nature structural & molecular biology

Structural basis for dimethylarginine recognition by the Tudor domains of human SMN and SPF30 proteins

Konstantinos Tripsianes[1,2], Tobias Madl[1,2], Martin Machyna[3], Dimitrios Fessas[4], Clemens Englbrecht[5], Utz Fischer[5], Karla M Neugebauer[3] & Michael Sattler[1,2]

Arginine dimethylation plays critical roles in the assembly of ribonucleoprotein complexes in pre-mRNA splicing and piRNA pathways. We report solution structures of SMN and SPF30 Tudor domains bound to symmetric and asymmetric dimethylated arginine (DMA) that is inherent in the RNP complexes. An aromatic cage in the Tudor domain mediates dimethylarginine recognition by electrostatic stabilization through cation-π interactions. Distinct from extended Tudor domains, dimethylarginine binding by the SMN and SPF30 Tudor domains is independent of proximal residues in the ligand. Yet, enhanced micromolar affinities are obtained by external cooperativity when multiple methylation marks are presented in arginine- and glycine-rich peptide ligands. A hydrogen bond network in the SMN Tudor domain, including Glu134 and a tyrosine hydroxyl of the aromatic cage, enhances cation-π interactions and is impaired by a mutation causing an E134K substitution associated with spinal muscular atrophy. Our structural analysis enables the design of an optimized binding pocket and the prediction of DMA binding properties of Tudor domains.

Arginine methylation has long been observed in RNA-binding proteins and histones, but it has only recently been implicated in a variety of cellular processes, including RNA processing, transcriptional regulation, trafficking, signal transduction and gametogenesis[1,2]. Members of a broad family of protein arginine methyltransferases (PRMTs) generate asymmetric DMA (aDMA; PRMT1) by placing two methyl groups on one of the terminal guanidino nitrogen atoms of arginine, or symmetric DMA (sDMA; PRMT5) by placing one methyl group on each of the terminal guanidino nitrogen atoms, whereas monomethylated arginine (MMA) is probably generated as an intermediate *en route* to dimethylation[1]. Despite the increasing number of protein targets for arginine methylation[3], the only modules known so far to interpret dimethylarginine signatures are the Tudor domains of survival motor neuron protein (SMN), survival of motor neuron–related splicing factor 30 (SPF30), staphylococcal nuclease domain–containing protein 1 (SND1; also known as p100) and Tudor domain–containing protein (TDRD) family members[4–7]. SMN and SPF30 (**Fig. 1a**) play critical roles in the assembly of uridine-rich small nuclear ribonucleoprotein complexes (U snRNPs)[8–10] and in pre-mRNA splicing[11,12], respectively. U snRNPs have in common a seven-membered ring of Sm proteins, which forms around a conserved RNA sequence. Both SMN and SPF30 contain single Tudor domains that have been shown to associate directly with a subset of Sm proteins through sDMA modifications in their arginine- and glycine-rich C-terminal tails (**Fig. 1b** and refs. 4–6,13–16). In addition, the presence of SMN in Cajal bodies (spherical suborganelles found in the nuclei of proliferative or metabolically active cells) has been shown to depend on sDMA modifications of an arginine- and glycine-rich region located at the C-terminal end of coilin[17] (one of the main components of Cajal bodies), suggesting that recognition of sDMA marks in coilin determines SMN's subnuclear localization in Cajal bodies[18].

More recently, it was found that PIWI-ribonucleoprotein complexes (piRNPs) consisting of PIWI proteins and PIWI-interacting RNAs (piRNAs) can silence transposable elements in the germline during early development. Biogenesis of piRNAs depends on the recognition of sDMA modifications in the N-terminal regions of PIWI proteins (**Fig. 1b**) by TDRD Tudor domains[2,19–21]. A common feature of Tudor proteins that participate in the piRNA pathway is an extended architecture, whereby the basic Tudor domain is fused to an oligonucleotide- and oligosaccharide-binding (OB) fold. Two recent structural studies have underscored the requirement of the OB fold in assisting the Tudor domain for binding sDMA-containing peptides by interaction with nonmethylated residues adjacent to the modification[22,23].

Here, we set out to address the structural and energetic determinants of dimethylarginine recognition by the prototypic Tudor domains of SMN and SPF30. These proteins contain only the basic Tudor fold, which alone is sufficient for binding arginine- and glycine-rich peptides containing sDMA and aDMA with low micromolar affinity. Our study highlights unprecedented features in the recognition mode of arginine dimethylation and should greatly assist in understanding the evolution and functional role of the Tudor fold.

[1]Institute of Structural Biology, Helmholtz Zentrum München, Neuherberg, Germany. [2]Munich Center for Integrated Protein Science at the Chair of Biomolecular NMR, Department Chemie, Technische Universität München, Garching, Germany. [3]Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [4]Dipartimento di Tecnologie Alimentari e Microbiologiche, Università di Milano, Milano, Italy. [5]Department of Biochemistry, Theodor Boveri Institute, University of Würzburg, Würzburg, Germany. Correspondence should be addressed to M.S. (sattler@helmholtz-muenchen.de).

