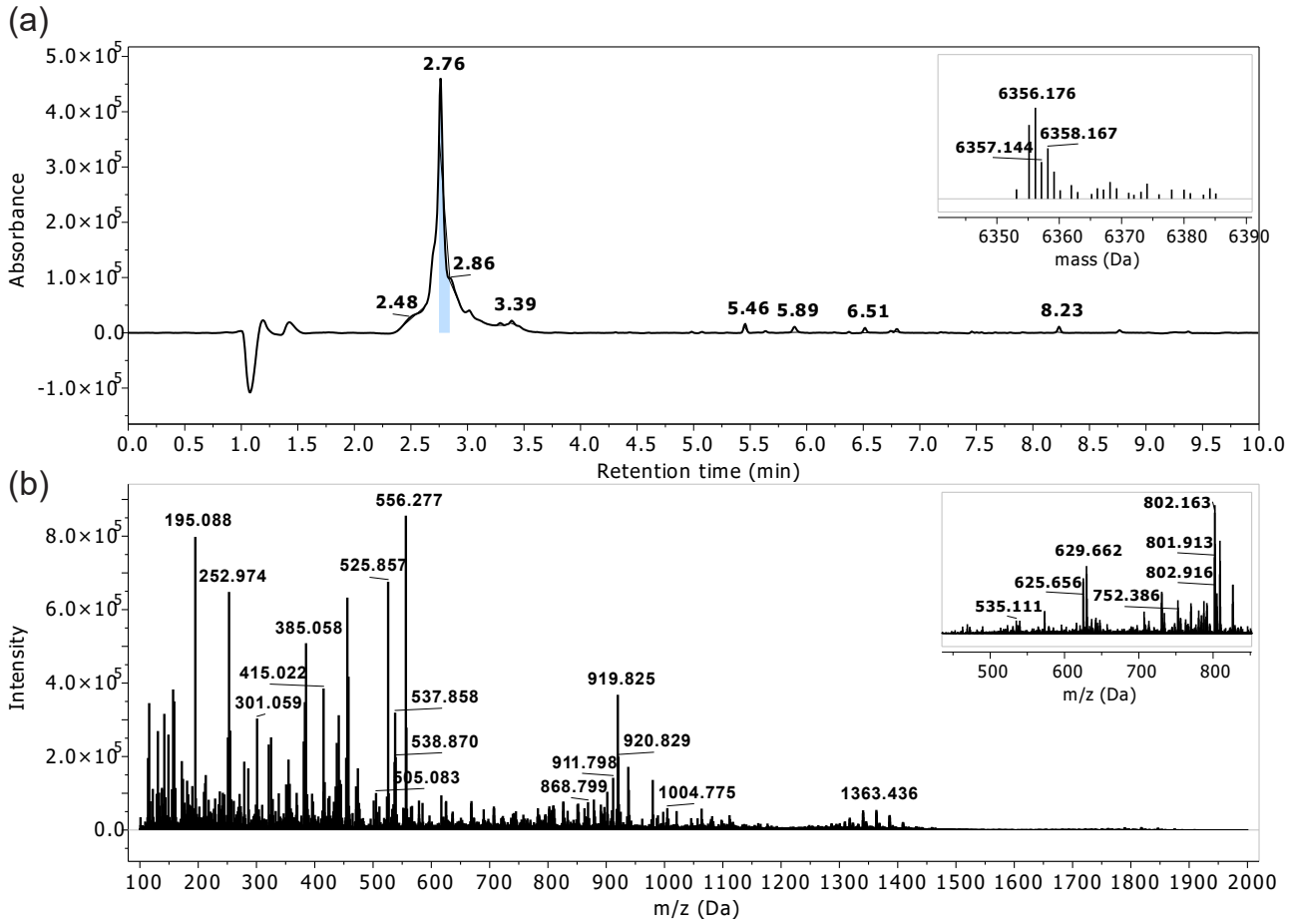

FIGURE S23. LCMS profile of crude RS-pS223. (a) Absorbance chromatogram of RS-pS223 at $\lambda = 214$ nm with an Rt of 2.76 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.76 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6353.1527 Da. LCMS Gradient B.

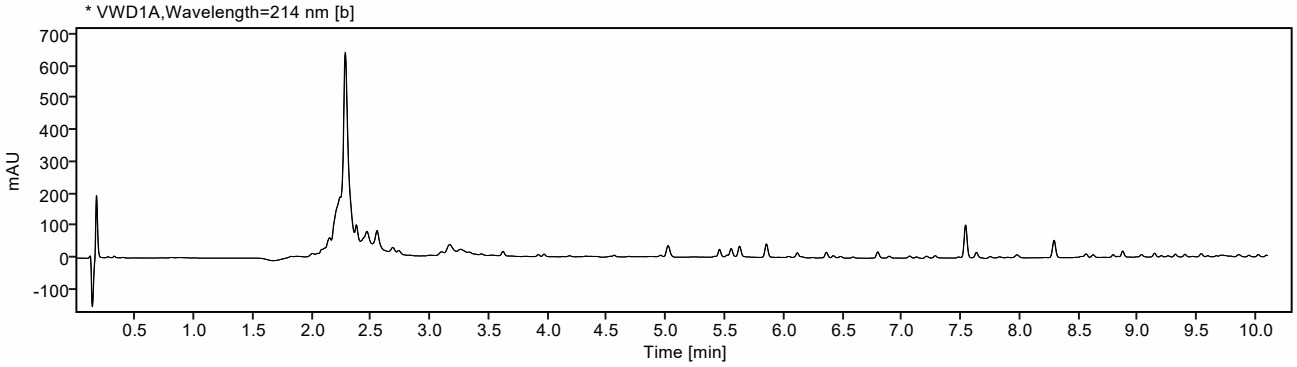

FIGURE S24. UHPLC profile of crude RS-pS223 using an column B and a gradient of 0 to 100 % MeCN over 10 min at 40 °C. The peptide eluted at 2.27 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 10 min yields 51 % purity.

