The crystal structure of the Split End protein SHARP adds a new layer of complexity to proteins containing RNA recognition motifs

Fabiana Arieti[1], Caroline Gabus[1], Margherita Tambalo[1], Tiphaine Huet[1], Adam Round[2] and Stéphane Thore[1],*

[1]Department of Molecular Biology, University of Geneva, Geneva 1211, Switzerland and [2]European Molecular Biology Laboratory, Grenoble Outstation and Unit for Virus Host-Cell Interactions, University Grenoble Alpes-EMBL-CNRS, Grenoble 38042, France

ABSTRACT

The Split Ends (SPEN) protein was originally discovered in *Drosophila* in the late 1990s. Since then, homologous proteins have been identified in eukaryotic species ranging from plants to humans. Every family member contains three predicted RNA recognition motifs (RRMs) in the N-terminal region of the protein. We have determined the crystal structure of the region of the human SPEN homolog that contains these RRMs—the SMRT/HDAC1 Associated Repressor Protein (SHARP), at 2.0 Å resolution. SHARP is a co-regulator of the nuclear receptors. We demonstrate that two of the three RRMs, namely RRM3 and RRM4, interact via a highly conserved interface. Furthermore, we show that the RRM3–RRM4 block is the main platform mediating the stable association with the H12–H13 substructure found in the steroid receptor RNA activator (SRA), a long, non-coding RNA previously shown to play a crucial role in nuclear receptor transcriptional regulation. We determine that SHARP association with SRA relies on both single- and double-stranded RNA sequences. The crystal structure of the SHARP–RRM fragment, together with the associated RNA-binding studies, extend the repertoire of nucleic acid binding properties of RRM domains suggesting a new hypothesis for a better understanding of SPEN protein functions.

INTRODUCTION

The Split Ends (SPEN) gene was discovered in the mid-1990s through genetic studies linked to homeotic phenotypes in *Drosophila* (1,2). The severe developmental problems observed in knockout animals demonstrated its essential role (3,4). The rat homolog, called Msx2-Interacting Nuclear Target (MINT), was later identified independently in a screening for interacting homeoprotein during osteogenesis (5). MINT was shown to localize in the nucleus, and, *in vitro*, its predicted RNA recognition motifs (RRMs) were shown to interact with the osteocalcin (OC) proximal promoter region as well as with several homopolymeric DNA sequences (5). MINT was also proposed as a transcription repressor of the OC promoter. Multiple SPEN/MINT protein functions have been systematically reported as being associated with negative effects on transcription (6–9). The capacity of SPEN proteins to recruit transcriptional repressors was later shown to be due to a conserved SPEN paralog and ortholog C-terminal domain (SPOC). This domain arbitrates the interaction with the silencing mediator for retinoid and thyroid-hormone repressor protein (SMRT/NCoR) as well as with the histone deacetylase HDAC1 (10–13). The human homolog of the SPEN protein was renamed SMRT/HDAC1 Associated Repressor Protein or SHARP (we use this name throughout the manuscript; 13). The structure of the SPOC domain of SHARP has been previously determined, thus, providing important molecular details regarding the recruitment of SMRT by the SPOC domain (10). SHARP has also a proven role in Notch signalling, where SHARP and Notch-1-IC seem to have an exclusive binding behaviour with their partner RBP-Jκ (12). The region responsible for the stable recruitment of SHARP to RBP-Jκ is located in the central part of the protein, between residues 2804 and 2816 (12). Simultaneously, several independent studies demonstrated the role of SHARP in nuclear receptor-mediated transcriptional responses (13–15). The three predicted RRMs in the N-terminal region of SHARP were shown to mediate its negative transcriptional activity in the nuclear receptor pathway, both *in vitro* and *in vivo* (13,15). These effects

were shown to occur via its association with the non-coding RNA produced by the steroid receptor activator gene (SRA RNA; 13,15). A particular region containing the H12–H13 substructure of SRA RNA was shown to be sufficient to mediate SHARP association (15). RRMs are the most abundant RNA-binding domains (RBDs) present in vertebrates (they have been found in 0.5%–1% of human genes; 16). Interestingly, biochemical and structural studies of RRMs have generally shown that every RRM plays its own specific role in cellular functions; this is in spite of their structural similarities (17,18).

We have determined the crystal structure of the three RRMs present in the N-terminal part of SHARP. The atomic model revealed domain architecture where RRM3 and RRM4 form a platform, with RRM2 being linked flexibly. The residues responsible for the interaction between RRM3 and RRM4 are highly conserved throughout the SPEN family. Moreover, while the RRM3 has the consensus amino acids for single-stranded RNA association, the RNA-binding surface of RRM4 is blocked by an α-helix located immediately downstream of the RRM fold—a situation reminiscent of the newly defined xRRM present in the LARP protein (19). The xRRMs have the atypical properties of binding base-paired RNA sequences. We then characterized the association of the RRMs of SHARP with the H12–H13 RNA. Point mutations in the RRM3 or deletion of the RRM4 strongly destabilize the interaction with the H12-H13 fragment. The RRM3/RRM4 platform is therefore crucial for the formation of a stable complex with the H12–H13 region of the SRA RNA. We suggest that the association of SHARP with the H12–H13 RNA sequence is specific and requires stable stem loops including base-paired sequences. Our structural and biochemical data highlight the unexpected properties of the SHARP RRMs, which bring a new layer of complexity in the RNA recognition mode of proteins containing multiple RRMs.

MATERIALS AND METHODS

Molecular Biology

The nucleotide sequence encoding residues 335 to 620 of the human SHARP was obtained by gene synthesis (Entelechon) and named R2–3–4h. The sequence was codon-optimized for protein expression in *Escherichia coli* (*E. coli*) and it was then subcloned in a pET-based vector containing N-terminal 9XHistidine and Glutathione S-Transferase tags followed by the Tobacco Etch Virus (TEV) protease cleavage site. The truncation of SHARP–RRM corresponding to the constructs R3 (residues 436–512), R2–3 (residues 336 to 513), and R3–4h (residues 436 to 620) were cloned in the same expression plasmid and produce using the same protocol. The point mutants (R3mut and R2mut-R3mut) were obtained by site-directed mutagenesis of the plasmid encoding R2–3–4h using a Quick-Change mutagenesis kit (Stratagene) according to the manufacturer's instructions. The R3mut construct contains the following mutations: F441A, K470A, Y478A, F480A, and K512A. The R2mut-R3mut construct contains the same mutations plus K338A, Q369A, H371A, E376A, L380A, and F382A. All constructs were verified using DNA sequencing prior to large-scale ex-

pression. Primer sequences used for the truncations or the mutations are provided in Supplementary Table S1. nullnull