Received 25 August 2010; accepted 14 October 2011; published online 20 November 2011; doi:10.1038/nsmb.2185

RESULTS

Thermodynamic analysis of Tudor-dimethylarginine interactions

Binding assays using isothermal titration calorimetry (ITC) demonstrated that the SMN and SPF30 Tudor domains do not bind arginine, as well as lysine or its (mono-, di- and tri-) methylated forms (data not shown). Arginine methylation, on the other hand, induced measurable affinities, with binding increasing from MMA to aDMA to sDMA (**Fig. 1** and **Supplementary Table 1a**). SMN showed higher affinity for all ligands compared with SPF30. Binding affinities for an Sm B/B′-derived pentapeptide harboring a single sDMA residue were nearly equivalent to those of the isolated sDMA. Similar binding affinities and stoichiometries were also observed for the Sm D1 hexapeptide containing two sDMAs. Thus, a second sDMA motif separated by a single glycine residue did not enhance the binding affinity (**Supplementary Fig. 1**). However, the picture changed substantially when the long peptide derived from the Sm D3 tail, which harbors four sDMA

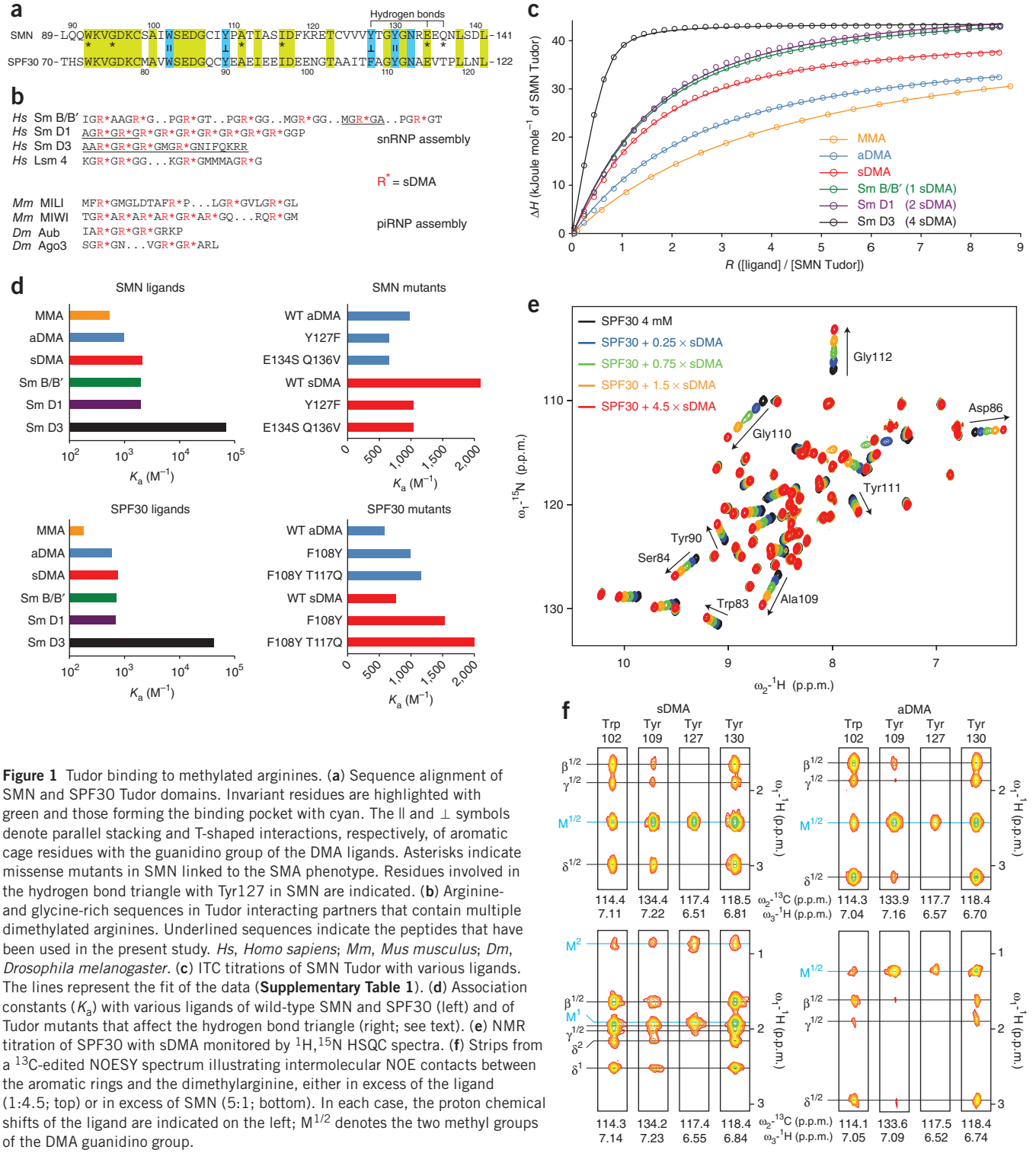

Figure 1 Tudor binding to methylated arginines. (**a**) Sequence alignment of SMN and SPF30 Tudor domains. Invariant residues are highlighted with green and those forming the binding pocket with cyan. The ∥ and ⊥ symbols denote parallel stacking and T-shaped interactions, respectively, of aromatic cage residues with the guanidino group of the DMA ligands. Asterisks indicate missense mutants in SMN linked to the SMA phenotype. Residues involved in the hydrogen bond triangle with Tyr127 in SMN are indicated. (**b**) Arginine- and glycine-rich sequences in Tudor interacting partners that contain multiple dimethylated arginines. Underlined sequences indicate the peptides that have been used in the present study. *Hs*, *Homo sapiens*; *Mm*, *Mus musculus*; *Dm*, *Drosophila melanogaster*. (**c**) ITC titrations of SMN Tudor with various ligands. The lines represent the fit of the data (**Supplementary Table 1**). (**d**) Association constants (K_a) with various ligands of wild-type SMN and SPF30 (left) and of Tudor mutants that affect the hydrogen bond triangle (right; see text). (**e**) NMR titration of SPF30 with sDMA monitored by $^{1}H,^{15}N$ HSQC spectra. (**f**) Strips from a ^{13}C-edited NOESY spectrum illustrating intermolecular NOE contacts between the aromatic rings and the dimethylarginine, either in excess of the ligand (1:4.5; top) or in excess of SMN (5:1; bottom). In each case, the proton chemical shifts of the ligand are indicated on the left; $M^{1/2}$ denotes the two methyl groups of the DMA guanidino group.

Table 1 NMR and refinement statistics for the Tudor–DMA complexes