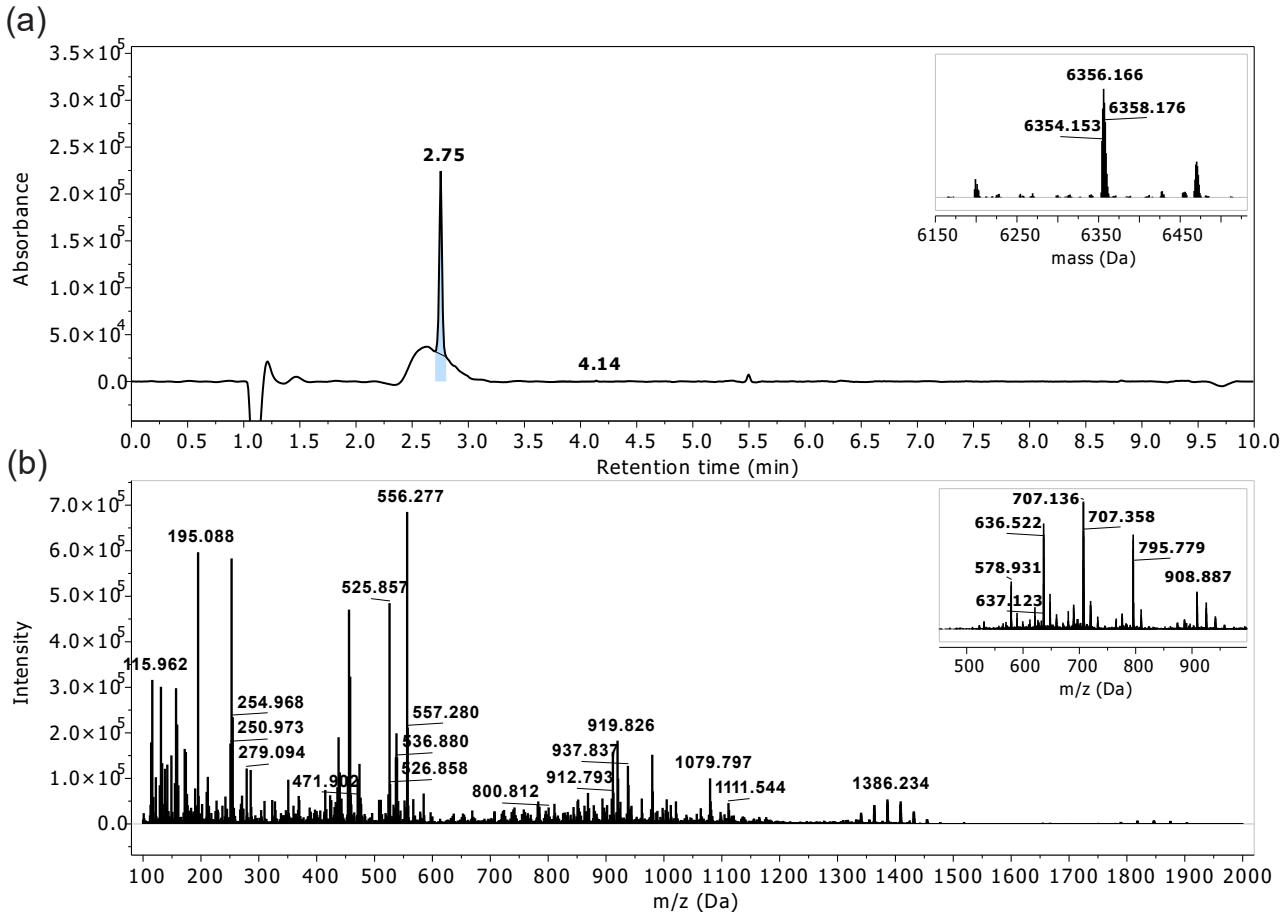

FIGURE S25. LCMS profile of pure RS-pS223. (a) Absorbance chromatogram of pure RS-pS223 at $\lambda = 214\,\text{nm}$ with an Rt of 2.75 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.75 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6353.1682 Da. LCMS Gradient B.

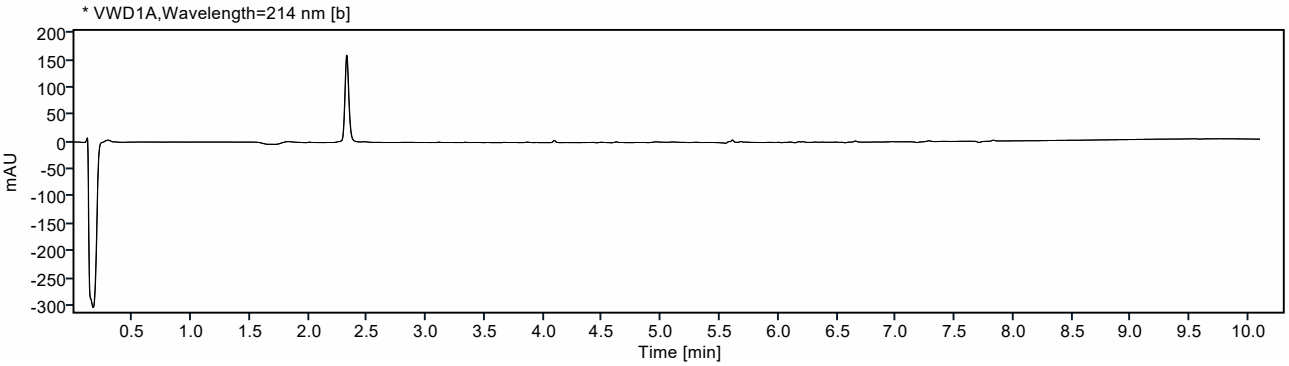

FIGURE S26. UHPLC profile of pure RS-pS223 using column B and a gradient of 0 to 100 % MeCN over 10 min at 40 °C. The peptide eluted at 2.33 min. Integration of the absorbance at $\lambda = 214\,\text{nm}$ between 1 to 10 min yields 95 % purity.