Protein expression and purification

All the proteins tested in the subsequent experiments were produced using the same protocol. Briefly, large cultures of *E. coli* BL21 Star cells, previously transformed with a given SHARP construct, were grown in Terrific Broth media (1.2% peptone, 2.4% yeast extract, 72 mM K_2HPO_4, 17 mM KH_2PO_4, and 0.4% glycerol) at 37°C for 4 h. This was followed by overnight induction at 18°C with 0.25 mM isopropyl-β-D-thiogalactopyranoside. Cells were harvested using centrifugation at 5200 rcf and resuspended in 50 mM HEPES buffer (pH 7.5) containing 300 mM NaCl, 20 mM imidazole, 0.1% X-Triton 100, DNase 1 (1 μg/ml), Lysozyme (1 μg/ml), 5 mM β-mercaptoethanol, and a cocktail of protease inhibitors (PhenylMethylSulfonyl Fluoride 1 mM, leupeptin 1 μg/ml, and pepstatin 2 μg/ml). Cells were lysed using an Emulsiflex system (Avestin) and cleared using centrifugation at 33,000 rcf for 30 min at 4°C. The soluble fraction was subjected to an initial affinity purification using a chelating HiTrap FF crude column (GE Healthcare) charged with Ni^{2+} ions. The protein was eluted with 250 mM imidazole and desalted against 50 mM HEPES (pH 7.5), 300 mM NaCl, 20 mM imidazole, and 5 mM β-mercaptoethanol. The TEV protease was added at a ratio of 1/50 (TEV/protein), and the sample was incubated at 4°C for 3 h. The cleaved protein was separated from the tags, the protease, and the contaminants by a second affinity purification on a chelating HiTrap FF crude column charged with Ni^{2+} ions. The flow-through was collected and directly loaded on a Heparin column (GE Healthcare) to re-

Table 1. Data collection and refinement statistics

Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, b, c (Å)	60.9, 69.8, 89.1
α, β, γ (°)	90, 90, 90
Resolution (Å)	50.0–2.0 (2.05–2.00)
R_{meas} (%)	5.1 (106.6)
$CC_{1/2}$ (%)	99.9 (72.8)
$I/\sigma I$	20.44 (2.21)
Completeness (%)	99.7 (100.0)
Redundancy	5.4 (5.6)
Refinement	
Resolution (Å)	28.7–2.00
Number of reflections	26035
R_{work} / R_{free} (%)	19.24/23.77
Number of atoms	
Protein	2325
Solvent	161
B-factors	
Protein	44.7
Solvent	45.1
Root mean square deviations	
Bond length (Å)	0.008
Bond angle (°)	1.096
Ramachandran plot (%)	
Favored region	97.9
Allowed region	2.1
Outliers region	0

move the remaining contaminants. Elution was performed with a linear salt gradient between 0.05 and 1 M of NaCl. The fractions containing the elution peak were analysed on a SDS-PAGE and pooled. The protein was further cleaned using gel filtration chromatography with a Superdex 75 column (GE Healthcare) in 20 mM HEPES (pH 7.5) containing 150 mM NaCl, 1 mM β-mercaptoethanol, 5% (v/v) glycerol, and 5 mM MgCl$_2$. The protein eluted from the gel filtration column as monomers and was concentrated (MWCO 30 KDa; Amicon) to 10 mg/ml. The concentrated protein was used immediately for crystallization or stored at −80°C. All protein samples were more than 95% pure as judged by Coomassie-stained SDS-PAGE gels.

Crystallization

The SHARP construct, encompassing residues 335 to 620, was crystallized at a concentration of 10 mg/ml (0.3 mM) using the sitting-drop vapour diffusion technique at 4°C, mixing an equal volume of the protein and the reservoir solution. Initial screening with nanolitre volumes of drops gave one crystal in one condition (Natrix screen, Hampton research); this was further optimized to obtain diffraction quality crystals. Final crystals were grown from a mixture of 0.5 μl of protein with 0.5 μl of mother liquor (0.1 M ammonium sulphate, 0.01 M magnesium chloride hexahydrate, 0.05M MES monohydrate pH 5.6, and 16% w/v polyethylene glycol 8000). Prior to flash freezing in liquid nitrogen, crystals were put into perfluoropolyether CryoOil (Hampton Research).

Data collection, structure determination, and refinement

Diffraction data were collected at the European Synchrotron Radiation Facility (ESRF, Grenoble) and processed using XDS (20). Data collection statistics are shown in Table 1. The structure of R2–3–4h was determined by molecular replacement using the PHASER programme from the CCP4 package (21). We used three ensembles, each one containing several RRM structures as models: Ensemble 1 (PDB codes: 3MD1, 2ADC, 2DNM, 2DGU, 2CQB), Ensemble 2 (PDB codes: X4AR, 2YTC, 4F26, 1WHY, 2CPZ), and Ensemble 3 (PDB codes: 2I38, 1WHY, 1×55, 2LCW, 2CPE). Because the quality of the original densities was not good enough to allow reconstruction of our protein model, we performed several cycles of density modification using the SOLOMON programme (22); this gave a readily interpretable electron density map. The final model was built using the COOT molecular graphics programme (23) and refined using routine phenix.refine from the PHENIX programme (24). We included TLS refinement using the following groups: group1, amino acids 335–411; group 2, amino acids 412–429; group 3, amino acids 430–513; group 4, amino acids 514–588; and group 5, amino acids 589–620 (25). The coordinates and the structure factors have been deposited in the PDB with accession code 4P6Q.

Electrophoretic mobility shift assay (EMSA)

Radiolabelled RNAs were synthesized *in vitro* using the T7 RNA polymerase (home made) in the presence of α^{32}P-UTP. The transcription reactions were performed as previously described (26). We used various DNA templates to produce the RNA sequences used in our EMSA. The purified and radiolabelled RNAs (0.1 pmol) were incubated with increasing concentrations of the different SHARP constructs for 20 min at room temperature in 20 mM HEPES (pH 7.5) containing 150 mM NaCl, 5 mM MgCl$_2$, and 0.5 U of Superase (Invitrogen). Samples were then mixed with native loading dye (50% glycerol, 0.2% bromophenol, and 0.2% xylene cyanol) and loaded on a 5% native polyacrylamide gel. Gels were run in standard 1X Tris-borate–EDTA buffer at 200 V for 2 h at 4°C and visualized using autoradiography.

For the competition assays, the cold RNAs corresponding to the H12–H13, H12, H13, and H7 sequences were synthesized *in vitro* using T7 RNA polymerase as indicated above. Purified RNAs were annealed at low concentration by heating them at 90°C for 5 min followed by a slow reduction of temperature to 30°C over an extra hour, and were subsequently kept on ice. The short single-stranded RNAs LA, LB, LC, UL, and USL were ordered (Thermo Scientific, Dharmacon), deprotected and resuspended in water before annealing as recommended by the manufacturers. The R2–3–4h/H12-H13 complex was prepared as described above. We then added the cold RNAs at different concentrations and incubated them for an additional 10 min. The samples were subsequently analysed on a 5% native gel as previously.