	SMN–sDMA	SPF30–sDMA	SMN–aDMA	SPF30–aDMA
NMR distance and dihedral restraints				
Distance restraints				
Total NOE	1,865	1,747	1,862	1,811
Intra-residue	374	310	382	332
Inter-residue	1,491	1,437	1,480	1,479
Sequential ($\|i-j\|=1$)	446	456	440	464
Medium-range ($\|i-j\|<4$)	246	256	239	271
Long-range ($\|i-j\|>5$)	724	646	726	671
Intermolecular	75	79	75	73
Hydrogen bonds	3		3	
Total dihedral angle restraints	85	97	85	97
ϕ	42	47	42	47
ψ	43	50	43	50
Structure statistics				
Violations (mean ± s.d.)				
Distance restraints (Å)	0.010 ± 0.00	0.009 ± 0.00	0.010 ± 0.00	0.009 ± 0.00
Dihedral angle restraints (°)	0.545 ± 0.05	0.542 ± 0.07	0.754 ± 0.06	0.521 ± 0.06
Max. dihedral angle violation (°)	2.68	3.05	4.73	3.05
Max. distance restraint violation (Å)	0.27	0.39	0.18	0.18
Deviations from idealized geometry				
Bond lengths (Å)	0.013 ± 0.00	0.012 ± 0.00	0.013 ± 0.00	0.013 ± 0.00
Bond angles (°)	1.206 ± 0.03	1.058 ± 0.03	1.156 ± 0.02	1.023 ± 0.03
Impropers (°)	1.204 ± 0.06	1.280 ± 0.05	1.211 ± 0.05	1.197 ± 0.08
Average pairwise r.m.s. deviation (Å)[a]				
Heavy	0.67 ± 0.10	0.67 ± 0.10	0.65 ± 0.09	0.67 ± 0.09
Backbone	0.18 ± 0.04	0.24 ± 0.05	0.19 ± 0.05	0.23 ± 0.05

[a]Pairwise r.m.s. deviation was calculated among 20 refined structures for residues 92–141 of SMN and 73–122 of SPF30.

marks distributed over a range of ten residues (**Fig. 1b**), was studied. Both SMN and SPF30 Tudor domains bound to this ligand with 50× the strength of the shorter peptides containing only one or two sDMA modifications (Sm B/B′, Sm D1), with affinities in the low micromolar range (**Fig. 1c**,**d** and **Supplementary Table 1a**). These affinities were comparable to the ones reported for other Tudor domains (tandem, hybrid or extended forms) that bind methylation marks (lysine or arginine) but require contacts with proximal residues for efficient recognition[22–26]. A series of NMR titrations using the SMN Tudor domain (residues 84–147) and a longer construct (residues 3–160) demonstrated that neither the unstructured region around the SMN Tudor fold nor the residues flanking the methylation marks contributed to binding (**Supplementary Fig. 1** and ref. 16). Analysis of the ITC data indicated that binding enhancement to the long peptide was of entropic nature (**Supplementary Fig. 2**). As the aromatic cage of the Tudor domain can only accommodate a single DMA residue (see below), the strongly increased binding likely reflected external cooperativity effects induced by the increased local concentration of multiple sDMAs presented in the long peptide (**Fig. 1c**,**d**). Consistently, known SMN-Tudor ligands comprise several sDMA modifications in arginine- and glycine-rich peptide motifs[4], implying that this is required for high (micromolar K_d) binding affinity.