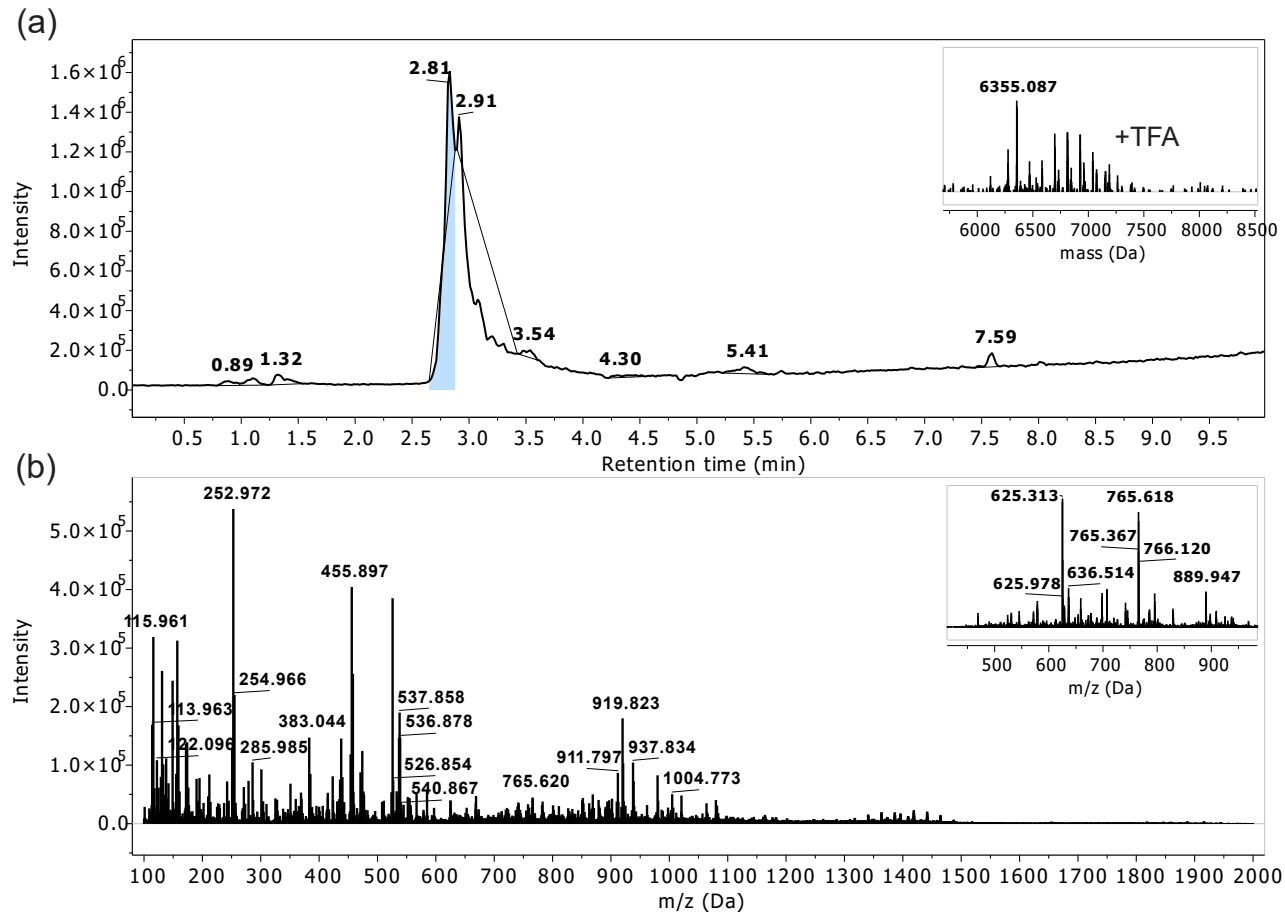

FIGURE S27. LCMS profile of crude RS-pS225. (a) Total Ion Chromatogram (TIC) of RS-pS225 at $\lambda = 214$ nm with an Rt of 2.81 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.81 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6353.077 Da. LCMS Gradient A. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

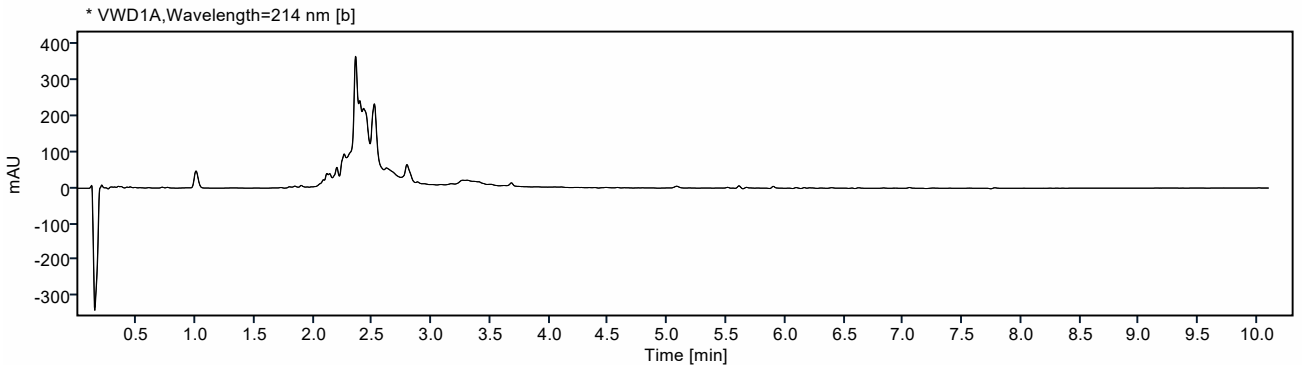

FIGURE S28. UHPLC profile of crude RS-pS225 using column B and a gradient of 0 to 100 % MeCN over 10 min at 40 °C. The peptide eluted at 2.36 min. Integration of the absorbance at $\lambda = 214$ nm between 0.5 to 10 min yields 32 % purity.

