The N-terminal RNA Recognition Motif of PfSR1 Confers Semi-specificity for Pyrimidines during RNA Recognition

Akshay Kumar Ganguly, Garima Verma and Neel Sarovar Bhavesh

Transcription Regulation group, *International Centre for Genetic Engineering and Biotechnology (ICGEB), Aruna Asaf Ali Marg, New Delhi 110067, India*

Correspondence to Neel Sarovar Bhavesh: *neelsb@icgeb.res.in*
https://doi.org/10.1016/j.jmb.2018.11.020
Edited by C. Kalodimos

Abstract

Alternative splicing confers a complexity to the mRNA landscape of apicomplexans, resulting in a high proteomic diversity. The *Plasmodium falciparum* Ser/Arg-rich protein 1 (PfSR1) is the first protein to be confirmed as an alternative splicing factor in this class of parasitic protists [1]. A recent study [2] showed a purine bias in RNA binding among cognate RNA substrates of PfSR1. Here, we have investigated the role played by the amino-terminal RNA recognition motif (RRM1) of PfSR1 from the solution structure of its complex with ACAUCA RNA hexamer to understand how its mechanism of RNA recognition compares to human orthologs and to the C-terminal RRM. RNA binding by RRM1 is mediated through specific recognition of a cytosine base situated 5′ of one or more pyrimidine bases by a conserved tyrosine residue on β_1 and a glutamate residue on the β_4 strand. Affinity is conferred through insertion of a 3′ pyrimidine into a positively charged pocket. Retention of fast dynamics and ITC binding constants indicate the complex to be of moderate affinity. Using calorimetry and mapping of NMR chemical shift perturbations, we have also ascertained the purine preference of PfSR1 to be a property of the carboxy terminal pseudo-RRM (RRM2), which binds RNA non-canonically and with greater affinity compared to RRM1. Our findings show conclusive evidence of complementary RNA sequence recognition by the two RRMs, which may potentially aid PfSR1 in binding RNA with a high sequence specificity.

Introduction

Alternative splicing events in the parasite genome generate alternate transcripts from a common pre-mRNA. Thus, *Plasmodium falciparum*, which has a mere 5700 genes [3], has been able to adopt a life cycle that is much more intricate than organisms with more complex genomes, such as *Saccharomyces cerevisiae* (http://www.yeastgenome.org/cache/genomeSnapshot.html), making it one of the most lethal parasites. A similar and more pronounced behavior is observed in mammals, wherein greater than 90% genes exhibit alternate splicing [4]. Alternative mRNA splicing has enabled plasmodia to possess a diverse proteome from a relatively small genome, thus increasing the functional repertoire of these parasites [5–7]. One of the initial studied targets of alternative splicing in plasmodia was the merozoite apical erythrocyte-binding ligand (MAEBL) [6,8], which plays an important role in mosquito salivary gland invasion through one of its splice variants [8]. However, mechanisms of alternative pre-mRNA splicing and post-transcriptional regulation for this gene and for plasmodia in general are still largely unexplored [9,10].

In eukaryotes, serine/arginine-rich (SR) proteins act in synchrony with the spliceosome and are key players among both constitutive and alternative splicing factors [11–14]. Similar to other RNA binding proteins, SR proteins exhibit modularity. An eponymous SR domain, unstructured and rich in serine and arginine residues, is the hallmark of this family. Besides this domain, one or more RNA recognition motifs (RRMs) are usually present that bind RNA [15].

The SR domains possess regulatory and trafficking roles that are adopted on phosphorylation of serine and arginine residues by kinases such as SRPK [12]. Among lower organisms, SR proteins have been characterized in *Schizosaccharomyces pombe* [16] and *Trypanosoma cruzi* [17]. The first functional

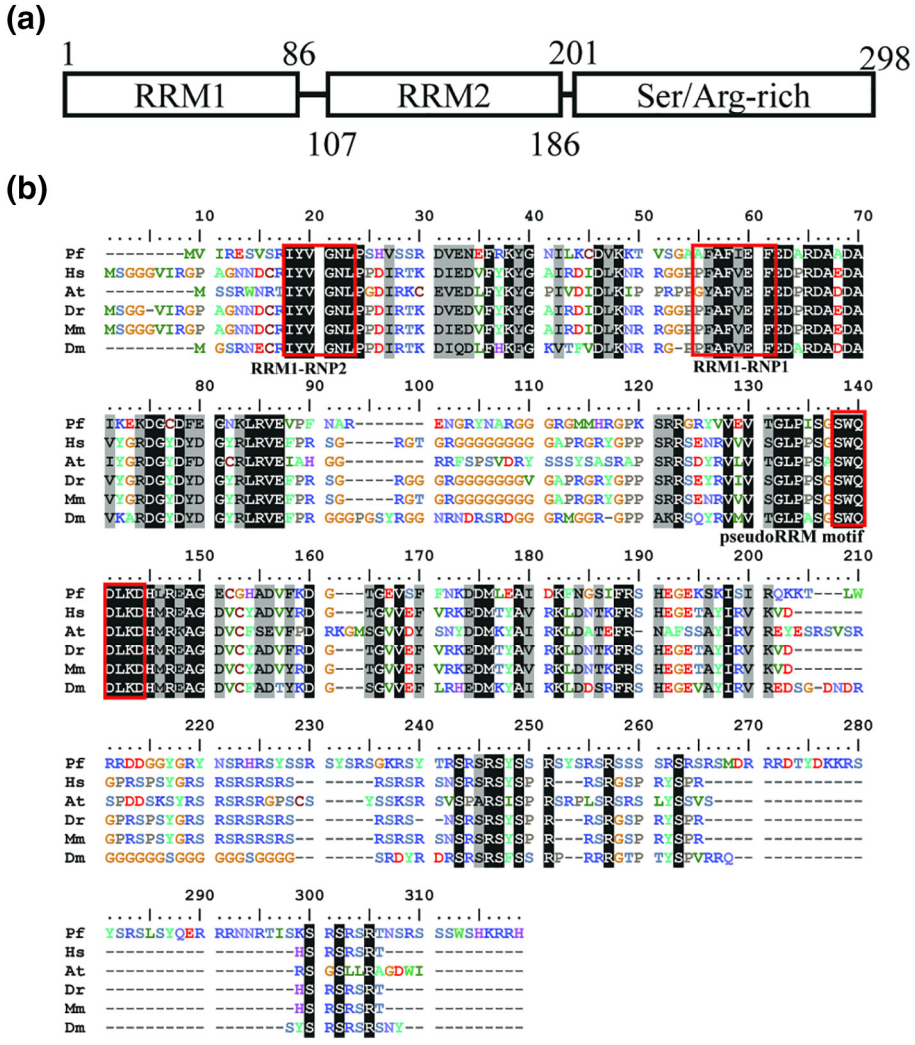

Fig. 1. Sequence analysis of *P. falciparum* Ser/Arg-rich protein 1. (a) Domain architecture of the alternate splicing factor PfSR1, showing the two RRMs and the Ser/Arg-rich carboxy-terminus. (b) ClustalW sequence alignment of *Pf*SR1 with SR alternate splicing factors from *H. sapiens* (Hs), *Arabidopsis thaliana* (At), *D. rerio* (Dr), *M. musculus* (Mm), and *Drosophila melanogaster* (Dm). Three boxed regions correspond to the two RNP sites on RRM1 and the pseudo-RRM RNA binding motif on RRM2.

alternate splicing factor found in apicomplexans was an SR protein called PfSR1 (PFE0865c) in *P. falciparum* [1]. The above study demonstrated that PfSR1 was able to interact with the spliceosome and influence splicing activity through interaction with splice sites, as well as localized to both the parasite nucleus and cytoplasm. PfSRPK1 kinase has been determined to be responsible for the phosphorylation of the SR domain of PfSR1, which affected its affinity toward RNA as well as mediated its nuclear transport [18]. A recent study has reported that the cognate RNA landscape of PfSR1 consists of GA- and GU-rich sequences [2]. However, this can probably be explained as the specificity of the C-terminal pseudo-RRM on PfSR1, given that human pseudo-RRMs such as SRSF1 exhibit similar binding specificities [19]. Being the first characterized alternative splicing factor in apicomplexans, it is crucial to delineate how each domain contributes to RNA recognition and how their mechanisms compare to what we know in the animal kingdom.

Here, we mainly focus on showing how the N-terminal RRM of PfSR1 semi-specifically recognizes pyrimidines on RNA using a combination of NMR-based structural biology and calorimetric approaches. Furthermore, we provide a comparison in RNA binding mechanism of RRM1 with the human ortholog SRp20 (PDB ID: 2I2Y, 43% identity with PfSR1-RRM1) and highlight how it differs from RRM2.