Tudor NMR structures in complex with sDMA and aDMA

In NMR titrations, the SMN and SPF30 Tudor domains bound sDMA and aDMA in the fast-exchange regime on the NMR chemical shift time scale (**Fig. 1e** and **Supplementary Fig. 3**), consistent with the binding affinities measured by ITC (**Supplementary Table 1a**). Based on a large number of total and intermolecular distance restraints for the SMN–sDMA, SPF30–sDMA, SMN–aDMA and SPF30–aDMA complexes (**Fig. 1f**, **Table 1** and **Supplementary Fig. 4**), we calculated high-quality structures (**Fig. 2**, **Table 1** and **Supplementary Fig. 5**). The characteristic strongly curved β-sheet fold of the Tudor domain did not show any appreciable structural perturbation upon binding to sDMA and aDMA (**Supplementary Fig. 5**). In either complex, the planar guanidino group of the dimethylated arginine was sandwiched between the aromatic side chains of a tryptophan and a tyrosine residue, whereas another tyrosine and a fourth aromatic residue (SMN Tyr127 and SPF30 Phe108) were positioned orthogonally to this arrangement constituting the aromatic cage (**Figs. 1a** and **2b**). An asparagine residue provided additional van der Waals contacts to the ligand and completed the binding pocket (**Fig. 2b**). Both sDMA and aDMA inserted into the cavity with their methyl groups facing the aromatic side chain (tyrosine or phenylalanine) located at the far end, as confirmed by the NOE contacts (**Supplementary Fig. 6**). Arginine dimethylation imparted hydrophobicity and bulkiness to the residue, but it did not neutralize the cationic charge (**Fig. 2a** and **Supplementary Table 2**). The increased bulk and hydrophobicity facilitated nonelectrostatic contacts with the cage and, most importantly, placed the cationic carbon in a favorable position for cation-π interactions (**Fig. 2c**). For sDMA, the cation was stacked parallel at van der Waals distance from the tryptophan and tyrosine ring centroids (face of the aromat with face of the guanidino group), maximizing cation-π stabilization. In addition, it occupied the focal point of the π-electron system of the two other aromatic side chains that line the cavity (**Fig. 2b**,**c**), thereby contributing additional cation-π interactions in a T-shaped geometry (face of the aromat with edge of the guanidino group)[27,28]. Such a coordination scheme efficiently utilized the polar interactions in an otherwise hydrophobic environment (**Fig. 2b**,**c**). The aDMA cation, on the other hand, was not perfectly perpendicular to the aromatic rings in the stacking geometry, thus reducing the electrostatic stabilization compared with sDMA, an observation consistent with the lower binding affinity. The T-shaped interactions with the other two aromatic side chains became more relevant due to shorter distances. As a common principle of DMA recognition, the steric bulk provided by the addition of two methyl groups to the arginine side chain guides the position of the cation relative to the rigid aromatic cage. This mode of recognition achieves a specific readout of the dimethylation marks and discriminates against nonmethylated and monomethylated arginines.

Contributions of aromatic cage residues to DMA specificity

To assess the contributions of the aromatic cage residues toward methylarginine binding, we individually replaced these in SMN Tudor with amino acids that are frequently observed in the sequences of human TDRD and the *Drosophila* TUDOR proteins (**Fig. 3a** and **Supplementary Fig. 7**). Substitution of any of the four aromatic residues by polar or aliphatic side chains abrogated binding. N132D had mild effects, and