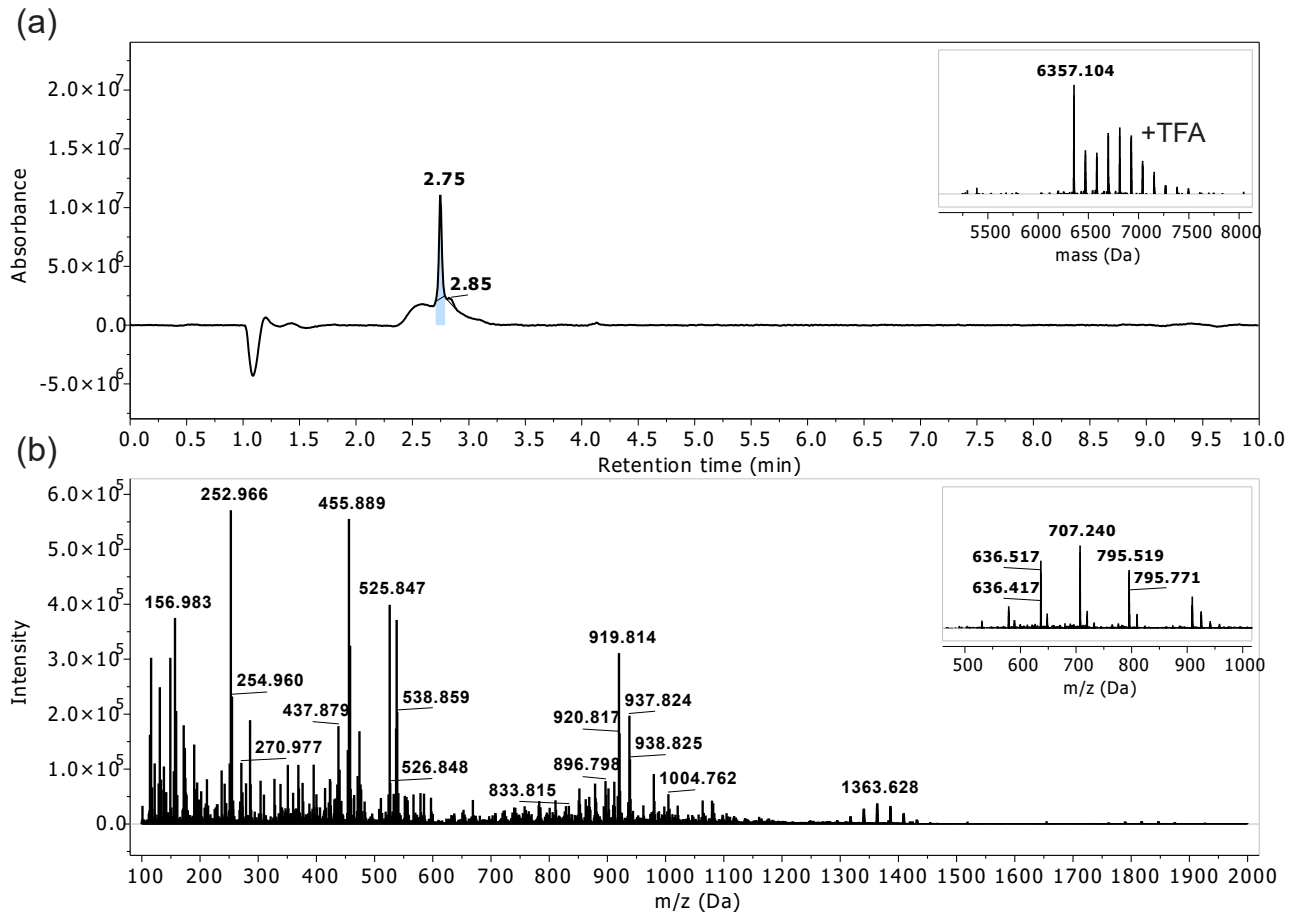

FIGURE S29. LCMS profile of pure RS-pS225. (a) Absorbance chromatogram of pure RS-pS225 at $\lambda = 214\,\text{nm}$ with an Rt of 2.75 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.75 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{423}N_{114}O_{81}PS_2$ is 6353.1701 Da; found 6353.0933 Da. LCMS Gradient B. Peptide ionizes with TFA, and mass corresponding to [Peptide + n*TFA] can be observed.

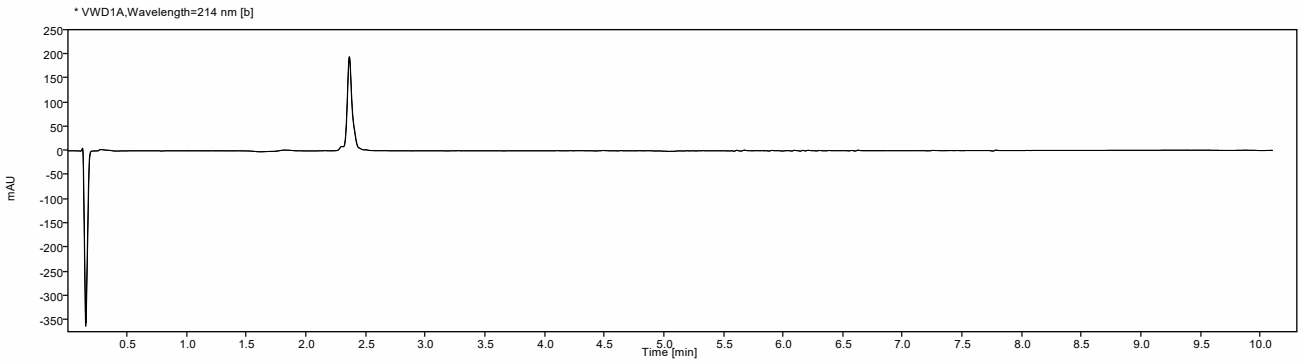

FIGURE S30. UHPLC profile of pure RS-pS225 using column B and a gradient of 0 to 100 % MeCN over 10 min at 40 °C. The peptide eluted at 2.36 min. Integration of the absorbance at $\lambda = 214\,\text{nm}$ between 1 to 10 min yields 96 % purity.