RESULTS

Structure of the SHARP–RRM fragment

The SHARP protein has three RRMs located in its N-terminal region. The same protein fragment has previously been shown to bind to and modulate the transcriptional activity induced by the SRA RNA (13). We determined the atomic structure of the protein construct encompassing residues 335 to 620 (SHARP–RRM or R2–3–4h) at a resolution of 2.0 Å. The model contains the three predicted RBDs plus two helices at the C-terminus (Figure 1A and B; see statistics in Table 1). The crystal structure revealed that the predicted RBDs fold into three individual RRMs with the typical βαββαβ topology. These RBDs were named RRM2, RRM3, and RRM4 as the N-terminal region of SHARP was predicted to have an additional RRM (encompassing residues 6 to 81). The SHARP protein fragment is organized in two blocks, made of the RRM2 on one hand and the RRM3-RRM4 on the other. A flexible linker connects these two blocks (Figure 1B). Furthermore, the C-terminal tail folds into two consecutive α-helices immediately after the RRM4 (Figure 1C and D). The first of these two helices interacts extensively with the RRM4 β-sheet surface, thereby covering the putative RNA-binding surface of this RRM (Figure 1C and D). The presence of a helical extension immediately following the predicted RBDs has been observed multiple times in other RRM-containing proteins (27–29). In a few cases, specific functions have been attributed to these C-terminal helices. For instance, the C-terminal helix C in the U1A protein is shown to reorient upon RNA association (30,31), and the helical extension
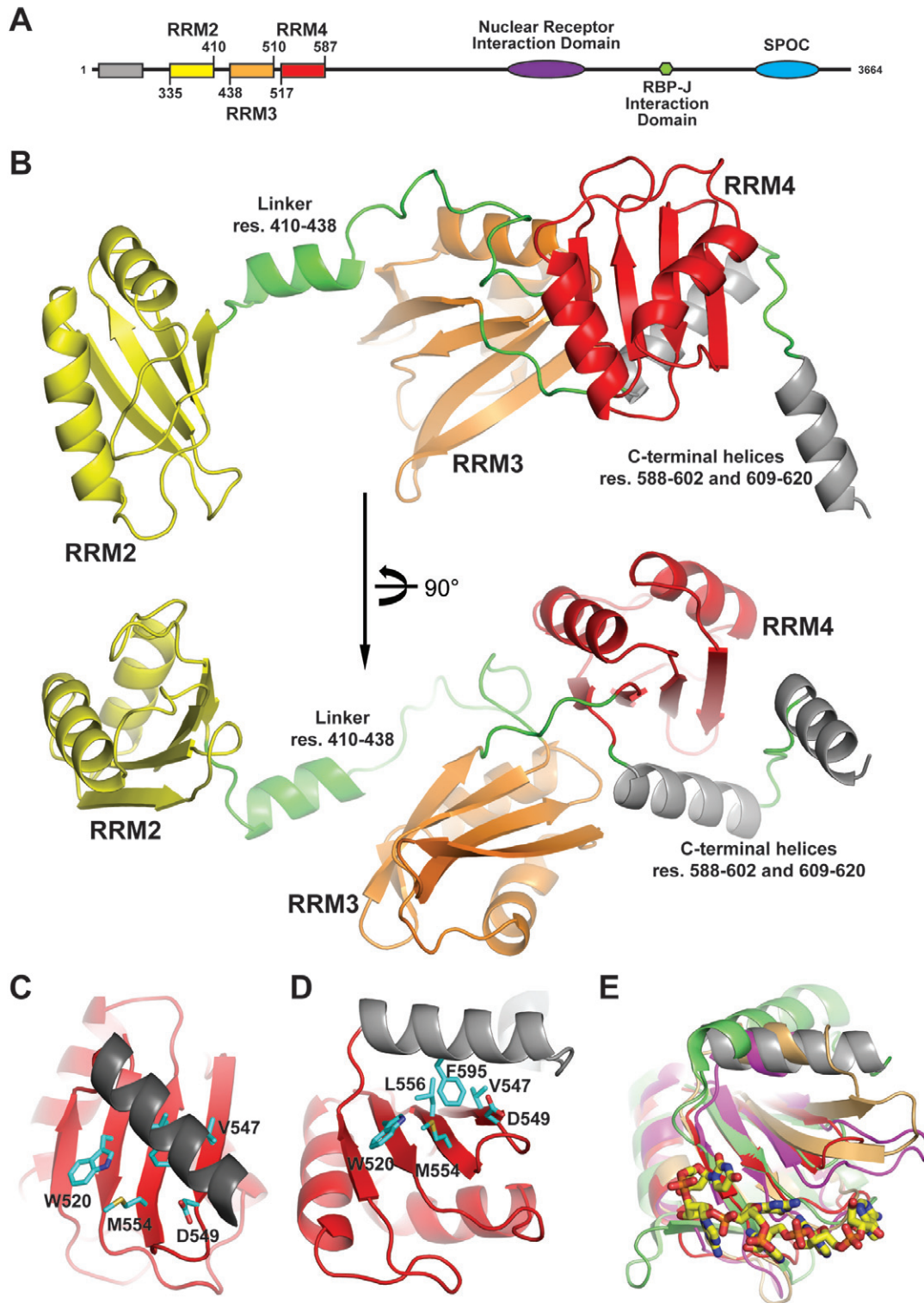
Figure 1. Overall structure of SHARP–RRMs. (**A**) Schematic representation of the domain organization of the SHARP protein. (**B**) The overall structure of R2–3–4h is shown: RRM2 is coloured in yellow, RRM3 in orange, RRM4 in red, C-terminal helices in grey, and linker regions are coloured in green. (**C**) and (**D**) Top (C) and side (D) views of the RRM4 plus the C-terminal α-helix. The protein is shown as a cartoon and coloured in red for the RRM and grey for the helix. (**E**) Superposition of various RRMs followed by C-terminal helices. The xRRM found in the p65 protein is coloured in green (PDB code 4EYT). The qRRM 1 and 2 from the hnRNP F in complex with the AGGGAU hexa-ribonucleotide are coloured in gold and magenta, respectively. The nucleotides are colored according to atom types (carbon: yellow; nitrogen: blue; oxygen: red; and phosphate: orange). SHARP is coloured in red with the helix coloured in grey. Proteins are shown as a cartoon and residues involved in intramolecular interactions in panels C and D are shown as sticks and coloured according to atom type (carbon: cyan; nitrogen: blue; and oxygen: red).

found in the splicing factor protein p14 is responsible for the tight association with the splicing factor SF3b155 (27). The hnRNP F protein contains three copies of the qRRM domains which have their β-sheet surfaces partially covered by a C-terminal helix (32). It was further shown that qR-RMs bind their RNA target using the loop regions rather than β-sheet surface (29). Another example is found in the recently described xRRM present in the p65 protein (19). This atypical type of RRM is present in members of the La protein family and is known to interact with specific sites of the telomerase RNA (33). The C-terminal helix found in the xRRM makes it capable of establishing a combination of single- and double-strand-mediated interactions with its RNA target (19). Structural superimposition of SHARP RRM4 including the C-terminal helix onto individual RRM domains from the hnRNP F or the p65 proteins clearly shows the conserved location of the additional secondary structure element found in the SHARP protein (Figure 1E). With the notable exception of U1A, every the RRMs that are followed by a C-terminal helix are unable to bind RNA through the β-sheet surface, suggesting that RRM4 could follow this trend (Figure 1D and E).