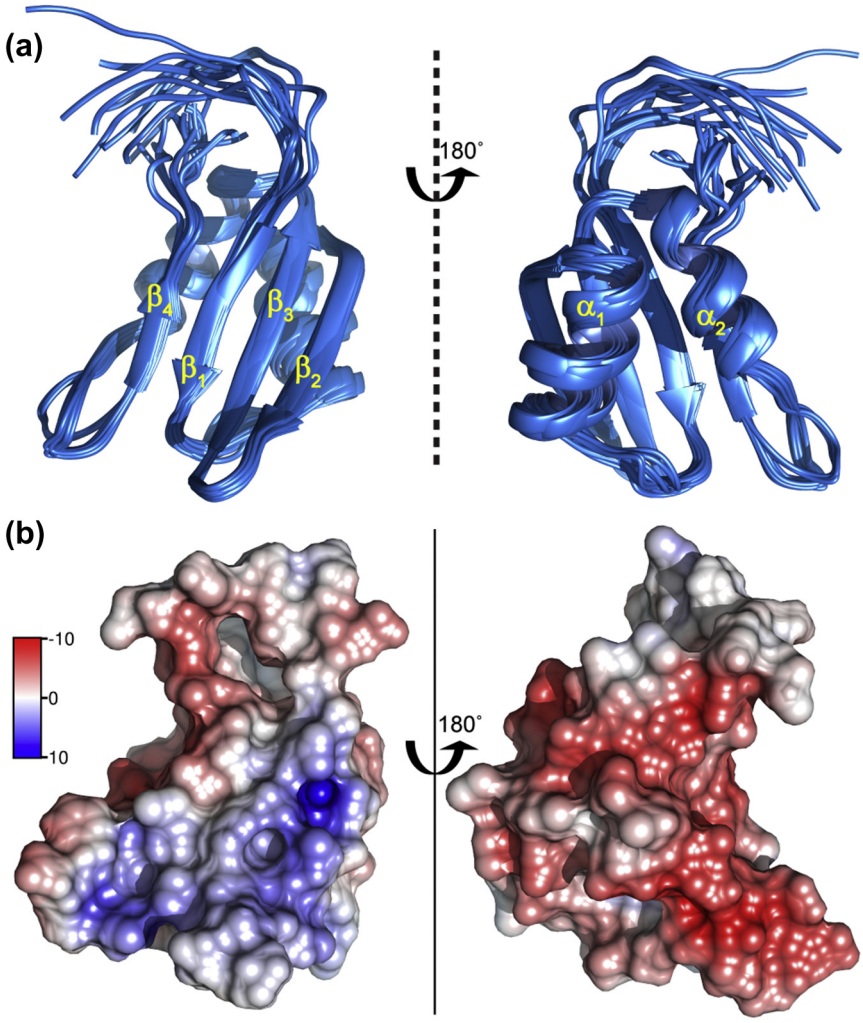

Fig. 2. Solution structure of PfSR1-RRM1 in free form. (a) RRM1 solution structure of 20 conformers (ribbon view) showing a canonical RRM fold (PDB ID: 2N3L). β-strands (left panel) and α-helices (right panel) are labeled. (b) APBS-derived electrostatic surface view of RRM1, showing a positively charged β-sheet (left panel) and negatively charged α-helical face (right panel). Charges are in units of $k_B T/e$, where k_B = Boltzmann constant, T = temperature in Kelvin, and e = charge of an electron.

Results

RRM1 of PfSR1 has a conserved structural architecture

PfSR1 possesses a domain architecture of two RRMs (RRM1 toward the amino terminus and RRM2 at the center of the polypeptide) separated by a short linker and a characteristic carboxy-terminus rich in serine and arginine residues (Fig. 1a). A comparison of polypeptide sequences across five model organisms (Fig. 1b) reveals two, broadly conserved regions, which correspond to the two RRMs. PfSR1 shows conservation of canonical RNA-binding ribonucleoprotein (RNP) sites: RNP1 and RNP2 [20] in RRM1. From sequence homology, RRM2 is ostensibly a pseudo-RRM that lacks RNP sites but contains a heptapeptide stretch (SWQDLKD) that has been previously implicated in non-canonical RNA recognition in human SRSF1 [19]. The differences mainly lie in the linker region between the two RRMs and the Ser/Arg-rich carboxy terminus. The linker showed a higher sequence complexity as compared to those in *Homo sapiens*, *Danio rerio*, and *Mus musculus* SRSF1, and the SR region was greater in length. Comparisons among members of *Plasmodium* sp., however, revealed a near-global sequence conservation, with variations occurring mainly in the SR-rich region (Supplementary Fig. S1A). In addition, a comparison of the two RRMs also showed the residue-wise differences that exist in the RNA-binding regions (Supplementary Fig. S1B). Notably, the absence of aromatic residues in the RNP1 and RNP2 sites for RRM2, as well as the absence of a tryptophan residue in the pseudo-RRM motif region of RRM1 indicated the mechanistic divergence of these two domains in RNA recognition.

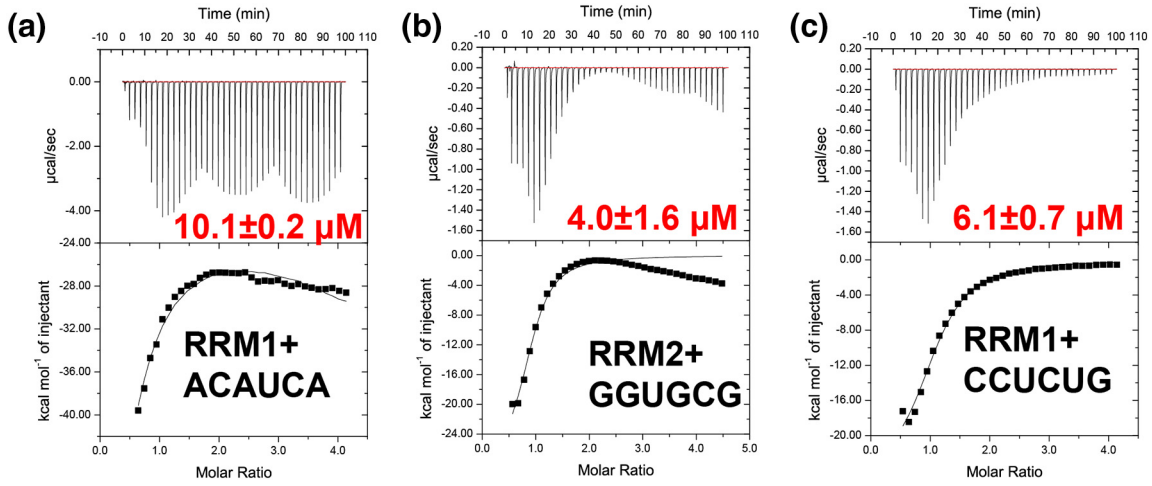

Fig. 3. Isothermal titration calorimetry of RRM–RNA binding: ITC thermograms of (a) RRM1 against ACAUCA*, (b) RRM1 against CCUCUG, and (c) RRM2 against GGUGCG. Dissociation constants (K_d) are written in red, along with corresponding standard errors. *Higher heats observed in the raw data after 40 min of titration were due to high heats of dilution of ACAUCA. Similar heats were also observed for UCUUCU and $(A)_6$ RNAs. These heats were subtracted from the calculated isotherm by running dilution controls.

In order to gain insight on structural conservation, we determined the three-dimensional structure of PfSR1-RRM1 using NMR spectroscopy. The solution structure of the N-terminal RNA recognition motif (RRM1) (Fig. 2a) was calculated using a total of 986 restraints, of which 167 were long range, ($|i–j| > 4$) conformationally restraining distance restraints (Supplementary Table S1). It exhibited a canonical β_1–α_1–β_2–β_3–α_2–β_4 RRM fold with a β-sheet composed of four β-strands and two α-helices arranged at an angle to each other. Residues comprising secondary structural elements and hydrophobic core are tabulated in Supplementary Table S2. Conserved RNP sites were situated on β_1 (RNP2; I_{10}YVGNL$_{15}$) and β_3 (RNP1; A_{47}FAFIEF$_{53}$) strands. The RRM1 surface exhibited a distinct electrostatic polarity, with the β-sheet being positively charged and the α-helical face being negatively charged (Fig. 2b).