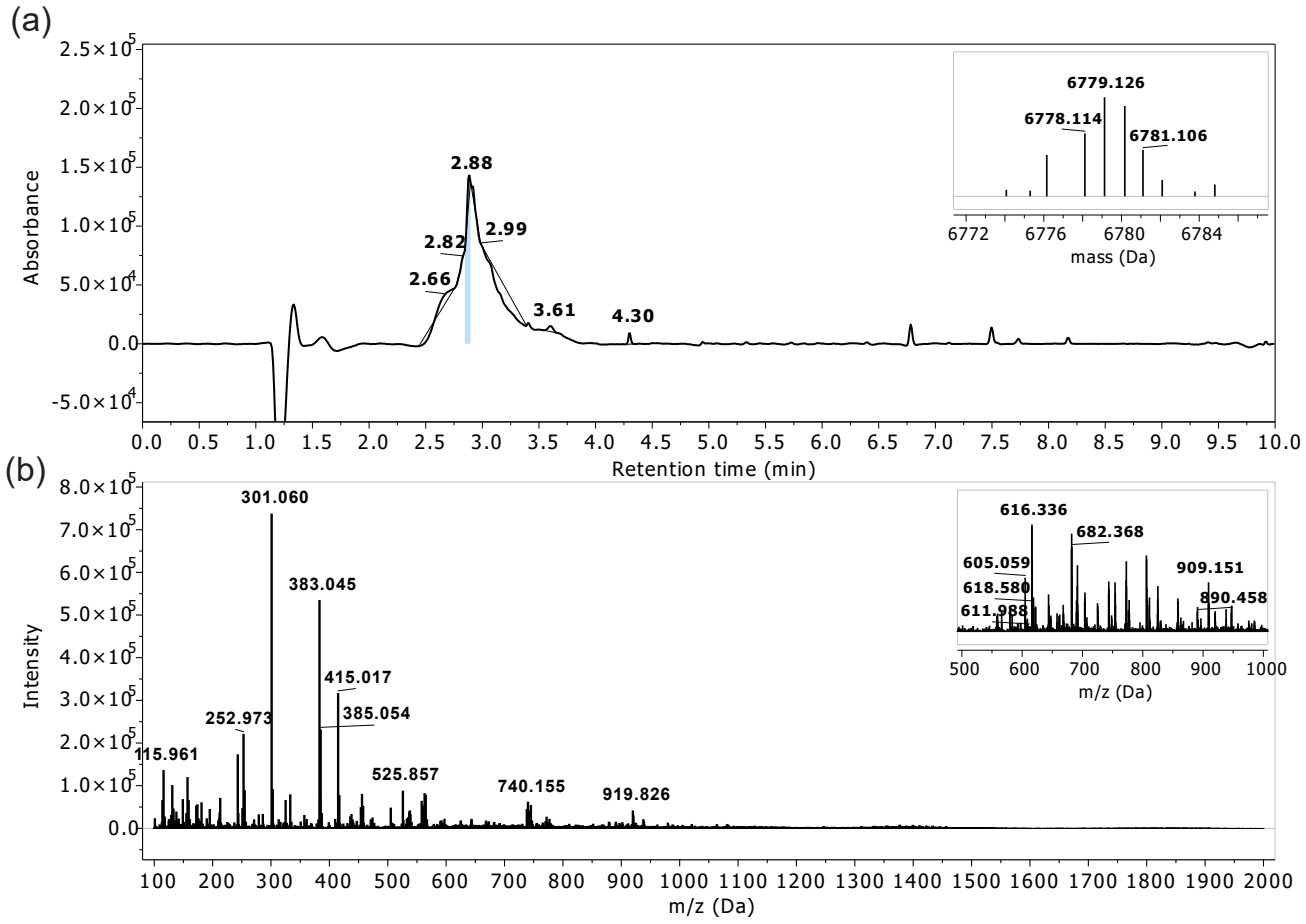

FIGURE S31. LCMS profile of crude RS-2×pS. (a) Absorbance chromatogram of RS-2×pS at $\lambda = 214$ nm with an Rt of 2.88 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.88 min). Monoisotopic mass (ESI+) calculated for $[C_{246}H_{424}N_{114}O_{85}P_2S_2$ + 2*TFA] is 6775.0401 Da; found 6775.324 Da. LCMS Gradient A.

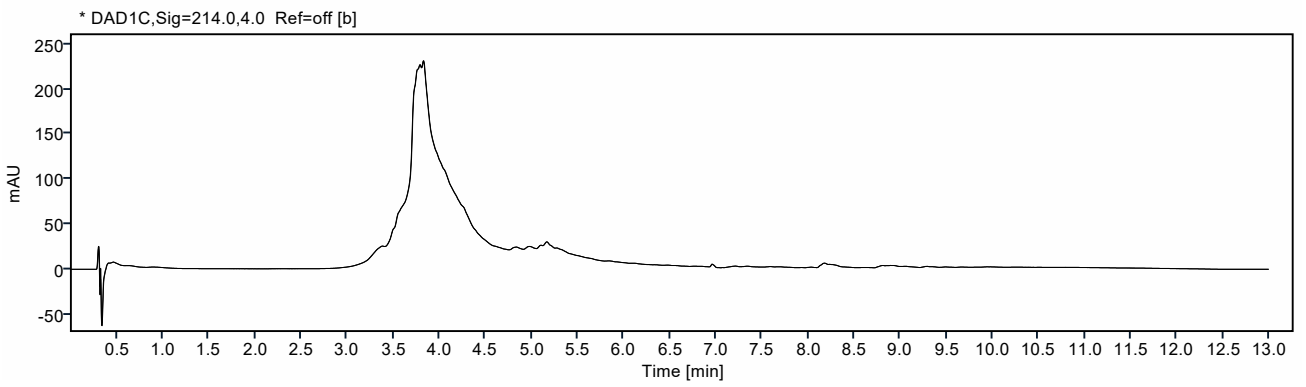

FIGURE S32. UHPLC profile of crude RS-2×pS using column A and a gradient of 5 to 95 % MeCN over 10 min at 40 °C. The peptide eluted at 3.79 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 10 min yields 24 % purity.