Comparison of the predicted domain organization of annotated SPEN protein family members, ranging from mammalian to plant species, underlines the conserved organization of the N-terminal region within this family (Supplementary Figure S1). This conserved region can be divided in two modules: one which contains an RRM (RRM2 in SHARP) and a second module containing two RRMs (RRM3 and RRM4 in SHARP). These two blocks are connected through a relatively long linker of ∼20–30 amino acids (Supplementary Figure S1). Interestingly, the plant proteins belonging to the SPEN family have an atypical primary sequence organization. In these species, the RRM corresponding to the human RRM2 is located after the RRM3/RRM4 block (Supplementary Figure S1). We suggest that, in the plant clade, the RRM2 coding sequence has been repositioned after the RRM3/RRM4-coding block at the primary sequence level. Such domain reorganization is not unusual in eukaryotic multidomain proteins, and may be due to the development of an alternative splicing mechanism (34,35). We used the Consurf server to map the sequence conservation onto the atomic surface of SHARP–RRM (36). This analysis clearly identified the RRM3/RRM4 interface as containing multiple residues highly conserved across eukaryotic species, with the plant proteins being the least conserved (Figure 2A, B and C; Supplementary Figure S2). This interface relies on amino acids from RRM3 and RRM4 domains and in particular residues R438, Q482, W520, and D585 (Figure 2D). Other residues connecting RRM2 to RRM3, RRM3 to RRM4 and to the C-terminal α-helix following RRM4 are also part of the interface (Figure 2D). Furthermore, the particular orientation observed between RRM3 and RRM4 appears to be unique when compared with other protein structures containing multiple RRMs (Supplementary Figure S3). Clearly, the respective orientation of RRM3 and RRM4 as well as the complex surface of interaction observed in our atomic model has not been reported before. Therefore, the presently described architecture of RRM3/RRM4 documents a new mode of interaction between RRMs.
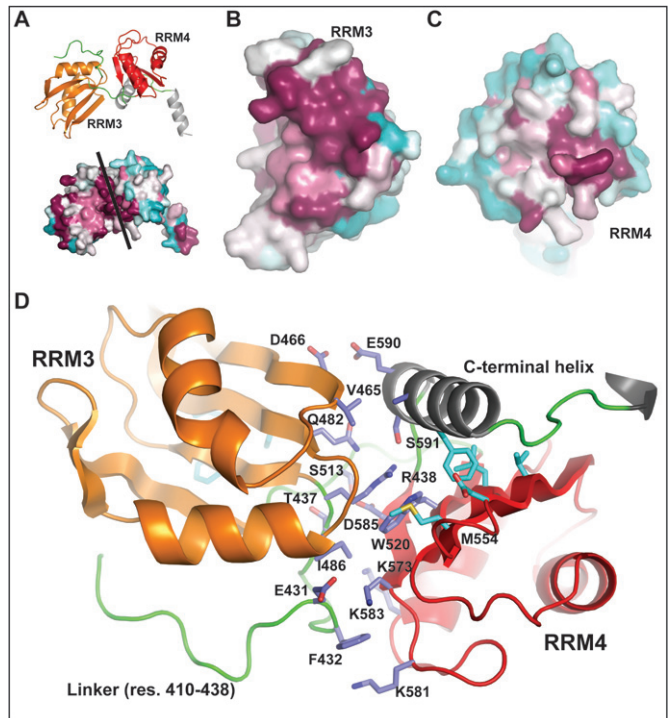

Figure 2. RRM3–RRM4 interface conservation. (**A**) RRM3 and RRM4 (shown as cartoon and coloured in orange and red, respectively) are in close contact. Bottom part shows the same protein fragment using a surface representation coloured according to surface conservation (cyan: low conservation and magenta: high conservation). (**B**) and (**C**) The RRM3–RRM4 interface is facing the reader, and the protein surface is coloured as in panel A. (**D**) Magnified view of the RRM3–RRM4 interface of SHARP: residues located at the interface are shown as sticks and coloured according to atom types (carbon: slate or cyan; nitrogen: blue; and oxygen: red).

To further validate the proposed modular organization of the SHARP N-terminal region consisting of the RRM2 loosely attached to the RRM3/RRM4 unit, we performed small angle X-ray scattering (SAXS) experiments on the SHARP–RRM fragments R2–3–4h and R3–4h (Supplementary Figure S4). The SAXS data are quite well fitted using either the atomic models presently obtained or using *ab initio* procedure (Supplementary Figure S4A). Docking the atomic structures corresponding to RRM2 and R3–4h fragments connected by a flexible linker into the envelope calculated with the R2–3–4h data is possible with a slight reorientation of RRM2 as expected from its loose attachment with the RRM3/RRM4 unit (Supplementary Figure S4B). The R3–4h fragment of the structure can be fitted into both envelopes without any changes, strongly suggesting that RRM3 and RRM4 domains form a stable unit even in the absence of RRM2 (Supplementary Figure S4B and C). We further suggest that this structural design likely reflects its importance for the SPEN protein function, most probably by governing their atypical RNA-binding properties (see below).

Identification of the RNA-binding properties of the SHARP–RRM

Using gel shift assays, we then investigated how the SHARP–RRM fragment could recognize nucleic acids. Typically, three aromatic side chains located on β-strands 1 and 3 are essential for RNA binding in canonical RRMs (for a review, see 17,37). In the SHARP protein, RRM2 (partially), RRM3 and RRM4 exhibit aromatic residues at the standard positions within their folded domains (Figure 3). However, the C-terminal helix of RRM4 is covering these conserved residues through multiple hydrophobic interactions (Figure 2D). The strong association of the α-helix with the β-sheet in RRM4 likely prevent RNA association at this location (Figures 2D and 3C). Moreover, the electrostatic surface potential of the SHARP–RRM fragment clearly indicate the existence of positively charged regions near the β-sheet of RRM3 as well as in the loop connecting RRM3 to RRM4 and in RRM4 (Supplementary Figure S5). To further understand how SHARP–RRM could interact with RNA, we measured RNA association with multiple SHARP constructs.

SHARP has two reported nucleic acid targets, the OC promoter region and the H12–H13 substructure found in the long non-coding SRA RNA (5,13,15). The later is an 85 nucleotide-long RNA element highly conserved which adopts a stable secondary structure (38). Based on our atomic model, we designed different protein mutants or truncations and tested their ability to bind to the H12–H13 RNA fragment (Figure 4). We were able to confirm that our crystallized construct was binding to the SRA RNA H12–H13 sequence with an apparent affinity constant in the low micromolar range ($K_{d,\ app} = 1.75\ \mu M$; Figure 4A and B). As the three RRMs can potentially bind nucleic acid molecules, we subsequently tested several constructions corresponding to different combinations of RRMs. We started with the construct R3–4h, where the RRM2 was deleted while keeping the RRM3–RRM4 block and the C-terminal helices intact (Figure 4C). This construct is able to recognize the H12–H13 RNA fragment with a similar affinity, clearly indicating the RRM2 is not essential for the stable recognition of the SRA RNA (Figure 4C). We then mutated several residues of the RRM3 predicted to be important for RNA binding: (i) three aromatic residues found in the RNP1 and RNP2 motifs (F441, Y478, and F480). (ii) Two lysine residues found in the β2–β3 loop and in the linker between RRM3 and RRM4 (K470 and K512, respectively). We verified that the construct with the RRM3 domain mutated (R3mut) was still stable and folded by gel filtration chromatography and circular dichroism (Supplementary Figures S6 and S7). Both measurements strongly suggested that the mutant is well folded (Supplementary Figures S6 and S7). Though folded, the R3mut construct was unable to form a stable complex with the H12–H13 RNA fragment, indicating that the mutated residues within RRM3 are important for RNA association (Figure 2D). We then tested RRM4 and the C-terminal helices. Unfortunately, the deletion of the C-terminal helices makes the protein fragment highly insoluble (data not shown). We thus decided to delete the entire region containing the RRM4 and the helices (R2–3 construct), which resulted in a sta-

ble and folded polypeptide (Supplementary Figures S6 and S7). We further expressed RRM3 alone and showed that it also behaved as a stable folded domain (Supplementary Figures S6 and S7). EMSA experiments with the R2–3 protein revealed that the RNA-binding capacity of this construct was impaired, indicating that the RRM4 domain and the helices also play a crucial role in RNA association (Figure 4E). Then, we tested if RRM4 could bind the RNA on its own by mutating the canonical residues for RNA binding in both RRM2 and RRM3 (R2mutR3mut construct, mutated residues are listed in the Materials and Methods section). The RNA association was strongly reduced to a background level, indicating that RRM4 and the C-terminal helices cannot bind stably the RNA on their own (Figure 4F).