Prior to structure calculation, we assigned ~97% backbone and ~79% side-chain resonances for RRM1 (Supplementary Fig. S2A). RRM2, on the other hand, exhibited exchange broadening with time, rendering complete resonance assignment impossible, despite several efforts at spectral optimization by varying temperature, buffers and through use of deuteration. As a result, we were able to assign ~85% backbone (H^N, N^H, C^α, and C^β) and ~10% side-chain resonances for RRM2 (Supplementary Fig. S2B). The combined construct of RRM1 and RRM2 showed extensive line broadening in 2D [^{15}N,^{1}H] HSQC spectrum and hence could not be assigned. This was probably an outcome of reduced flexibility of the plasmodial linker as compared to its human counterpart, given its higher sequence complexity and hydropathy (Fig. 1b). A less flexible linker would cause the individual domains to tumble as a single, large entity, resulting in shorter T_2 relaxation times [21].

RRM1 and RRM2 show complementary and contrasting sequence specificities

We used isothermal titration calorimetry to quantify interactions of previously identified RNA targets with the two RRMs of PfSR1 [2,18,20,22,23]. The first RNA target to be identified for PfSR1 was a partial transcript of the PF13_0082 gene, encompassing the whole of intron 3 and segments of exons 3 and 4 [18]. In a subsequent study, SR1 binding motifs were found to be enriched up to 96.6% in exons and also composed of GA- or GU-rich sequences [2]. Consequently, we assayed the binding of a 16-nt oligomer (referred herein as 16A) from the exon4 of PF13_0082, containing a portion of intron3, (UACAG/UUCUUGGAGGU, the "/" marks the 3′ splice site of intron 3) to RRM1 and RRM2. Both domains bound to this RNA sequence, with RRM1 binding at a 10-fold weaker affinity (K_d = 1.8 µM) as compared to RRM2 (K_d = 170 nM) (Supplementary fig. S3A, B). Furthermore, the RIP-ChIP-based screening cited above [2] revealed purine-rich sequences such as GAAGAA being recognized by PfSR1. Taking this into consideration, the purine-rich region of 16A downstream of the splice junction (GGAGGU) was titrated against both RRMs. Surprisingly, only RRM2 bound to this RNA sequence (K_d = 8.9 µM) (Supplementary Fig. S3C, D), indicating that RRM1 probably bound closer to the 3′ splice site of intron 3. Binding of RRM2 toward the hexameric GGAGGU sequence was found to be nearly 50-fold weaker as compared

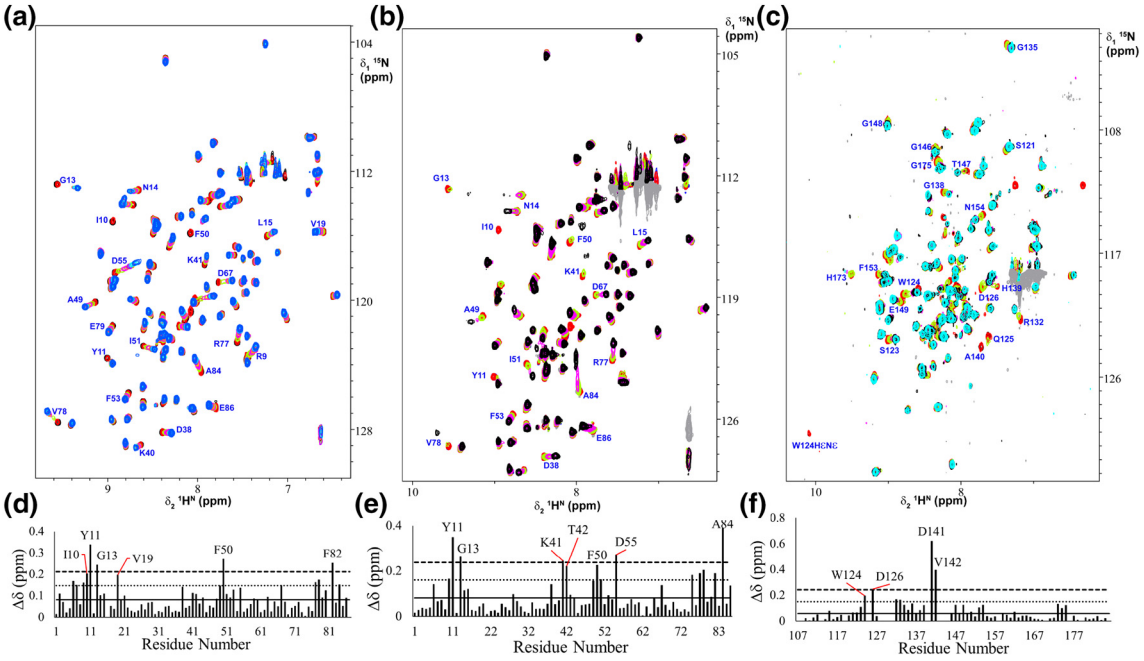

Fig. 4. Mapping of residues on RRM1 and RRM2 that interact with RNA. (a) Overlay of 2D [^{15}N, ^{1}H] HSQC spectra of RRM1 titrated against zero (black), 0.2 (red), 0.4 (green), 0.6 (pink), 0.8 (ochre), 1.0 (purple), 1.2 (cyan), and 1.4 (blue) molar ratios of ACAUCA and (b) against zero (blue), 0.2 (red), 0.6 (green), 1.0 (pink), and 1.4 (black) molar ratios of CCUCUG. (c) Overlay of 2D [^{15}N, ^{1}H] HSQC spectra of RRM2 titrated against zero (navy blue), 0.2 (red), 0.4 (green), 0.6 (pink), 1.0 (black), and 1.4 (cyan) molar ratios of GGUGCG. Residue-wise chemical shift perturbation (Δδ) profile of (d) RRM1 against ACAUCA, (e) RRM1 against CCUCUG, and (f) RRM2 against GGUGCG, with significantly perturbed residues labeled. Horizontal lines correspond to mean (solid), mean + 1σ (dotted), and mean + 2σ (dashed) levels.

to 16A. At this juncture, we can assume that this is due to either increased slippage of the hexamer from RRM2 or an expanded number of non-specific contacts with the longer RNA. Our attempts to quantify interaction with RNAs GAUGAUGA, GAUGA, and GUUGA sequences did not succeed due to protein precipitation during titration. In humans, SRp20, a canonical SR-RRM having a 43% sequence identity with PfSR1-RRM1, shows a preference toward the consensus sequence (A/U)-C-(A/U)-(A/U)-C [23]. Subsequently, we tested three hexameric RNAs containing CAUC, CCUC, and CUUC motifs for binding to RRM1. All three RNAs showed binding to RRM1 with dissociation constants of 10.1 μM (Fig. 3a), 6.1 μM (Fig. 3B), and 10.0 μM (Supplementary Fig. S3E), respectively. To check whether recognition was broadly pyrimidine-specific, RRM1 was titrated against $(U)_6$ RNA (Supplementary Fig. S3F). Interestingly, no binding was detected in this instance, indicating that the initial cytosine is probably important in mediating recognition by RRM1. Moreover, RRM2 failed to bind to CUUC motif (Supplementary Fig. S3G), implying that there was no cross-over between specificities of the two RRM domains. A similar test for purine specificity of RRM2 was performed by titrating GGUGCG and $(A)_6$ RNAs against it. RRM2 bound GGUGCG with a moderate affinity (K_d = 4.0 μM) (Fig. 3c) but failed to bind $(A)_6$ (Supplementary Fig. S3H), showing that the initial presence of guanine residues is likely to be crucial for recognition. Since PfSR1 would likely bind longer mRNAs *in vivo*, affinities of binding were tested for both RRMs against a longer, 16-mer synthetic RNA containing motifs for both RRMs (A**CAUC**AU**UCUUGGAG**U), referred to as 16B here onward. While RRM1 showed a nearly 5-fold increase in affinity (K_d = 1.8 μM) similar to its binding for 16A, RRM2 exhibited an extremely strong association with its cognate sequence (K_d = 20 nM) (Supplementary Fig. S3I, J), around 8-fold stronger than what was observed for the biological 16A sequence. This clearly explains the purine bias in PfSR1 targets that were determined in the earlier RIP-ChIP study [2]. The use of immunoprecipitation without cross-linking as in the case of RIP-ChIP would bias the selected RNAs toward high-affinity targets of RRM2.