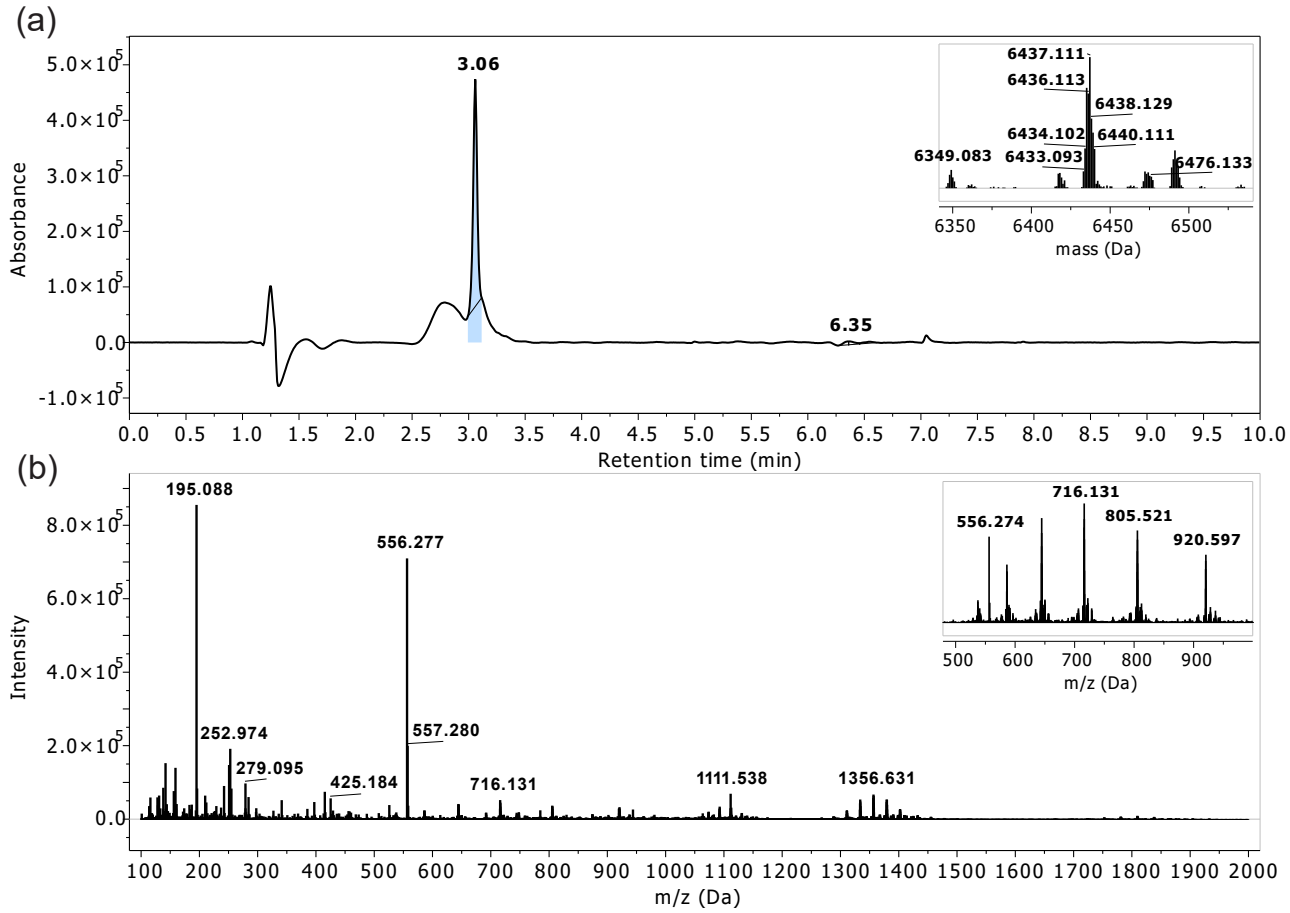

FIGURE S33. LCMS profile of pure RS-2×pS. (a) Absorbance chromatogram of pure RS-2×pS at $\lambda = 214$ nm with an Rt of 3.06 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 3.06 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{424}N_{114}O_{84}P_2S_2$ is 6433.1327 Da; found 6433.1041 Da. LCMS Gradient A.

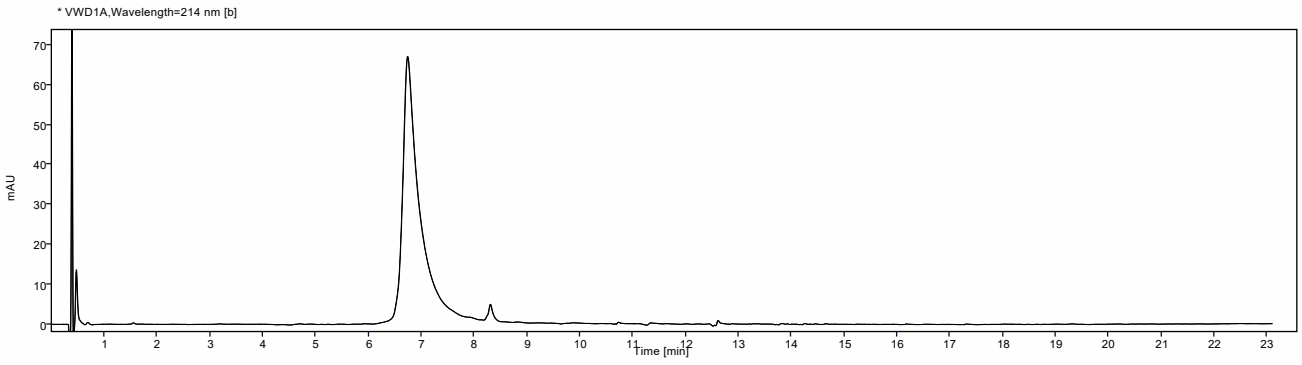

FIGURE S34. UHPLC profile of pure RS-2×pS using column B and a gradient of 0 to 100 % MeCN over 20 min at 40 °C. The peptide eluted at 6.74 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 20 min yields 98 % purity.