As we show above, RRM3 is tightly associated with RRM4 in our crystal structure and this association appears conserved within the SPEN family. These RNA-binding experiments strongly suggest that the RNA-binding properties of SHARP are mostly due to the newly characterized RRM3/RRM4 block (Figures 2B and 4C). Although the exact role of each RRM is not yet fully apprehended, we propose that there is a strong cooperative effect between the two RRMs based on our atomic structure and on the RNA-binding experiments. The precise positioning of the two RRMs in space through their specific surface of interaction leads to high affinity binding properties, as previously observed for other multi-RRM-containing protein, such as PTB (39,40).

The RNA structural context is important for SHARP–RRM/SRA interaction

To further understand the RNA-binding capacities of SHARP, we investigated the RNA sequence driving the association, i.e. the 85 nucleotide-long H12–H13 fragment of SRA. We measured the ability of different SRA RNA fragments to displace the preformed R2–3–4h/H12–H13 complex (Figure 5A). We first tested the full-length H12–H13, which competed very efficiently, while a single-stranded A_{15} oligomer did not displace the preformed SHARP–RRM/SRA RNA complex (Figure 5B, lanes 5–7 and 8–10). Most RRM domains have been shown to bind to short unstructured ssRNAs (41,42). The CISBP-RNA server was used to predict the best RNA-binding motif for the RRM3 (43). The best RNA-binding sequence predicted is GUGUG, and the second best predicted sequence is ACACA. Based on these two predictions, an RNA sequence with an alternative purine-pyrimidine sequence is likely to bind to RRM3. Such a consensus sequence is partially found in the LA loop. We tested three individual loop regions of H12–H13, i.e. LA, LB, and LC region, which did not displace the preformed complex even when added in 100-fold molar excess (Figure 5B, lanes 11–19). This indicates that the SHARP association with RNA is unlikely to be driven by non-specific interaction mediated by charged residues, as observed with proteins bound non-specifically to RNA (44,45). These experiments with individual loops also suggested that the RNA sequence responsible for the stable and specific association with SHARP–RRMs is not be restricted to single-stranded regions. We then split the H12–H13 RNA into two parts, namely H12
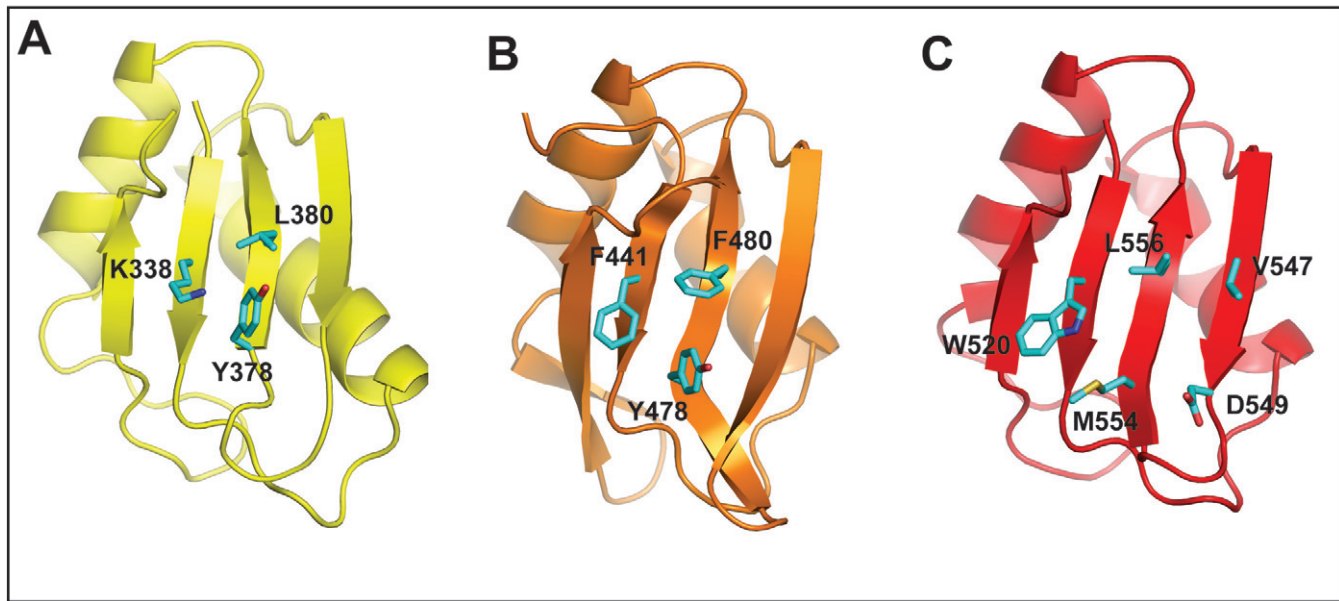
Figure 3. Individual RRM structures and the consensus amino acids for RNA binding. (**A**), (**B**), and (**C**) RRM2, RRM3, and RRM4 are shown as a cartoon and coloured in yellow, orange, and red, respectively. Amino acids at the canonical position for RNA interaction are shown as sticks and coloured as in Figure 1.

and H13, and performed similar experiments in an attempt to narrow down our search for the specific determinant behind SHARP/RNA recognition (Figure 5C). Both the H12 and the H13 RNA fragments were able to displace the preformed complex, although they required larger molar excess than the full H12–H13 fragment (Figure 5C, lanes 7–10 and 11–14). This result indicated that the LA but also the LB or LC sequences, when embedded in stem-loop structures, were efficiently binding to SHARP. In order to test for the possibility that the specific SHARP/RNA interactions were due to a folded H12–H13 RNA, we performed competition experiments using three unrelated sequences with diverse primary and secondary structures. We chose another fragment of SRA (H7) as a further example of RNA with multiple predicted stem-loop secondary structures (38). We used a 17 nt ssRNA sequence (UL), which has multiple purine-pyrimidine dinucleotides. The third RNA corresponded to the same UL sequence plus a short, well-characterized stem loop (USL). We observed that the H7 and USL fragments competed with the preformed protein/RNA complex, while the ssRNA sequence UL did not compete for the R2–3–4h association (Figure 5D). This experiment clearly indicated that secondary structures within the RNA are necessary for SHARP/RNA interaction. Notably, the H7 and the USL sequences needed to be added in larger molar excess than the H12–H13 fragment to achieve the same competitive effect, indicating that SHARP–RRM/H12–H13 association is specific (compare lanes 5, 9, and 17 in Figure 5D). Examined together, these experiments point towards a complex mode of association between SHARP–RRMs and the H12–H13 fragment: a mode which is not limited to an association mediated by a 3–4 nucleotide single-stranded sequence likely located in the loop LA and the β-sheet surface of RRM3. The H12–H13 secondary structure was recently shown to be identical

both in isolation and within the full-length SRA RNA (46). It seems therefore unlikely that SHARP–RRM recognizes a particular tertiary fold adopted by the H12–H13 fragment. We also tested the competing capacities of the single-stranded and duplex DNA sequences found in the OC promoter region, previously identified as a specific binding site for the SHARP protein (5). Even though SHARP is able to bind to the OC promoter ssDNA or duplex DNA (Supplementary Figure S8), none of the DNA sequences displaced the preformed protein–RNA complex in our competition assays (Supplementary Figure S8). This further validates the binding specificity of SHARP–RRMs for the SRA RNA H12–H13 element. In an ultimate effort to identify the contact points between SHARP and the SRA RNA, we decided to crosslink the H12–H13 SRA RNA to the R3–4h protein construct. After UV irradiation of the complex, we isolated and primer extended the crosslinked RNA (Supplementary Figure S9). Several weak crosslinks were observed in the single-stranded regions of the H12–H13 RNA (Supplementary Figure S9) reinforcing the idea that SHARP has multiple contact points within the RNA. The major crosslinking events happened at the dinucleotides UG of the loop LA (positions 36–37 and 39–40 of the H12–H13 element; Supplementary Figure S9), which corroborate the predicted binding site for the RRM3, i.e. alternating pyrimidine/purine sequence.