2D [^{15}N, ^{1}H] HSQC spectra were used to identify residues on the proteins involved in RNA binding. *U*-^{15}N RRM1 and RRM2 were independently titrated against increasing molar ratios of corresponding hexamers. Changes in chemical shifts (chemical shift perturbations, CSPs) were monitored using 2D [^{15}N, ^{1}H] HSQC spectra recorded at each molar ratio. RRM1 was titrated against ACAUCA (Fig. 4a) and CCUCUG (Fig. 4b) RNAs, while RRM2 was titrated

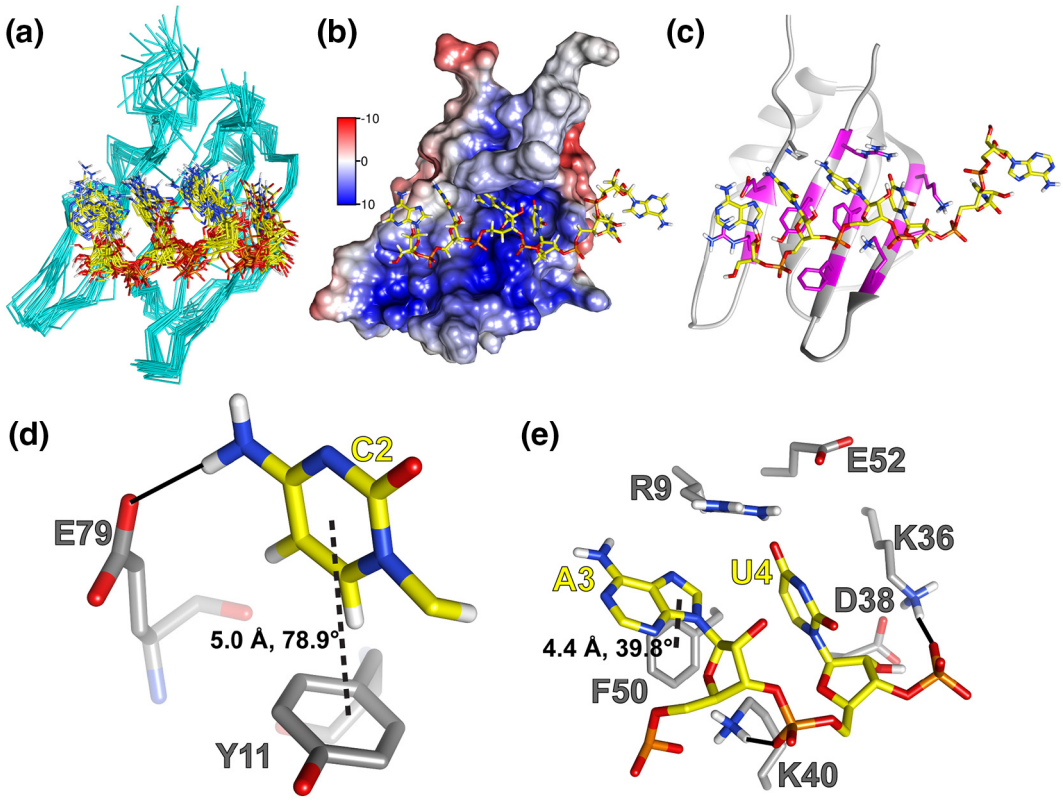

Fig. 5. Solution structure of RRM1 in complex with ACAUCA. (a) Structural ensemble of 20 conformers energy-minimized in AMBER (PDB ID: 2N7C). RNA is in yellow, while RRM domain is shown in cyan. Flexible residues Gly-2 to Glu5 and bases C5 and A6 are removed for clarity. (b) Structure of the complex (model 1 of the ensemble) shown with APBS-derived electrostatic surface of RRM1 with RNA (yellow). Charge units are in $k_B T/e$. (c) Ribbon representation of RRM–RNA complex with interfacial amino acids colored in magenta and nitrogen bases shown in yellow. Solid black lines represent hydrogen bonds. (d, e) Major interacting amino acids (gray) and nitrogen bases (yellow) shown in isolation. Solid black lines represent H-bonds and dashed black lines indicate stacking interactions. Distances between centers of geometry and inter-planar angles are shown for the two stacking interactions.

against GGUGCG (Fig. 4c). Binding of RRM1 to ACAUCA and CCUCUG mainly caused changes in the chemical shifts of residues within RNP1 and RNP2 sites, along with a few residues near the carboxy terminus (Fig. 4d). In the RNP1 site, Phe50 showed the largest deviation, indicating its direct involvement in RNA binding. Ile10, Tyr11 and Gly13 in the RNP2 site also exhibited a high perturbation. On the β_4 strand, Arg77 and Val78 showed CSPs higher than 1 SD above the mean, while Asp38 was the lone interacting amino acid on the β_2 strand. The CSPs were thus a clear indication of RNA binding across the β-sheet of the RRM in a canonical manner. Aside from the β-sheet, Phe82 on the C-terminal tail and Val19 in the β_1-α_2 loop exhibited CSPs higher than 1 SD from the mean. Binding was accompanied by a global decrease in ^{15}N T_2 relaxation times, with a change in rotational auto-correlation time from 6.35 ns to 7.69 ns. Fast time-scale motions (T_1 and $\{^1H\}$–^{15}N nOe) were largely unperturbed (Supplementary Fig. S4), indicating that the complex was dynamic, giving credence to the moderate-level affinity observed in ITC. Similar contributions to RNA binding were observed for CCUCUG binding to RRM1 (Fig. 4e) but with additional CSPs seen for Lys41 and Thr42 on the junction between β_2 and β_2–β_3 loop; and a high perturbation of Ala84 on the C-terminal tail, indicating that its amide proton was probably hydrogen-bonded to RNA.

In contrast, the largest CSPs for RRM2 bound to GGUGCG were observed for Asp141 and Val142, rather than any aromatic residue (Fig. 4f). In addition, Ala140, Gln125, and Trp124 H N resonance showed extensive broadening of the resonances beyond 0.4 molar equivalents of RNA which, though non-quantifiable, may constitute binding nonetheless. Other residues showing CSPs greater than mean + 1σ were Trp124, Asp126, Arg132, and Glu133. The involvement of negatively charged residues in binding implied that the RNA was bound in a phosphate-out conformation, with hydrogen bonding between bases and negatively charged amino acids. The only possible stacking interaction would be with Trp124. This combination of interacting residues pointed to a

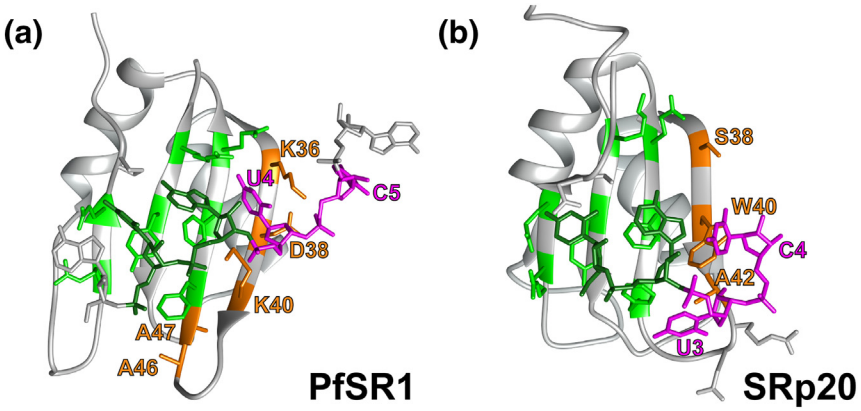

Fig. 6. Mechanistic differences between (a) PfSR1 and human (b) SRp20. A structural comparison of the two proteins is shown with residues highlighted by their similarity across the two orthologs. Amino acids depicted in orange exhibit major differences with the human ortholog. Similarly, the nucleotides showing a different binding mechanism are highlighted in magenta. Residues and nucleotides showing identity and similarity in their binding mechanisms are represented in light and dark green, respectively.

non-canonical mode of binding similar to that observed for the pseudo-RRM of human SRSF1 [19] with the conserved SWQDLKD motif being partially involved in RNA recognition (Supplementary Fig. S5).

Binding of RRMs to 16-mer RNAs as well as binding of GGAGGU to RRM2 caused severe broadening of amide proton resonances that were undetectable even at higher temperatures or with TROSY pulse schemes [24]. Hence, these were excluded from NMR analyses.