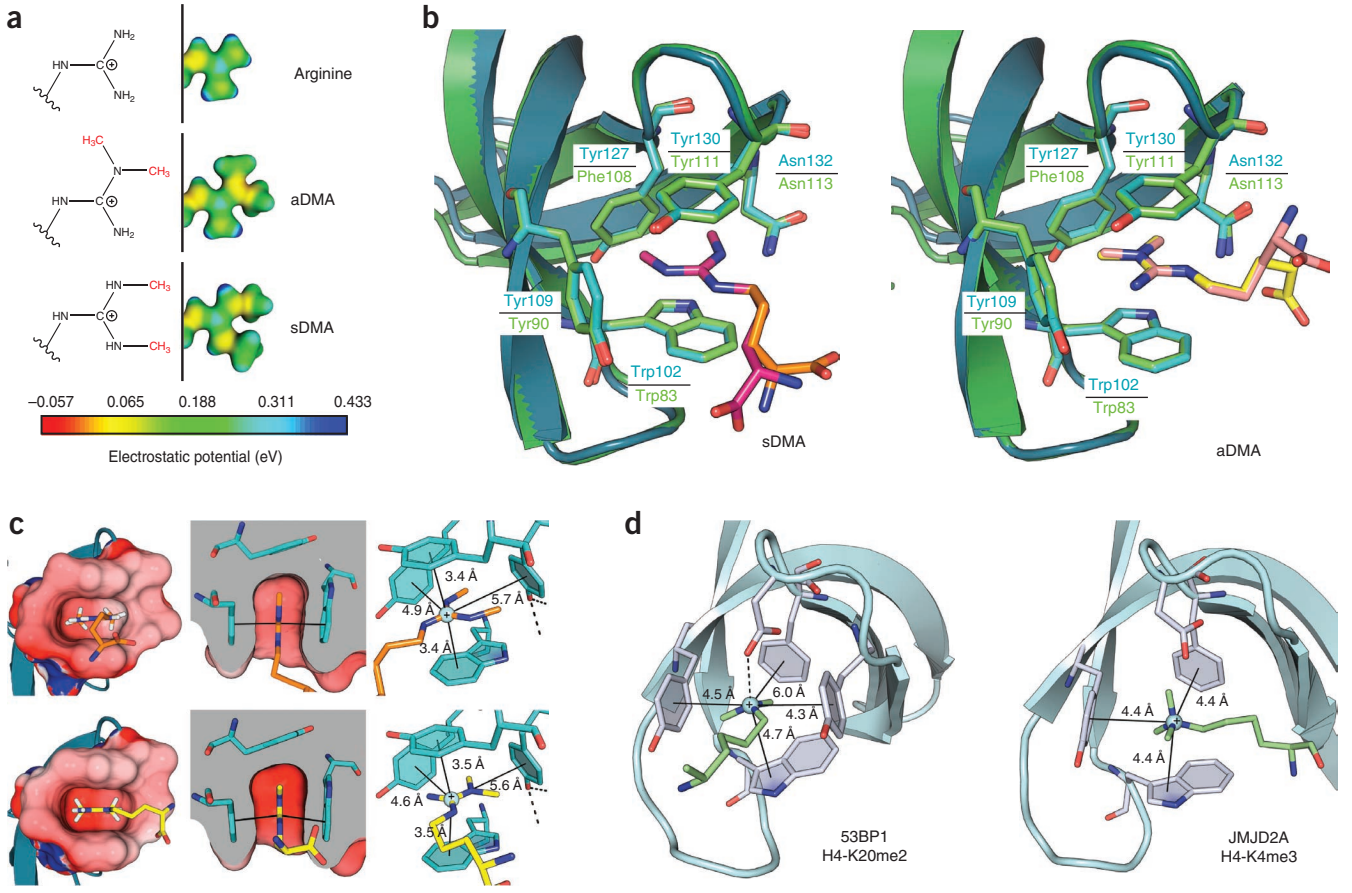

Figure 2 Recognition of sDMA and aDMA by SMN and SPF30 Tudor domains. (**a**) Molecular electrostatic potentials for arginine, aDMA and sDMA guanidino moieties (for details see Online Methods). (**b**) Overlay of SMN (blue) and SPF30 (green) structures in complex with sDMA (orange-purple with SMN-SPF30) and aDMA (yellow-pink with SMN-SPF30). The models represent the pairs with the lowest pairwise coordinate r.m.s. deviation (range 0.15–1.5 Å) from a (20 × 20) matrix of pairwise superpositions of the lowest-energy structures. (**c**) Aromatic cage electrostatics, stacked geometry and cation-π interactions for sDMA and aDMA recognition by SMN are shown at top and bottom, respectively. (**d**) Dimethyl- and trimethyllysine recognition by the Tudor domains of 53BP1 (PDB: 2IG0) and JMJD2A (PDB: 2GFA), respectively. For clarity only the Tudor domain that forms the aromatic cage and the methyllysine ligands are shown.

N132S reduced binding by 90% (**Supplementary Table 1b**). These data are consistent with the roles of interactions described above.

We examined binding in the TDRD2 protein (residues 327–420), which represents the trypophan-to-leucine substitution in our panel of aromatic cage mutations and has been structurally characterized[29]. Consistent with our structural analysis, no sDMA binding was observed for the TDRD2 Tudor domain by NMR or ITC (**Supplementary Fig. 8**), supporting the importance of this residue for DMA binding. Based on these observations, it can be predicted that about half of the Tudor domains in the TDRD family and the TUDOR *Drosophila* protein may not support efficient recognition of the dimethylarginine modification (**Fig. 3a** and **Supplementary Fig. 7**). Thus, it is likely that only a subset of Tudor domains in multi-TUDOR proteins mediates sDMA binding, whereas other Tudor domains may be involved in alternative interactions. Note that some members of the PHD finger and chromodomain families have also been shown to recognize nonmethylated ligands[26].