DISCUSSION

The SHARP protein is an important regulator of various transcriptional processes such as nuclear receptor-mediated responses, Notch-mediated transcriptional activation, or osteoblast differentiation (5,12,13). SHARP belongs to a protein family characterized by a SPOC domain at the C-terminus, and at least three RBDs at the N-terminus (Fig-
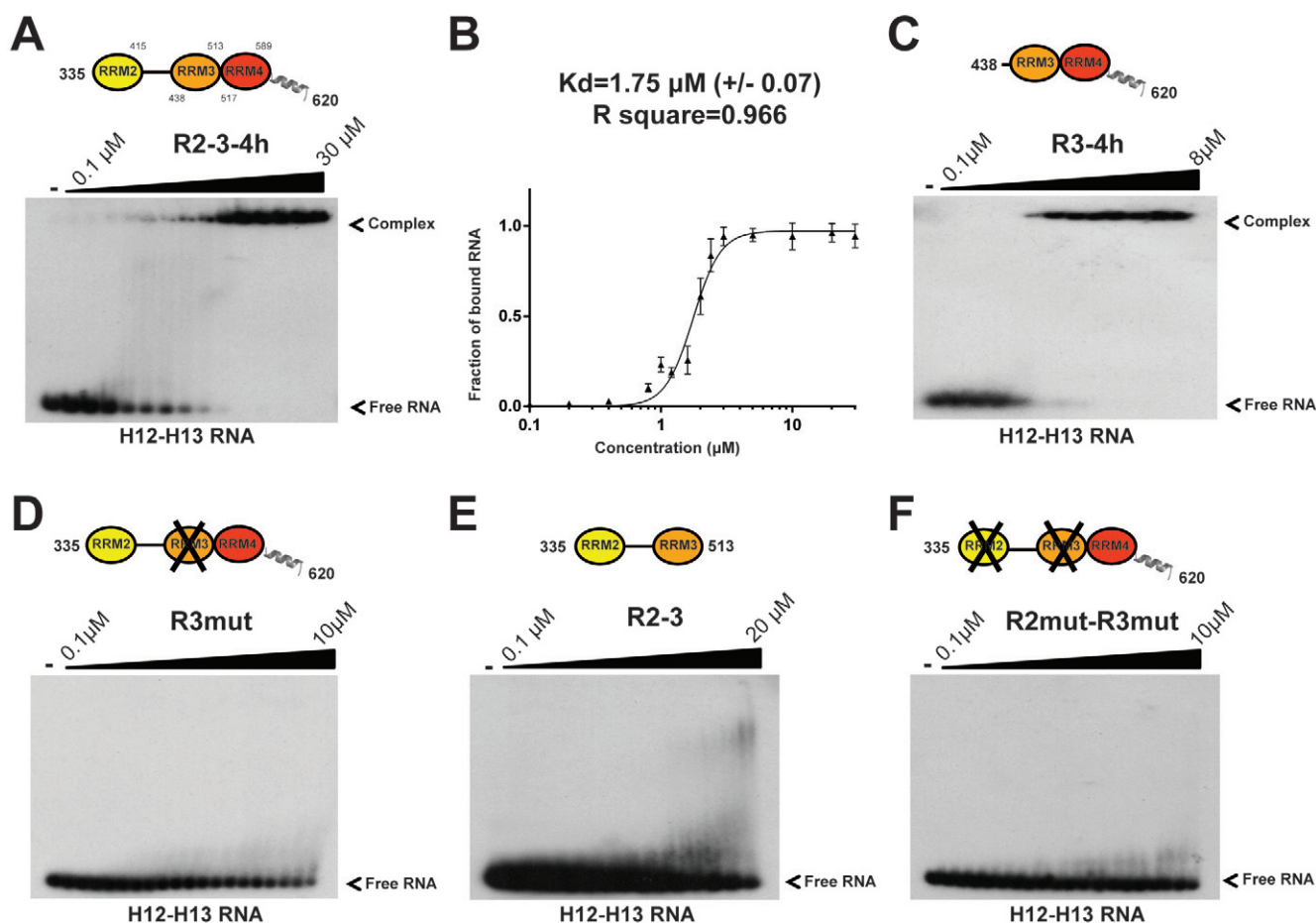
Figure 4. Identification of the protein region responsible for the SRA RNA association. (**A**) Representative EMSA used to quantify R2–3–4h construct association with the SRA RNA fragment H12–H13. (**B**) Quantification of the EMSA experiment shown in panel A. The bound fraction was quantified and analysed using the Hill Equation, and the fit is shown as a solid line. The calculated dissociation constant ($K_{d,\,app}$) and the quality of the fit are indicated. (**C**) EMSA autoradiogram obtained when measuring the RNA-binding capacity of the R3-R4h construct. (**D**) EMSA autoradiogram showing the very weak RNA-binding activity of the R3mut construct. (**E**) EMSA autoradiogram showing the poor RNA-binding activity of the R2–3 construct. (**F**) EMSA autoradiogram obtained using the R2mutR3mut construct and indicating that the RRM4 plus the two helices participate only marginally in RNA association.

ure 3A). It is a large multidomain protein of ∼400 kDa, which renders its structural characterization very difficult. We have solved the structure of the N-terminal protein fragment containing three RRM domains, enhancing our understanding of proteins containing RRM that are associated with negative transcriptional regulation (47–49).