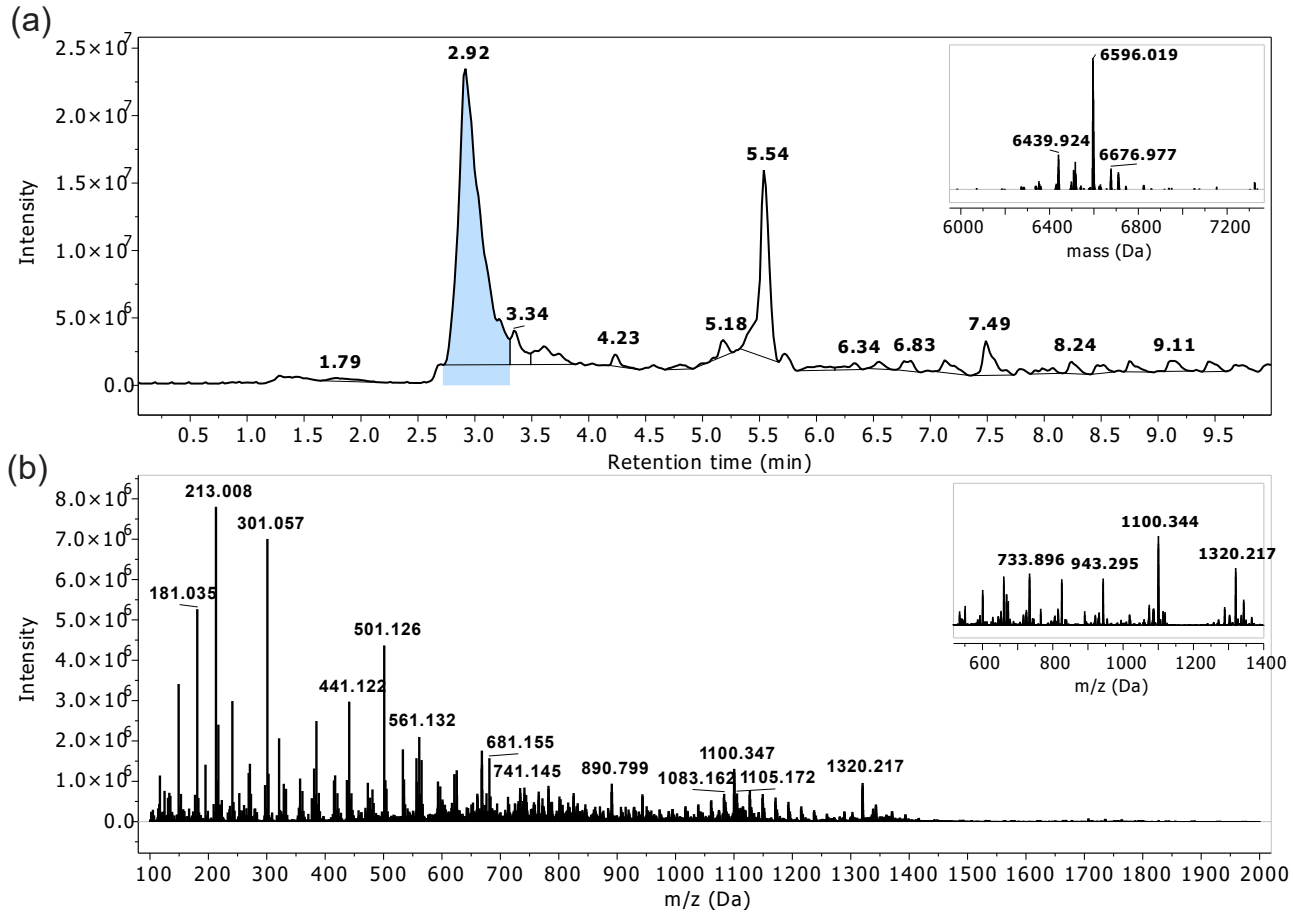

FIGURE S35. LCMS profile of crude RS-4×pS. (a) TI) of RS-4×pS at $\lambda = 214$ nm with an Rt of 2.92 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 10 min (insert: convoluted spectra at Rt 2.92 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{426}N_{114}O_{90}P_4S_2$ is 6593.0654 Da; found 6593.0037 Da. LCMS Gradient A.

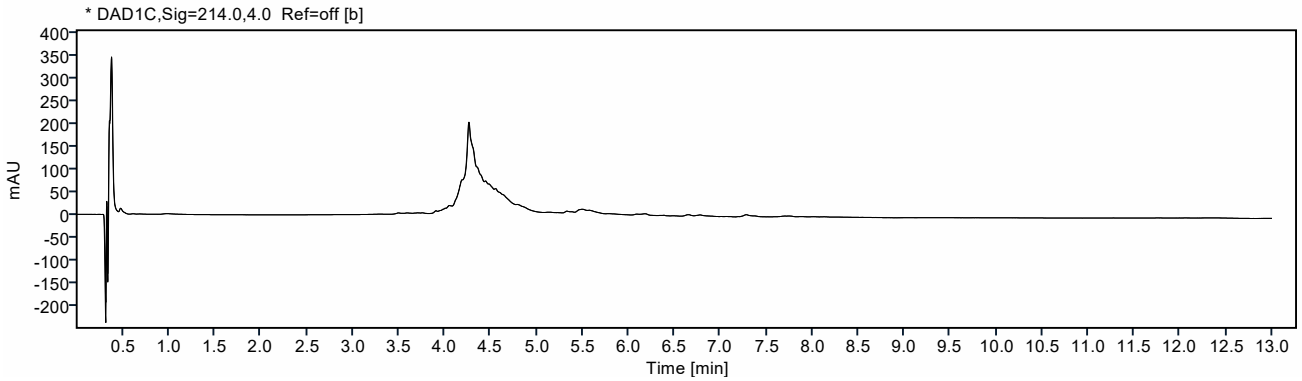

FIGURE S36. UHPLC profile of crude RS-4×pS using column A and a gradient of 5 to 95 % MeCN over 10 min at 40 °C. The peptide eluted at 4.27 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 10 min yields 28 % purity.