Design of Tudor-DMA binding specificity

A key difference between the aromatic cages of SMN and SPF30 is the aromatic residue at the backside of the cavity (SMN Tyr127 and SPF30 Phe108; **Figs. 1a** and **2b**). Tyrosine and phenylalanine are equally effective in cation-π binding, unless the hydroxyl group of the tyrosine is hydrogen bonded, in which case its cation-π binding strength increases substantially[27,28]. The structures of SMN Tudor bound to sDMA and aDMA revealed that the hydroxyl of the aromatic cage residue Tyr127 was engaged in a hydrogen bond triangle with the side chains of Glu134 and Gln136 (**Fig. 3b** and **Supplementary Fig. 9**), which is also present in the apo SMN Tudor domain[15,16]. To probe the role of this hydrogen bond for methylarginine binding, we substituted the corresponding residues and measured the binding affinities for the DMA ligands (**Supplementary Fig. 2**). Mutations that abolished the hydrogen bond triangle (SMN Y127F and E134S Q136V) attenuated binding to sDMA and aDMA to a similar extent (**Fig. 1d** and **Supplementary Table 1a**). On the other hand, introduction of a tyrosine at the corresponding position of SPF30 (F108Y) strengthened the interaction for both ligands. SPF30 already contains a glutamate (Glu113) at the position corresponding to Glu134 of SMN that can form a hydrogen bond to the tyrosine hydroxyl. Notably, the SPF30 F108Y T115Q double mutant strengthened the interaction to be comparable with that of SMN Tudor (**Fig. 1d** and **Supplementary Table 1a**), consistent with a fully completed hydrogen bond triangle (**Supplementary Fig. 9**).

SMA-linked point mutations in the SMN Tudor domain

Spinal muscular atrophy (SMA) is a neurodegenerative disease that results from mutations in the *SMN1* gene. Six pathogenic missense