With the increasing amount of structural information on multidomain proteins, it has become clear that proteins with multiple RRMs generally adopt unique architectures; this helps them achieve higher binding affinity and specificity for their RNA targets (50). One central question associated with such proteins is the functional relevance of these interdomain contacts. We identified an unexpected and stable interaction between RRM3 and RRM4. The surface of interaction found between these two protein subdomains is highly conserved throughout evolution. The RRM3 and RRM4 domain association is stable as observed by SAXS measurement (Supplementary Figure S4). Furthermore, we show that both RRMs are needed in order to achieve a stable RNA association (Figure 4C–F). The RRM3 domain contains the canonical residues allow-

ing this domain to bind RNA. However, the association of the SHARP–RRM fragment is not limited to the short single-stranded RNA-binding region present in the RRM3 as none of the single-stranded loops present in the H12–H13 sequence can compete efficiently with the binding of the full-length RNA fragment. We revealed that the presence of double-stranded RNA sequences is important to displace the preformed complex between SHARP–RRM and the SRA RNA. The β-sheet surface of RRM4 strongly interacts with an α-helix a situation reminiscent to the one encountered in various other RRMs, some of them shown to bind a double-stranded region within their RNA targets (19). We propose that the RRM3/RRM4 unit contributes to the overall RNA-binding affinity of SHARP, with a preference for alternating UG sequence located in structured loop (stem loop). We also show that SHARP affinity is higher for ribonucleotide sequences as ssDNA or duplex DNA sequences previously shown to interact with SHARP *in vitro* (5) were not able to compete efficiently for the binding (Supplementary Figure S8). Our structural and biochemical studies unravelled a new architecture for a pro-

Figure 5. SHARP–RRM association with RNA relies on stem-loop structure. (**A**) Secondary structures and sequences of the RNA molecules used for the competition assays reported in panels B, C, and D. The structural architecture of the H12–H13 fragment is based on that established by Novikova *et al.* (38). The loop regions and the single-strand RNA are coloured in blue (LA), yellow (LB), red (LC), and green (UL). (**B**) RNA-binding competition assays to determine the loop sequences mediating association between SHARP–RRM (R2–3–4h construct) and the SRA RNA H12–H13. Free RNA and R2–3–4h complex are shown in lanes 1 and 2. Competition assays were performed with the following RNAs: loop LA (lanes 11–13), loop LB (lanes 14–16), and loop LC (lanes 17–19). (**C**) Similar competition experiments performed with individual substructures H12 and H13 from the SRA RNA. Competition assays were performed with H12 RNA (lanes 7–10) and H13 RNA (lanes 11–14). (**D**) Competition experiments performed with various unrelated RNAs showing the importance of a stem region for the competitive capacity of the RNA. Competition assays were performed with cold H7 RNA (lanes 7–10), unrelated loop RNA (UL, lanes 11–14), and unrelated stem-loop RNA (USL, lanes 15–18). Concentrations of the cold RNAs were added in molar excess as indicated in each panel. Positive and negative controls for efficient competitions are identical in panels B, C, and D (labelled H12–H13 or A_{15}, respectively).

tein containing multi-RRM, increasing the already broad repertoire of nucleic acid recognition modes of these domains and underlined the preference of this protein for UG-containing structured stem loops.

Although the exact function of the SHARP protein is still the subject of intense research, we would like to speculate that the key to understanding its various reported interacting partners lies in the highly atypical nucleic acid binding properties presently observed (51,52). SHARP is a protein scaffold that is found in various transcriptional regulatory complexes acting around the polymerase holoenzyme. Moreover, long ncRNAs have been proposed to mimic an open promoter structure to regulate polymerase II-mediated transcription (53). We propose that the SRA RNA may mimic an active transcription unit, which would explain its capacity to recruit SHARP and possibly could help understand the many functions attributed to this long ncRNA.

ACCESSION NUMBER

Accession code 4P6Q.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGMENTS

FUNDING

REFERENCES

1. Dickson,B.J., van der Straten,A., Dominguez,M. and Hafen,E. (1996) Mutations modulating Raf signaling in Drosophila eye development. *Genetics*, **142**, 163–171.
2. Kolodziej,P.A., Jan,L.Y. and Jan,Y.N. (1995) Mutations that affect the length, fasciculation, or ventral orientation of specific sensory axons in the Drosophila embryo. *Neuron*, **15**, 273–286.
3. Wiellette,E.L., Harding,K.W., Mace,K.A., Ronshaugen,M.R., Wang,F.Y. and McGinnis,W. (1999) Spen encodes an RNP motif protein that interacts with Hox pathways to repress the development of head-like sclerites in the Drosophila trunk. *Development*, **126**, 5373–5385.
4. Yabe,D., Fukuda,H., Aoki,M., Yamada,S., Takebayashi,S., Shinkura,R., Yamamoto,N. and Honjo,T. (2007) Generation of a conditional knockout allele for mammalian Spen protein Mint/SHARP. *Genesis*, **45**, 300–306.
5. Newberry,E.P., Latifi,T. and Towler,D.A. (1999) The RRM domain of MINT, a novel Msx2 binding protein, recognizes and regulates the rat osteocalcin promoter. *Biochemistry*, **38**, 10678–10690.
6. Obata,J., Yano,M., Mimura,H., Goto,T., Nakayama,R., Mibu,Y., Oka,C. and Kawaichi,M. (2001) p48 subunit of mouse PTF1 binds to RBP-Jkappa/CBF-1, the intracellular mediator of Notch signalling, and is expressed in the neural tube of early stage embryos. *Genes Cells*, **6**, 345–360.
7. Rebay,I., Chen,F., Hsiao,F., Kolodziej,P.A., Kuang,B.H., Laverty,T., Suh,C., Voas,M., Williams,A. and Rubin,G.M. (2000) A genetic screen for novel components of the Ras/Mitogen-activated protein kinase signaling pathway that interact with the yan gene of Drosophila identifies split ends, a new RNA recognition motif-containing protein. *Genetics*, **154**, 695–712.
8. Kuang,B., Wu,S.C., Shin,Y., Luo,L. and Kolodziej,P. (2000) Split ends encodes large nuclear proteins that regulate neuronal cell fate and axon extension in the Drosophila embryo. *Development*, **127**, 1517–1529.
9. Chen,F. and Rebay,I. (2000) Split ends, a new component of the Drosophila EGF receptor pathway, regulates development of midline glial cells. *Curr. Biol.*, **10**, 943–946.
10. Ariyoshi,M. and Schwabe,J.W. (2003) A conserved structural motif reveals the essential transcriptional repression function of Spen proteins and their role in developmental signaling. *Genes Dev.*, **17**, 1909–1920.
11. Mikami,S., Kanaba,T., Ito,Y. and Mishima,M. (2013) NMR assignments of SPOC domain of the human transcriptional corepressor SHARP in complex with a C-terminal SMRT peptide. *Biomol. NMR Assign*, **7**, 267–270.
12. Oswald,F., Kostezka,U., Astrahantseff,K., Bourteele,S., Dillinger,K., Zechner,U., Ludwig,L., Wilda,M., Hameister,H., Knochel,W. *et al.* (2002) SHARP is a novel component of the Notch/RBP-Jkappa signalling pathway. *EMBO J.*, **21**, 5417–5426.
13. Shi,Y., Downes,M., Xie,W., Kao,H.Y., Ordentlich,P., Tsai,C.C., Hon,M. and Evans,R.M. (2001) Sharp, an inducible cofactor that integrates nuclear receptor repression and activation. *Genes Dev.*, **15**, 1140–1151.
14. Shi,Y., Hon,M. and Evans,R.M. (2002) The peroxisome proliferator-activated receptor delta, an integrator of transcriptional repression and nuclear receptor signaling. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 2613–2618.
15. Hatchell,E.C., Colley,S.M., Beveridge,D.J., Epis,M.R., Stuart,L.M., Giles,K.M., Redfern,A.D., Miles,L.E., Barker,A., MacDonald,L.M. *et al.* (2006) SLIRP, a small SRA binding protein, is a nuclear receptor corepressor. *Mol. Cell*, **22**, 657–668.
16. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al.* (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
17. Clery,A., Blatter,M. and Allain,F.H. (2008) RNA recognition motifs: boring? Not quite. *Curr. Opin. Struct. Biol.*, **18**, 290–298.
18. Muto,Y. and Yokoyama,S. (2012) Structural insight into RNA recognition motifs: versatile molecular Lego building blocks for biological systems. *Wiley Interdiscip. Rev. RNA*, **3**, 229–246.
19. Singh,M., Wang,Z., Koo,B.K., Patel,A., Cascio,D., Collins,K. and Feigon,J. (2012) Structural basis for telomerase RNA recognition and RNP assembly by the holoenzyme La family protein p65. *Mol. Cell*, **47**, 16–26.
20. Kabsch,W. (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 125–132.
21. McCoy,A.J., Grosse-Kunstleve,R.W., Adams,P.D., Winn,M.D., Storoni,L.C. and Read,R.J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.*, **40**, 658–674.
22. Abrahams,J.P. and Leslie,A.G. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallogr. D Biol. Crystallogr.*, **52**, 30–42.
23. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.
24. Adams,P.D., Afonine,P.V., Bunkoczi,G., Chen,V.B., Davis,I.W., Echols,N., Headd,J.J., Hung,L.W., Kapral,G.J.,

Grosse-Kunstleve,R.W. *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 213–221.