Structural basis for RNA recognition by PfSR1-RRM1

To determine the RNA sequence to be used for understanding the RNA recognition mechanism of RRM1, several considerations had to be taken into account. (1) There were no pyrimidine-rich, cytosine-containing consensus sequences identified for PfSR1 in past literature; (2) a mechanistic comparison of RNA binding with the closest human ortholog SRp20 (43% identity) would require an RNA close in sequence to CAUC (PDB ID: 2I2Y); (3) the binding of the RNA would have to be sufficiently strong to be a plausible binder *in vivo*; and (4) the binding would have to be of a sufficiently high chemical exchange rate between bound and unbound forms, which is commonly exploited to determine solution structures of RRM–RNA complexes owing to the retention of sharp linewidths even in the bound form [22,25]. The ACAUCA RNA was chosen as the ideal candidate to determine the mechanism by which RRM1 recognizes RNA as it exhibited all of the above properties. Previous literature on PfSR1 has not indicated at ACAUCA at being a consensus sequence, probably due to the bias of identified RNA sequences toward purine-rich RRM2 binders.

The solution structure of RRM1 in complex with ACAUCA RNA was subsequently calculated (Fig. 5a). Detailed structural statistics are shown in Supplementary Table S1. The SR1–RRM1–ACAUCA complex exhibited RNA binding across its β-sheet in a canonical manner. The first four residues (ACAU) of the RNA bound to a total surface area of 589.2 Å^2 on the positively charged face of the RRM. All four bound bases were in the *anti* configuration with C3′-endo sugar pucker. Of the first four bases, C2, A3, and U4 were found to possess more than 25 inter-atomic contacts with RRM1 on an average across the ensemble (Supplementary Fig. S6A). The RNA underwent 33.4% loss of solvent exposed area upon binding with a total interfacial area of 596.2 Å^2. Binding was predominantly on the β-sheet, with a minor contribution from the N-terminal amino acids. The RNA was secured in place through a combination of stacking interactions and hydrogen bonding (Fig. 5a). Stacking interactions included a π–π stack in the T-configuration between Tyr11 and C2 with a distance of 5.0 Å between centers of ring geometries and an interplanar angle of 78.9°. Owing to a hydrogen bond between C2 N^4H^4 and Glu79 O as well as a substantial dehydration of its surface area to the extent of 38.9 Å^2, aromatic $C^\delta H^\delta$ and $C^\varepsilon H^\varepsilon$ resonances of Tyr11 showed significant upfield shifts (Supplementary Fig. S7A). The approximate dissociation constant for Glu79 upon nucleotide binding was found to be 56 μM (Supplementary Fig. S8). Glu79, in turn, was secured by a hydrogen bond between its O^ε and $N^\eta H^\eta$ of Arg77. The role of Glu79 in RNA recognition was further corroborated by a loss of binding affinity for a E79A mutant (Supplementary Fig. S9B). The C2 ring was angled toward Tyr11 along the C^5H^5 and C^6H^6 edge, as evidenced by the presence of a nOe cross-peak between the H^6 atom and Tyr11 H^ε (Supplementary Fig. S7B). This stacking interaction was found to have a dissociation constant around 56 μM, similar to that for Glu79 (Supplementary Fig. S8). Net dehydration for C2 was 137.6 Å^2. The combination of stacking and

hydrogen-bonding interactions ensured that the binding was specific for cytosine. The contacts between C2 and E79 as well as C2 and Y11 were found to be relatively consistent across the ensemble, existing in 65–90% of the structural models (Supplementary Fig. S6C).

The second stacking interaction was a displaced π–π stacking between Phe50 and the five-membered ring of A3 (Buried surface area = 147.5 $Å^2$). This explained the large CSP shown by Phe50 during the titration (Fig. 4d). A salt bridge between Lys40 $N^{\zeta}H^{\zeta}$ and $O1^{P}$ of its phosphate group provided a non-specific tether. Interestingly, although A3 was found to have a high number of contacts with RRM1 across the ensemble, no specific inter-atomic contact was found to be conserved across the ensemble (Supplementary Fig. S6). The orientation of A3 was thus found to be dynamic throughout the ensemble, indicating the lack of specificity for an adenine at this position. This was seen when RRM1 was found to bind to UCUUCU and CCUCUG RNAs, which did not contain any adenine bases. Furthermore, the sigmoid kinetics observed for F50 upon nucleotide binding (Supplementary Fig. 8) demonstrated that the contact with A3 was stabilized only at higher concentrations of RNA and did not play a role in initial RNA tethering. U4 was found to be inserted into a cleft made of Lys36, Asp38, Lys40, Phe50, and Glu52, corroborated by a strong intermolecular nOe between U4 H^6 and Asp38 H^{α} as well as others between U4 H^6 and H^{α} resonances of Val39 and Ile51 (Supplementary Fig. S7C, D). Blocking this cleft by mutating Asp38 with a bulky Trp residue similar to Trp40 of SRp20 (Fig. 6b) resulted in a loss of affinity toward RNA (Supplementary Fig. S9A). Backbone of U4 was anchored by a hydrogen bond between $O1^{P}$ of its backbone phosphate and $N^{\zeta}H^{\zeta}$ of Lys40. This resulted in a higher desolvation for U4 as compared to other bases (buried surface area = 173.3 $Å^2$) and the highest number of contacts with RRM1, across the structural ensemble (Supplementary Fig. S6A, B). In terms of order of binding, residue-wise kinetics for the binding amino acids indicated that the charged residues Arg9, Lys36, and Glu52 were involved in initial encounter complex formation. The other residues that were actually found to bind bases in the atomic structure, showed distinctly slower kinetics, with stacking interactions actually showing sigmoid profiles (Supplementary Fig. S8).

The remainder of the 3′ end, C5, and A6 were largely flexible, as evidenced by the absence of intermolecular nOes and line broadening on NOESY spectra. Almost no contacts were found to be conserved for these two bases across the ensemble (Supplementary Fig. S6). The 5′ adenine was non-specifically bound by a loose cation–π interaction between its base and the guanidinium side-chain of Arg77, which was found to exist in more than 40% of conformers (Supplementary Fig. S6C). Owing to its location near the edge of the β-sheet, A1 remained ~75% solvent exposed. Binding of ACAUCA to RRM1 was thus partially sequence specific, mainly showing a preference toward cytosine near the 5′ end, followed by a pyrimidine nucleotide two bases downstream.

Discussion

Recent explorations of alternative splicing events in *P. falciparum* point to a significant role played by this phenomenon in post-transcriptional regulation of gene expression in this parasite [5–7]. The *P. falciparum* Ser/Arg-rich protein 1 is the first protein to be functionally characterized as an alternative splicing factor in apicomplexans [1]. Two studies have also determined cognate RNA partners for PfSR1 [2,18]. However, the consensus sequence of RNA bound by PfSR1 *in vivo* is yet to be confirmed in the literature, and neither does existing literature shed light on how each domain of PfSR1 recognizes RNA. In our study, rather than arrive at a consensus sequence of cognate RNA, we have performed a comparative mechanistic analysis of RNA recognition between PfSR1 and its well-characterized human orthologs SRp20 and SRSF1. As screening methods such as RIP-ChIP are yet to yield a clear consensus cognate sequence for PfSR1, we have used an outward approach beginning from the molecular mechanism of RNA binding for a known cognate sequence in the animal kingdom and extrapolating our results toward a potential consensus mechanism.

Given that prior studies on RNA binding were performed using the full-length protein, it was unknown how the individual RRMs of PfSR1 contribute to RNA binding. Architecturally, PfSR1 has an amino-terminal canonical RNA recognition motif (RRM1), a central pseudo-RNA recognition motif (RRM2) and an eponymous SR-rich domain at the C-terminus. The identified purine-rich GA and GU-containing cognate sequences in literature were likely to be bound only by RRM2, given the recognition of such sequences by the human pseudo-RRM ortholog SRSF1 [19]. In other words, the role of the entire N-terminal domain, RRM1, has remained a mystery. Our study provides a biophysical and structural characterization of RNA-binding preferences of the amino-terminal RNA recognition motif (RRM1) of PfSR1. Here, we show that RRM1 preferences are broadly pyrimidine-specific and that it specifically recognizes a cytosine through two conserved residues. In addition, we have shown evidence that the apparent purine specificity of PfSR1 [2] is conferred by RRM2 by a mechanism that is likely to be very similar to human SRSF1.

SR proteins are traditionally known to modulate splice site selection by binding to *cis*-acting exonic (ESEs) and intronic splicing enhancers on RNA [12]. PfSR1, similar to human SRSF1, also shows a

functional versatility by binding to intron-less RNA, which may play a role in mRNA stability [2]. In our study, both RRMs were found to bind tightly to an exonic site on the PF13_0082 gene, which was the first cognate RNA substrate of PfSR1 to be identified [18]. This gene encodes for a putative cop-coated vesicle membrane protein p24 precursor, implying that alternative splicing could play a key role in plasmodial intracellular secretory pathways. With GA-rich sequences being identified as PfSR1-binding motifs [2] and also being commonly found in human ESEs [26], the purine-rich GGAGGU sequence within the PF13_0082 mRNA 16-mer was tested for binding. As expected, similar to SRSF1 [27], only the pseudo-RRM (RRM2) of PfSR1 bound to this sequence. RRM1, on the other hand showed a preference for pyrimidine bases by binding to CUUC, CAUC and CCUC motifs *via* its β-sheet, implying that the 5′-UCUU-3′ motif upstream of GGAGGU was recognized by RRM1. Overall, this indicated binding to an ESE site, given that both RNA sequences were on exon4.