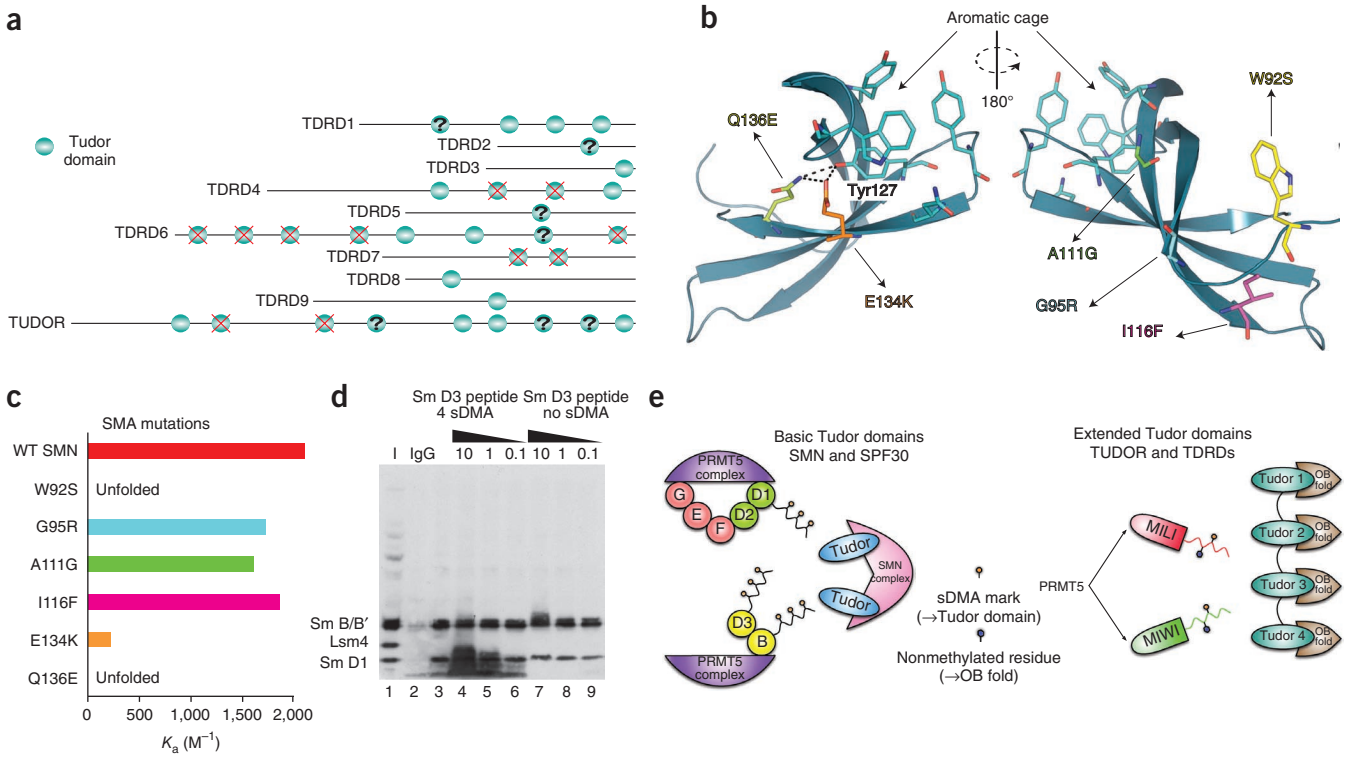

Figure 3 Implications of sDMA recognition in snRNP and piRNP assembly pathways. (**a**) Tudor domains in the sequences of mammalian TDRD proteins and TUDOR *Drosophila*. Question marks refer to Tudors with single substitutions, and red crosses refer to Tudors with two to four substitutions of aromatic cage residues. The latter are not expected to bind DMA. (**b**) Ribbon representation of SMN Tudor highlighting the side chains of SMA-related mutant proteins. The hydrogen bond triangle involving the side chains of Glu134 and Gln136 and the Tyr127 hydroxyl is indicated by a dashed line. (**c**) Association constants of wild-type and mutant SMN Tudors for sDMA binding (**Supplementary Table 1b**). (**d**) Total extracts from HeLa cells stably expressing SMN-GFP (green fluorescent protein) from an integrated bacterial artificial chromosome were analyzed by immunoprecipitation with anti-GFP, followed by western blotting with Y12 antibody that specifically recognizes sDMA marks. Lanes: (1) input, 5% of the total cell lysate; (2) proteins co-immunoprecipitated with nonimmune goat IgG; (3) proteins co-immunoprecipitated with SMN using anti-GFP and (4–9) incubated with the indicated concentrations (mM) of each peptide for competing the Tudor-sDMA interactions in the SMN complex. Note that the detection of additional low-molecular-weight bands in lanes 4–6 is due to the presence of the sDMA peptide, which reacts with Y12. (**e**) Possible roles for basic and extended Tudor domains in snRNP and piRNP pathways, respectively, as assembly platforms for ligands harboring sDMA modifications.

mutations in humans occur in the SMN Tudor domain[30–32]. We examined how these mutations affected the Tudor fold and the binding capacity of the protein to sDMA *in vitro* (**Fig. 3c** and **Supplementary Table 1b**). Two of the mutants (W92S and Q136E) were unfolded, as judged by the fingerprint NMR spectra (**Supplementary Fig. 10**) and did not bind sDMA. Notably, the E134K mutation reduced binding affinity for sDMA by an order of magnitude. Glu134, together with Gln136, participates in the hydrogen bond triangle with the aromatic cage Tyr127 hydroxyl (**Fig. 3b**). The glutamate to lysine substitution affected this hydrogen network and strongly reduced the sDMA interaction. Other SMA-linked single point mutations (G95R, A111G and I116F) perturbed the Tudor domain fold (**Supplementary Fig. 10**), but sDMA binding was similar to the wild-type protein (**Fig. 3c** and **Supplementary Table 1b**). These mutations may indirectly affect the protein stability and/or reduce the cellular levels of functional SMN protein in these patients. SMA mutants involving residues Glu134 (and presumably Trp92 and Gln136) strongly reduced sDMA binding by SMN Tudor, providing a possible link to the SMA phenotypes.

Role of SMN-sDMA recognition in snRNP assembly

Stepwise U snRNP assembly *in vitro* involves the association of Sm D3–B, D1–D2 and E–F–G hetero-oligomers onto the snRNA[33]. *In vivo*, U snRNP biogenesis depends on the PRMT5 and the SMN complexes[13,14,34]. First, the PRMT5 complex pre-organizes the Sm proteins and symmetrically dimethylates the arginine- and glycine-rich tails of Sm B/B′, D1 and D3. Subsequently, the SMN complex, which contains at least two copies of the SMN protein and several Gemin proteins, receives the Sm hetero-oligomers and promotes the formation of the heptameric Sm core of the U snRNPs. An array of Tudor domains thus presented by the SMN complex may serve as a platform for the recruitment of Sm subcomplexes by specific recognition of sDMA marks in the Sm tails (**Fig. 3d**,**e**). It has previously been shown that only sDMA-modified, but not unmodified, peptides representing the C-terminal arginine- and glycine-rich tails of Sm D1 and Sm D3 are able to pull down SMN from HeLa cell extracts[5]. Our co-immunoprecipitation experiments using overexpressed (**Fig. 3d**) or endogenous SMN protein (**Supplementary Fig. 11**) demonstrated that the Sm proteins present in the SMN complex were symmetrically dimethylated, supporting the role of the sDMA modification for the SMN interaction. Note that competition with the Sm D3 peptide, either unmodified or containing four sDMA residues, was not able to dissociate the methylated SMN partners from the SMN complex, indicating that the SMN complex is stabilized by various additional interactions. Indeed, a recently reported crystal structure has demonstrated that Gemin2 makes extensive contacts with an N-terminal helix of SMN and the pentamer of Sm proteins

(D1–D2–F–E–G)[35]. Although the sDMA modification of Sm proteins appears nonessential *in vitro*[34], the requirement for the PRMT5 complex[14,34,36] and the methylation status of the Sm proteins *in vivo* (**Fig. 3d** and **Supplementary Fig. 11**) argue that the readout of the sDMA modification by the SMN Tudor domain (**Fig. 2**) may regulate the kinetics and fidelity during early steps of ribonucleoprotein assembly processes *in vivo*.

DISCUSSION

Recently, crystal structures of extended Tudor domains (TUDOR *Drosophila* and SND1 proteins) in complex with symmetrically dimethylated peptides have been reported[22,23]. In these structures the basic Tudor fold is flanked by additional secondary structure elements. There are notable differences in the recognition of sDMA ligands by the extended Tudor domains compared to the prototypic Tudor domains of SMN and SPF30.