25. Winn,M.D., Isupov,M.N. and Murshudov,G.N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement. *Acta Crystallogr. D Biol. Crystallogr.*, **57**, 122–133.

26. Thore,S., Leibundgut,M. and Ban,N. (2006) Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science*, **312**, 1208–1211.

27. Schellenberg,M.J., Edwards,R.A., Ritchie,D.B., Kent,O.A., Golas,M.M., Stark,H., Luhrmann,R., Glover,J.N. and MacMillan,A.M. (2006) Crystal structure of a core spliceosomal protein interface. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 1266–1271.

28. Rupert,P.B., Xiao,H. and Ferre-D'Amare,A.R. (2003) U1A RNA-binding domain at 1.8 A resolution. *Acta Crystallogr. D Biol. Crystallogr.*, **59**, 1521–1524.

29. Dominguez,C., Fisette,J.F., Chabot,B. and Allain,F.H. (2010) Structural basis of G-tract recognition and encaging by hnRNP F quasi-RRMs. *Nat. Struct. Mol. Biol.*, **17**, 853–861.

30. Avis,J.M., Allain,F.H., Howe,P.W., Varani,G., Nagai,K. and Neuhaus,D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: the role of C-terminal residues in structure stability and RNA binding. *J. Mol. Biol.*, **257**, 398–411.

31. Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and Nagai,K. (1994) Crystal structure at 1.92 A resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

32. Dominguez,C. and Allain,F.H. (2006) NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res.*, **34**, 3634–3645.

33. Singh,M., Choi,C.P. and Feigon,J. (2013) xRRM: A new class of RRM found in the telomerase La family protein p65. *RNA Biol.*, **10**, 353–359.

34. Szilagyi,A., Zhang,Y. and Zavodszky,P. (2012) Intra-chain 3D segment swapping spawns the evolution of new multidomain protein architectures. *J. Mol. Biol.*, **415**, 221–235.

35. Light,S. and Elofsson,A. (2013) The impact of splicing on protein domain architecture. *Curr. Opin. Struct. Biol.*, **23**, 451–458.

36. Landau,M., Mayrose,I., Rosenberg,Y., Glaser,F., Martz,E., Pupko,T. and Ben-Tal,N. (2005) ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.*, **33**, W299–302.

37. Daubner,G.M., Clery,A. and Allain,F.H. (2013) RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.*, **23**, 100–108.

38. Novikova,I.V., Hennelly,S.P. and Sanbonmatsu,K.Y. (2012) Structural architecture of the human long non-coding RNA, steroid receptor RNA activator. *Nucleic Acids Res.*, **40**, 5034–5051.

39. Song,Y., Tzima,E., Ochs,K., Bassili,G., Trusheim,H., Linder,M., Preissner,K.T. and Niepmann,M. (2005) Evidence for an RNA chaperone function of polypyrimidine tract-binding protein in picornavirus translation. *RNA*, **11**, 1809–1824.

40. Maynard,C.M. and Hall,K.B. (2010) Interactions between PTB RRMs induce slow motions and increase RNA binding affinity. *J. Mol. Biol.*, **397**, 260–277.

41. Cook,K.B., Kazan,H., Zuberi,K., Morris,Q. and Hughes,T.R. (2011) RBPDB: a database of RNA-binding specificities. *Nucleic Acids Res.*, **39**, D301–308.

42. Auweter,S.D., Oberstrass,F.C. and Allain,F.H. (2006) Sequence-specific binding of single-stranded RNA: is there a code for recognition? *Nucleic Acids Res.*, **34**, 4943–4959.

43. Ray,D., Kazan,H., Cook,K.B., Weirauch,M.T., Najafabadi,H.S., Li,X., Gueroussov,S., Albu,M., Zheng,H., Yang,A. *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. *Nature*, **499**, 172–177.

44. Sengoku,T., Nureki,O., Nakamura,A., Kobayashi,S. and Yokoyama,S. (2006) Structural basis for RNA unwinding by the DEAD-box protein Drosophila Vasa. *Cell*, **125**, 287–300.

45. Westover,K.D., Bushnell,D.A. and Kornberg,R.D. (2004) Structural basis of transcription: separation of RNA from DNA by RNA polymerase II. *Science*, **303**, 1014–1016.

46. Novikova,I.V., Dharap,A., Hennelly,S.P. and Sanbonmatsu,K.Y. (2013) 3S: Shotgun secondary structure determination of long non-coding RNAs. *Methods*, **63**, 170–177.

47. DeAngelo,D.J., DeFalco,J., Rybacki,L. and Childs,G. (1995) The embryonic enhancer-binding protein SSAP contains a novel DNA-binding domain which has homology to several RNA-binding proteins. *Mol. Cell Biol.*, **15**, 1254–1264.

48. Bertolotti,A., Lutz,Y., Heard,D.J., Chambon,P. and Tora,L. (1996) hTAF(II)68, a novel RNA/ssDNA-binding protein with homology to the pro-oncoproteins TLS/FUS and EWS is associated with both TFIID and RNA polymerase II. *EMBO J.*, **15**, 5022–5031.

49. Zhang,D., Paley,A.J. and Childs,G. (1998) The transcriptional repressor ZFM1 interacts with and modulates the ability of EWS to activate transcription. *J. Biol. Chem.*, **273**, 18086–18091.

50. Mackereth,C.D. and Sattler,M. (2012) Dynamics in multi-domain protein recognition of RNA. *Curr. Opin. Struct. Biol.*, **22**, 287–296.

51. Malovannaya,A., Lanz,R.B., Jung,S.Y., Bulynko,Y., Le,N.T., Chan,D.W., Ding,C., Shi,Y., Yucer,N., Krenciute,G. *et al.* (2011) Analysis of the human endogenous coregulator complexome. *Cell*, **145**, 787–799.

52. Sierra,O.L., Cheng,S.L., Loewy,A.P., Charlton-Kachigian,N. and Towler,D.A. (2004) MINT, the Msx2 interacting nuclear matrix target, enhances Runx2-dependent activation of the osteocalcin fibroblast growth factor response element. *J. Biol. Chem.*, **279**, 32913–32923.

53. Oyoshi,T. and Kurokawa,R. (2012) Structure of noncoding RNA is a determinant of function of RNA binding proteins in transcriptional regulation. *Cell Biosci.*, **2**, 1.