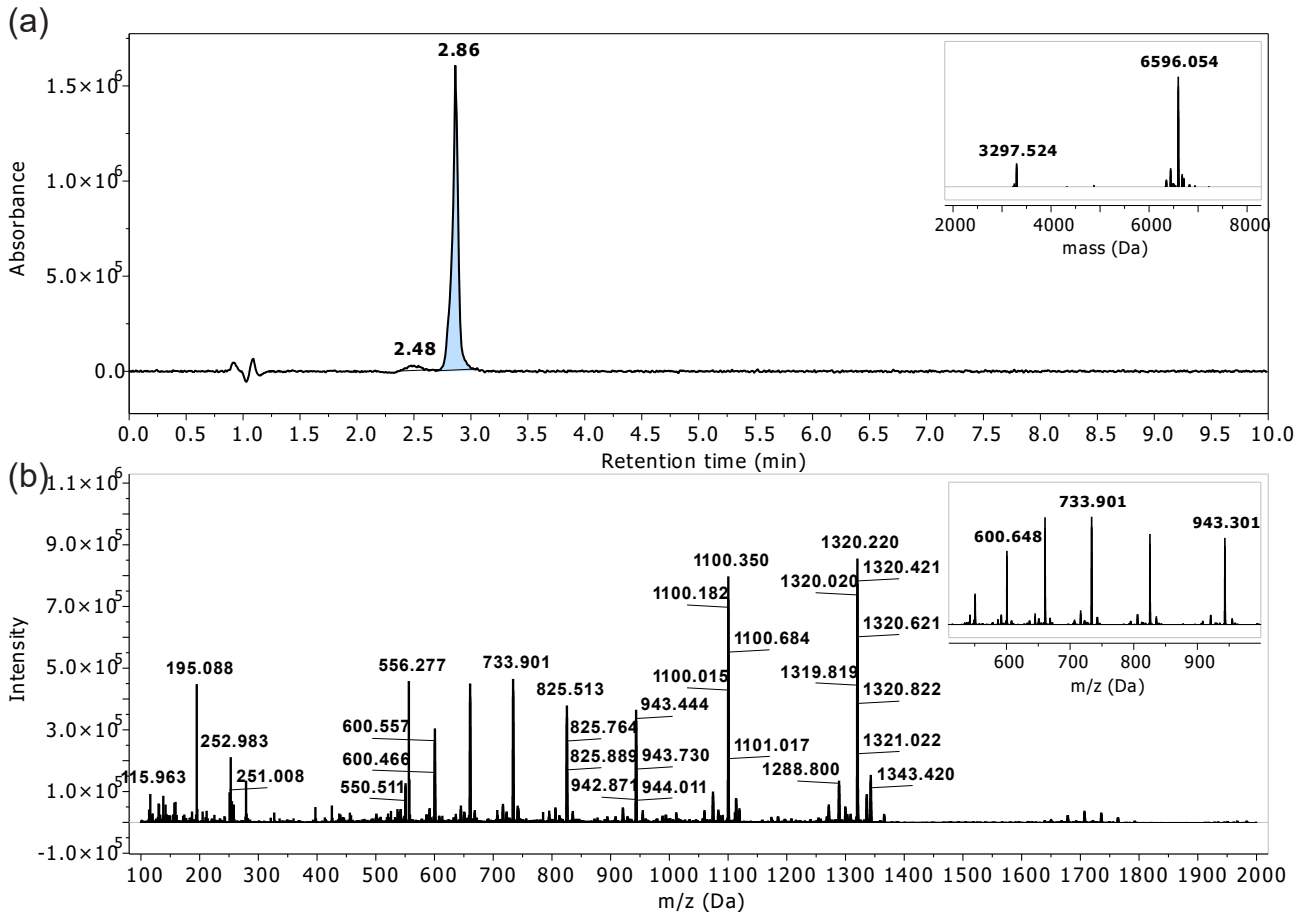

FIGURE S37. LCMS profile of pure RS-4×pS. (a) Absorbance chromatogram of pure RS-4×pS at $\lambda = 214$ nm with an Rt of 2.86 min (insert: deconvoluted masses). (b) ESI-TOF spectrum found within Rt 2 to 9 min (insert: convoluted spectra at Rt 2.86 min). Monoisotopic mass (ESI+) calculated for $C_{245}H_{426}N_{114}O_{90}P_4S_2$ is 6593.0654 Da; found 6593.0495 Da. LCMS Gradient A.

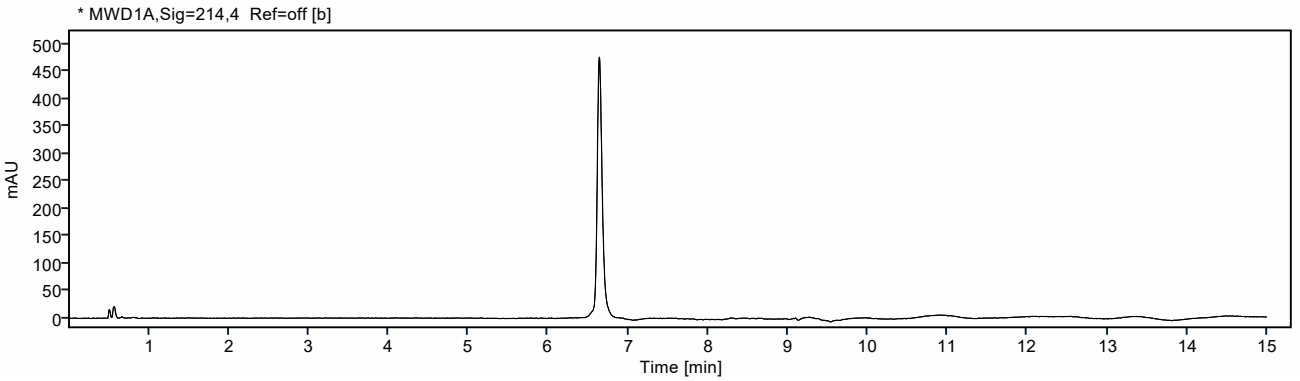

FIGURE S38. UHPLC profile of pure RS-4×pS using column C and a gradient of 0 to 100 % MeCN over 15 min at 40 °C. The peptide eluted at 6.64 min. Integration of the absorbance at $\lambda = 214$ nm between 1 to 15 min yields 99 % purity.