The presence of a cytosine base was found to be necessary for association with RRM1, given that a hexa-uridine RNA failed to show any binding. This could also explain why PfSR1 bound to cognate RNA motifs despite serial replacements with cytosine [2]. As previous literature did not indicate any putative cytosine-containing consensus sequence that could be bound by PfSR1-RRM1, nor is there any such existing mechanistic information of RNA recognition in apicomplexans, the choice of sequences for RRM1 was mainly influenced by the degenerate consensus sequence for SRp20 [27], the closest human ortholog with a well-defined RNA-recognition mechanism. The PF13_0082 RNA also contains a CAUC motif in exon3, with a GAGU motif downstream, which could also potentially be a binding site for PfSR1 through a concerted binding of both domains. The tethering of C2 in ACAUCA by a T-stack with Tyr11 and a hydrogen bond with the side-chain carboxy group of Glu79 provided clear evidence of cytosine specificity, in a manner similar to SRp20 [22]. Unlike its human counterpart, however, all bound bases (ACAU) were in the *anti* configuration as opposed to the *syn* configuration of a uridine base seen in SRp20. Semi-specificity for a pyrimidine was conferred by a cleft formed of five amino acids on β_2 and β_3 strands. An additional nucleotide could possibly be accommodated between cytosine and a downstream pyrimidine, which would be oriented away from the RRM in the *syn* form, similar to SRp20. Moreover, the high purine bias in the identified RNA motifs by Eshar *et al.* [2] could be explained by the nano-molar affinity of RRM2 to 5′-GGAGGU and 5′-GGAGU containing 16-mer RNAs, which would remain associated through a RIP-ChIP assay. The binding of RRM2 to RNA was of a non-canonical nature similar to SRSF1, mediated by residues outside the RNP sites (Supplementary Fig. S5) [19]. The necessity of an initial guanine was indicated by the lack of binding of RRM2 to hexa-adenosine RNA. Weaker binding was observed for smaller hexameric RNAs owing to their higher flexibility, which could potentially cause slippage of bound protein. Conversely, both RRMs exhibited strong affinity interactions with longer RNAs. This affinity would likely further increase *in vivo* in case of full-length PfSR1 protein binding to pre-mRNA. However, the increase in affinity for RRM2 (more than 50-fold) was much more pronounced as compared to that for RRM1 (~5-fold). One possible explanation can be an increased number of non-specific electrostatic contacts between RNA and RRM2, owing to the distinct difference in surface electrostatics of RRM2 and RRM1. RRM1 being of a more canonical type has a positively charged face lined with aromatic residues for RNA recognition, with the remainder of the domain being negatively charged (Supplementary Fig. S10A, C). The negative charge would preclude non-specific binding of negatively charged RNA. RRM2 on the other hand, exhibited CSPs that indicated that it would bind RNA through a small negatively charged region containing a single tryptophan residue. The remainder of the RRM2 domain is positively charged and could potentially provide a non-specific electrostatic platform for extended RNA binding (Supplementary Fig. S10B, C). From a biological perspective, a strong affinity binding would be crucial in preventing the rapid switching of cognate RNA targets *in vivo*, allowing tighter regulation of splicing for a specific target by a single SR1 molecule.

Such cognate RNA sequences would likely be similar to those found in ESEs on pre-mRNAs in the human genome [27]. A similar, transcriptome-wide study of ESEs in plasmodia would likely reveal similar sequence compositions, containing degenerate combinations of 5′-CN(N)Y-3′ and 5′-GPNG-3′ (N, any nucleotide; Y, pyrimidine; P, purine) motifs. The strong affinity along with the partial sequence specificity of the two domains indicates a lack of redundancy in splicing factors within *P. falciparum*, with SR1 being the go-to alternate splicing factor in most cases.

In conclusion, this study provides the first atomic-resolution structural characterization of an alternate splicing factor–RNA complex from apicomplexans. We have described the molecular mechanism of RNA recognition by RRM1 of PfSR1 and demonstrated the striking contrast between specificities of the two RRMs of PfSR1. The typical role of binding to ESEs constitutes only one of the myriad functions of this versatile class of proteins, which include mRNA transport, translation and stability [11,12] as well as association with other proteins such as histone H3 [28] and even RNA polymerase II [29]. An exploration into such moonlighting functions of SR proteins is still required in apicomplexans. Moreover, the presence of several other SR protein isoforms in *P. falciparum* also indicates a broad spectrum of activity. Our findings constitute a detailed look into one aspect of SR protein

functionality in the malarial parasite and provide valuable insight into the diversity of its cognate mRNA landscape. In addition, this structure of an alternate splicing factor–RNA complex from *P. falciparum* would help provide a seed to future atomic investigations into the transcriptional and post-transcriptional biology of parasites from apicomplexans class.

Methods

Molecular cloning and recombinant protein preparation

PfSR1 (Uniprot ID Q8I3T5) coding sequence was optimized for expression in *Escherichia coli* and synthesized from GeneArt (Life Technologies). Three sub-clones were prepared corresponding to the two separate RRMs (residues 1–86 and 107–186) and the combined construct (1–186). Primers used for amplification are enlisted in Supplementary Table S3. Both constructs were cloned into pET28a expression vector, downstream of a thrombin-cleavable 6x His-tag, and the vector plasmid was transformed in *E. coli* BL21(DE3) *CodonPlus* cells for protein expression.

Recombinant protein expression was carried out by growing *E. coli* cells up to an OD_{600} of ~0.8 in Luria–Bertani broth or M9 minimal medium, followed by induction with 0.5 mM IPTG at 18 °C for 30 h. For preparation of *U*-^{15}N or *U*-^{13}C–^{15}N-labeled recombinant proteins, $^{13}C_6$–D-glucose and $^{15}NH_4Cl$ were used as sole carbon and nitrogen sources in M9 minimal media. Proteins were purified using a combination of Ni-NTA gravity flow affinity chromatography and gel permeation chromatography. 6x-His tag was cleaved using thrombin prior to gel permeation chromatography. Final NMR buffer solution for protein preparations was 25 mM sodium phosphate (pH 5.9) containing 100 mM NaCl, 50 mM L-arginine, 25 mM L-glutamic acid, 0.01% NaN_3 (w/v), 3 mM TCEP, and 5% D_2O (v/v). Fractions were concentrated to ~1 mM using centrifugal filters of molecular weight cutoff of 3 kDa. Protease inhibitor cocktails (Roche) were added to the mix and aliquots were stored at −80 °C.

Solution-state NMR spectroscopy

All experiments were carried out on Bruker *Avance* III spectrometers equipped with cryogenic triple-resonance probes, operating at field strengths (^{1}H Larmor frequency) of 500 and 700 MHz. For protein resonance assignment, a set of standard double- and triple-resonance spectra [30], namely, 2D [^{15}N,^{1}H]-HSQC, 3D HNCA, 3D CBCAcoNH, 3D HNCACB, 3D HNCO, 3D CcccoNH, 3D HcccoNH, 2D [^{13}C,^{1}H]-HSQC [aliphatic (−1 to 5 ppm $^1H^{ali}$) and aromatic (4.7 to 10 ppm $^1H^{aro}$)], 3D ^{15}N-edited [^{1}H,^{1}H]-NOESY, 3D $^{13}C^{ali}$-edited [^{1}H,^{1}H]-NOESY, and 3D $^{13}C^{aro}$-edited [^{1}H,^{1}H]-NOESY (NOESY mixing time: 100 ms) were measured at 303 K. Backbone dynamics of RRM1 in free and RNA bound form were measured using 2D ^{15}N–{^{1}H}-heteronuclear NOE and ^{15}N T_1, T_2 relaxation experiments. Sixteen delays ranging from 10 to 1100 ms were used for T_1, while 11 delays from 10 to 210 ms were used for T_2. In addition, a 3D $^{13}C^{ali}$-edited [^{1}H,^{1}H]-NOESY experiment was acquired in 99.9% D_2O for unambiguous nOe assignment of $^1H^{\alpha}/^{13}C^{\alpha}$ resonances. Topspin 2.1 and 3.5 (Bruker AG) software was used for acquisition, Fourier transformation and processing of time-domain data. Nitrogen spectral widths were adjusted to include all observable backbone amide peaks. Backbone and side-chain resonances were assigned manually using Computer Aided Resonance Assignment (CARA) software [31] with ^{1}H, ^{13}C, and ^{15}N shifts referenced indirectly to the DSS methyl proton resonance at 0 ppm in all spectra. Backbone (φ, ψ) and side-chain (χ) dihedral angle values were derived from observed chemical shifts using TALOS-N [32,33] (http://spin.niddk.nih.gov/bax/software/TALOS-N/).

Unlabeled RNA oligos were obtained in pure, 2′ desalted and 2′ de-protected form from GE Dharmacon. Oligos used were as follows: ACAUCA, UCUUCU, CCUCUG, GGAGGU, GGUGCG, AAAAAA (A_6), UUUUUU (U_6), UACAGUUCUUGGAGGU (16A), and ACAUCAUUCUUGGAGU (16B). Lyophilized RNAs were dissolved in NMR buffer prepared in diethyl-pyrocarbonate-treated water. Resonance assignments were carried out for 5′ ACAUCA 3′ hexamer by acquiring 2D [^{1}H,^{1}H] NOESY spectra (t_{mix} = 60 ms, 150 ms, 250 ms, 350 ms), 2D [^{1}H,^{1}H] TOCSY (t_{mix} = 50 ms), and 2D [^{13}C,^{1}H] HSQC spectra (50–100 ppm carbon spectral width, $^1J_{HC}$ = 150 Hz for sugars and 90–160 ppm carbon spectral width, $^1J_{HC}$ = 190 Hz for nitrogen bases). All spectra were acquired in ~99.9% D_2O and at 273.2 K. $^1H^C$, $^{13}C^{Base}$, and $13C^{Sugar}$ resonances were assigned using CARA.

NMR spectroscopy of protein–RNA complexes

Interface mapping of protein–RNA complexes was carried out by titrating 0.2 mM *U*-^{15}N protein against increasing molar ratios of RNA from 1:0 to 1:1.4 in steps of 0.2. Changes in chemical shifts (chemical shift perturbations, Δδ) of backbone amide protons were tracked on 2D [^{15}N,^{1}H] HSQC spectra recorded at each molar ratio. CSPs were calculated using the equation [34] –

$$\Delta\delta_{^{15}N^H,H^N} = \sqrt{\left(\frac{\Delta\delta_{^{15}N^H}}{5}\right)^2 + \left(\Delta\delta_{H^N}\right)^2}$$

where $\Delta\delta_H^N$ and $\Delta\delta^{15}{}_N^H$ are the changes in backbone amide chemical shifts for $^1H^N$ and ^{15}N resonances,

respectively. Interacting bases were identified on ACAUCA RNA in complex with a 1.1 fold molar excess of *U*-^{13}C–^{15}N *Pf*SR1-RRM1 by acquiring 2D ^{13}C-[ω_1, ω_2]-filtered [^{1}H,^{1}H] NOESY (t_{mix} = 350 ms) [35] and 2D [^{13}C,^{1}H] HSQC spectra for sugars and bases. RNA spectra were acquired in ~99.9% D_2O and at 273.2 and 298 K.

Isothermal titration calorimetry

Dissociation constants were determined by performing isothermal titration calorimetry at 298 K. Proteins and RNAs were prepared in 20 mM HEPES (pH 7.0) containing 50 mM NaCl and 3 mM β-mercaptoethanol in diethyl-pyrocarbonate-treated water. Concentrations were measured at 280 and 260 nm for proteins and RNAs, respectively. ITC experiments were carried out in a GE MicroCal iTC_{200} calorimeter with 0.05 or 0.1 mM protein inside the ITC cell and 1 to 2.5 mM RNA in the syringe. Sequential injections of RNA (20 injections of 2 μl each or 40 injections of 1 μl each) gave heats that were fit to a 1:1 binding model using Origin 7 software supplied with the calorimeter. Heats of dilution were subtracted from final isotherms by performing control injections of RNA into buffer.

Calculation of solution structure

Solution structure of *Pf*SR1-RRM1 was calculated using ATNOS-CANDID [36,37] derived inter-proton restraints from the four 3D NOESY spectra mentioned in an earlier section. Torsion angle restraints were prepared from TALOSN-derived "Strong" φ and ψ angles by providing a 40° range about the mean. Two hundred preliminary structures was calculated using seven iterations of ATNOS-CANDID followed by distance geometry-simulated annealing protocol of 10,000 steps in CYANA 3.97. Top 50 structures with least violations were further refined by restrained energy minimization in explicit water using CNS [38]. The preliminary structural ensemble was used to identify potential hydrogen bond acceptors, which along with hydrogen-deuterium exchange studies, enabled the inclusion of hydrogen bond restraints in CYANA 3.97 and CNS. Weak nOes overlooked by the automated process were picked manually and added as restraints. Re-calculation resulted in a tighter structural ensemble after CNS refinement. Final ensemble contained 20 of the lowest energy structures.

Solution structure of RRM1-ACAUCA complex was determined with the aid of 41 intermolecular and 65 intra-RNA distance restraints along with two hydrogen bond restraints (C2 N^3H^3–Glu79 O^{ε} and U4 O^{1P}–Lys40 $N^{\zeta}H^{\zeta}$) identified from a preliminary structure, to improve convergence. Selection of the latter two restraints was justified by the loss of RNA binding upon mutation into alanine (For Glu79), a 20% increase in Het-nOe upon RNA binding (For Lys40) and strong nOes between distal side-chain H^C protons and RNA that indicated the orientation of the terminal ends of these residues toward RNA. These restraints were derived from 2D [^{1}H,^{1}H] NOESY spectra of the complex with ^{13}C decoupling in both dimensions. Distance restraints were estimated using the spin-pair approximation [39] using the intensity of the H^5–H^6 cross-peak of CYT 5 that corresponds to a theoretical distance of 2.4 Å. Intermolecular nOe restraints for the complex were obtained from the 3D $^{13}C^{Ali}$-resolved ^{13}C[ω_1]-filtered/^{13}C[ω_2]-edited [^{1}H,^{1}H]-NOESY spectrum. Additional intermolecular nOes were obtained by differentiating from intra-RNA nOes as well as $^{13}C^{aro}$–$^1H^{aro}$ resonances arising from the protein by comparing the spectrum against 2D $^{13}C^{aro}$-[ω_2]-edited [^{1}H,^{1}H] NOESY (t_{mix} = 100 ms) and 2D ^{13}C-[ω_1,ω_2]-filtered [^{1}H,^{1}H] NOESY (t_{mix} = 350 ms). Preliminary structures were calculated using CYANA 3.97, similar to free RRM1. Refinement was performed by simulated annealing and energy minimization in explicit solvent in AMBER14 [40]. All models were validated using PROCHECK-NMR [41] and MolProbity [42] and visualized using Chimera v1.10 [43]. Model 1 of the ensemble was chosen as the representative structure for all analyses. Interface analysis was performed using PISA webserver (www.ebi.ac.uk/pdbe/pisa/).

Accession codes

Solution structures of PfSR1-RRM1 in free and ACAUCA RNA-bound form have been deposited in the Protein Data Bank under accession numbers PDB ID: 2N3L and PDB ID: 2N7C, respectively. The corresponding depositions of resonance assignments in the BioMagResBank have been made under accession numbers BMRB ID: 25650 and BMRB ID: 25800. Partial backbone resonance assignments of PfSR1-RRM2 have been deposited in the BMRB under accession number BMRB ID: 25779.

Acknowledgments

A.K.G. was a recipient of DBT senior research fellowship. G.V. is a recipient of UGC senior research fellowship. The study was supported by the ICGEB, India core funds. The authors wish to thank DBT and ICGEB, New Delhi, for providing financial support for the high field NMR spectrometers at the ICGEB, New Delhi, and NII, New Delhi. The Department of Biotechnology, Government of India funded the ITC instrument through grant number BT/PR13018/BRB/10/731/2009 and computational facility through grant number BT/PR6400/BRB/10/1138/2012 to N.S.B.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmb.2018.11.020.

Received 31 August 2018;
Received in revised form 17 November 2018;
Accepted 20 November 2018
Available online 28 November 2018

Keywords:
alternative splicing;
malaria;
protein–RNA complex;
structure;
NMR

Present affiliation: A.K. Ganguly, Sysmex Corporation, 4-4-4 Takatsuka-dai, Nishi-Ku, Kobe, Hyogo Prefecture 651 2271, Japan.

Abbreviations used:
SR, serine/arginine-rich; RRM, RNA recognition motif; RNP, ribonucleoprotein; ESE, exonic splicing enhancer.

References

[1] S. Eshar, E. Allemand, A. Sebag, F. Glaser, C. Muchardt, Y. Mandel-Gutfreund, et al., A novel *Plasmodium falciparum* SR protein is an alternative splicing factor required for the parasites' proliferation in human erythrocytes, Nucleic Acids Res. 40 (2012) 9903–9916.

[2] S. Eshar, L. Altenhofen, A. Rabner, P. Ross, Y. Fastman, Y. Mandel-Gutfreund, et al., PfSR1 controls alternative splicing and steady-state RNA levels in *Plasmodium falciparum* through preferential recognition of specific RNA motifs, Mol. Microbiol. 96 (2015) 1283–1297.

[3] M.J. Gardner, N. Hall, E. Fung, O. White, M. Berriman, R.W. Hyman, et al., Genome sequence of the human malaria parasite *Plasmodium falciparum*, Nature 419 (2002) 498–511.

[4] H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and function, Nat. Rev. Genet. 11 (2010) 345–355.

[5] H. Iriko, L. Jin, O. Kaneko, S. Takeo, E.T. Han, M. Tachibana, et al., A small-scale systematic analysis of alternative splicing in *Plasmodium falciparum*, Parasitol. Int. 58 (2009) 196–199.

[6] N. Singh, P. Preiser, L. Renia, B. Balu, J. Barnwell, P. Blair, et al., Conservation and developmental control of alternative splicing in maebl among malaria parasites, J. Mol. Biol. 343 (2004) 589–599.

[7] D.K. Muhia, C.A. Swales, U. Eckstein-Ludwig, S. Saran, S.D. Polley, J.M. Kelly, et al., Multiple splice variants encode a novel adenylyl cyclase of possible plastid origin expressed in the sexual stage of the malaria parasite *Plasmodium falciparum*, J. Biol. Chem. 278 (2003) 22014–22022.

[8] F.E. Saenz, B. Balu, J. Smith, S.R. Mendonca, J.H. Adams, The transmembrane isoform of *Plasmodium falciparum* MAEBL is essential for the invasion of Anopheles salivary glands, PLoS One 3 (2008), e2287.

[9] P. Horrocks, K. Dechering, M. Lanzer, Control of gene expression in *Plasmodium falciparum*, Mol. Biochem. Parasitol. 95 (1998) 171–181.

[10] J.S. Sims, K.T. Militello, P.A. Sims, V.P. Patel, J.M. Kasper, D.F. Wirth, Patterns of gene-specific and total transcriptional activity during the *Plasmodium falciparum* intraerythrocytic developmental cycle, Eukaryot. Cell 8 (2009) 327–338.

[11] P.J. Shepard, K.J. Hertel, The SR protein family, Genome Biol. 10 (2009) 242.

[12] J.C. Long, J.F. Caceres, The SR protein family of splicing factors: master regulators of gene expression, Biochem. J. 417 (2009) 15–27.

[13] M. Chen, J.L. Manley, Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches, Nat. Rev. Mol. Cell Biol. 10 (2009) 741–754.

[14] D.L. Black, Mechanisms of alternative pre-messenger RNA splicing, Annu. Rev. Biochem. 72 (2003) 291–336.

[15] P. Zuo, J.L. Manley, Functional domains of the human splicing factor ASF/SF2, EMBO J. 12 (1993) 4727–4737.

[16] G. Ast, How did alternative splicing evolve? Nat. Rev. Genet. 5 (2004) 773–782.

[17] D. Portal, J.M. Espinosa, G.S. Lobo, S. Kadener, C.A. Pereira, M. De La Mata, et al., An early ancestor in the evolution of splicing: a *Trypanosoma cruzi* serine–arginine-rich protein (TcSR) is functional in cis-splicing, Mol. Biochem. Parasitol. 127 (2003) 37–46.

[18] A. Dixit, P.K. Singh, G.P. Sharma, P. Malhotra, P. Sharma, PfSRPK1, a novel splicing-related kinase from *Plasmodium falciparum*, J. Biol. Chem. 285 (2010) 38315–38323.

[19] A. Clery, R. Sinha, O. Anczukow, A. Corrionero, A. Moursy, G.M. Daubner, et al., Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) E2802–E2811.

[20] G. Dreyfuss, M.S. Swanson, S. Pinol-Roma, Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation, Trends Biochem. Sci. 13 (1988) 86–91.

[21] G.M. Clore, A.M. Gronenborn, NMR structure determination of proteins and protein complexes larger than 20 kDa, Curr. Opin. Chem. Biol. 2 (1998) 564–570.

[22] Y. Hargous, G.M. Hautbergue, A.M. Tintaru, L. Skrisovska, A.P. Golovanov, J. Stevenin, et al., Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8, EMBO J. 25 (2006) 5126–5137.

[23] Y. Cavaloc, C.F. Bourgeois, L. Kister, J. Stevenin, The splicing factors 9G8 and SRp20 transactivate splicing through different and specific enhancers, RNA 5 (1999) 468–483.

[24] M. Salzmann, G. Wider, K. Pervushin, H. Senn, K. Wüthrich, TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins, J. Am. Chem. Soc. 121 (1999) 844–848.

[25] F.C. Oberstrass, S.D. Auweter, M. Erat, Y. Hargous, A. Henning, P. Wenter, et al., Structure of PTB bound to RNA: specific binding and implications for splicing regulation, Science 309 (2005) 2054–2057.

[26] W.G. Fairbrother, R.F. Yeh, P.A. Sharp, C.B. Burge, Predictive identification of exonic splicing enhancers in human genes, Science 297 (2002) 1007–1013.

[27] C.F. Bourgeois, F. Lejeune, J. Stevenin, Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA, Prog. Nucleic Acid Res. Mol. Biol. 78 (2004) 37–88.

[28] R.J. Loomis, Y. Naoe, J.B. Parker, V. Savic, M.R. Bozovsky, T. Macfarlan, et al., Chromatin binding of SRp20 and ASF/SF2 and

dissociation from mitotic chromosomes is modulated by histone H3 serine 10 phosphorylation, Mol. Cell 33 (2009) 450–461.
[29] T. Misteli, D.L. Spector, RNA polymerase II targets pre-mRNA splicing factors to transcription sites in vivo, Mol. Cell 3 (1999) 697–705.
[30] A. Bax, S. Grzesiek, Methodological advances in protein NMR, Acc. Chem. Res. 26 (1993) 131–138.
[31] R. Keller, The Computer Aided Resonance Assignment Tutorial, Cantina, Goldau, 2004.
[32] Y. Shen, A. Bax, Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks, J. Biomol. NMR 56 (2013) 227–241.
[33] M.V. Berjanskii, D.S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts, J. Am. Chem. Soc. 127 (2005) 14970–14971.
[34] M.P. Williamson, Using chemical shift perturbation to characterise ligand binding, Prog. Nucl. Magn. Reson. Spectrosc. 73 (2013) 1–16.
[35] C. Zwahlen, P. Legault, S.J. Vincent, J. Greenblatt, R. Konrat, L.E. Kay, Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex, J. Am. Chem. Soc. 119 (1997) 6711–6721.
[36] T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS, J. Biomol. NMR 24 (2002) 171–189.
[37] T. Herrmann, P. Guntert, K. Wuthrich, Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA, J. Mol. Biol. 319 (2002) 209–227.
[38] A.T. Brunger, P.D. Adams, G.M. Clore, W.L. Delano, P. Gros, R.W. Grosse-Kunstleve, et al., Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr. D Biol. Crystallogr. 54 (1998) 905–921.
[39] K. Wüthrich, NMR of Proteins and Nucleic Acids, Wiley, New York, 1986.
[40] B. Xia, V. Tsui, D.A. Case, H.J. Dyson, P.E. Wright, Comparison of protein solution structures refined by molecular dynamics simulation in vacuum, with a generalized Born model, and with explicit water, J. Biomol. NMR 22 (2002) 317–331.
[41] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, J. Appl. Crystallogr. 26 (1993) 283–291.
[42] I.W. Davis, A. Leaver-Fay, V.B. Chen, J.N. Block, G.J. Kapral, X. Wang, et al., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, Nucleic Acids Res. 35 (2007) W375–W383.
[43] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, et al., UCSF chimera—a visualization system for exploratory research and analysis, J. Comput. Chem. 25 (2004) 1605–1612.