First, efficient binding by the extended Tudor domains requires additional contacts between the OB-fold extension and neighboring, non-methylated amino acids flanking the sDMA within the bound peptide[22,23]. This is in stark contrast to the SMN and SPF30 Tudor domains, where the basic Tudor fold alone is sufficient to recognize exclusively the sDMA marks in cognate peptides (**Supplementary Fig. 1**). Even though SMN and SPF30 show only the basic Tudor fold, they achieve low-micromolar affinities by external binding cooperativity, which results from the presence of multiple sDMA residues within a short stretch of ligand residues as presented by the Sm proteins (**Fig. 1** and **Supplementary Fig. 2**). This external binding cooperativity may compensate for the lack of additional interactions with residues adjacent to the sDMA mark seen in the complex structures of extended Tudor domains.

A second important difference between extended and basic Tudor domains concerns the conformation of the bound sDMA and the rotameric state of the asparagine residue that completes the binding pocket (**Supplementary Fig. 12**). In both reported extended Tudor domains[22,23], the asparagine side chain adopts the *gauche*(+) χ_1 rotameric state (**Supplementary Fig. 12**). In this arrangement the cavity accommodates the *anti-syn* conformation of sDMA, albeit with lower intrinsic affinity. In the basic Tudor domains of SMN and SPF30 presented here, the *gauche*(−) χ_1 rotameric state of the asparagine, which has been measured experimentally for the bound forms (**Supplementary Fig. 13**), enables binding to the *anti-anti* conformation of sDMA with higher intrinsic affinity (**Fig. 2**). This may additionally correlate with the tryptophan residue at the bottom of the cavity that is a distinct feature of SMN and SPF30 Tudor domains (**Supplementary Figs. 7** and **12**). Compared with the phenylalanine (or tyrosine in other instances) that occupies the equivalent position in the recently reported structures of extended Tudor domains[22,23], the tryptophan indole ring enhances the interaction with the methylarginine for two reasons: first, the six-membered ring of the tryptophan side chain mediates stronger π-cation interactions[27,28], and second, the connected six-and five-membered aromatic rings of the indole moiety provide a larger surface able to accommodate both methyl groups of the sDMA (if presented in an *anti-anti* conformation), thereby substantially increasing the van der Waals contacts[27,28].

Note that the *anti-syn* conformation of the sDMA is sterically incompatible with the asparagine *gauche*(−) χ_1 rotamer found in the basic Tudor domains and, vice versa, the asparagine *gauche*(+) χ_1 rotamer observed in the extended Tudor domains would exclude the *anti-anti* conformation of sDMA from the aromatic cage. Density function theory (DFT) calculations indicate that the ground-state energies for the two (free) sDMA conformations differ by 2.8 kcal mol^{-1}, with that of the *anti-syn* conformation being the most favorable. However, the activation energy for rotation about the C^{ζ}-N^{η} bond is 14 kcal mol^{-1} based on the DFT calculations (**Supplementary Fig. 13**). Thus, at room temperature[37], rotations about the C^{ζ}-N^{η} bond of the dimethylated guanidino group will allow for an (*anti-anti*) binding conformation that is distinct from the lowest-energy (*anti-syn*) conformation of the dimethylated arginine side chain when free in solution.

An important role for Tudor-sDMA interactions in RNP biogenesis has been demonstrated for the assembly of piRNA pathway components during germline development. The recognition of sDMA modifications in N-terminal regions of PIWI proteins (such as MILI and MIWI) by TDRD proteins is an interaction in the germline conserved by evolution[2,19–21]. However, TDRD proteins have acquired extensions in the Tudor fold accompanied by new intrinsic features in the recognition of methylarginines (**Supplementary Fig. 12**). The linkage of the Tudor domain to the OB fold has expanded the available interaction surface around the aromatic cage, offering the possibility of additional contacts with residues flanking the sDMA (**Fig. 3e**). Indeed, the OB fold is indispensable for the recognition of methylarginine peptides by the extended Tudor domains[22,23]. It not only assists the aromatic cage in the readout of the sDMA in the *anti-syn* conformation but also provides another level of selectivity for the peptide sequence flanking the modification. Given that TUDOR *Drosophila* and some TDRDs contain functional repeats of the extended Tudor domain, they may be capable of sorting out sDMA-containing effector proteins at specific points of the piRNA pathway (**Fig. 3e**).

Our structural findings show that the same general principles allow the basic Tudor domains of SMN and SPF30 to bind aDMA with subtle differences reflecting the slightly lower affinity for aDMA (**Supplementary Fig. 12**). These observations are consistent with a role for aDMA in Tudor-mediated interactions. For example, SMN interaction with the splicing factor CA150 is subject to regulation by aDMA modification[38]. In addition, the Tudor domain of TDRD3 selectively recognizes a peptide from the C-terminal domain of RNA polymerase II that is asymmetrically modified[39]. Finally, the *Drosophila* Vasa protein and its mouse homolog contain both sDMA and aDMA, but the sDMA modifications were found to be dispensable for the association with their Tudor binding partners[40]. Thus, it is possible that Vasa interaction with Tudor-containing partners relies on aDMA modification as well. The comparable binding affinities and similar modes of recognition of the SMN and SPF30 Tudor domain for sDMA and aDMA (**Fig. 2**) are consistent with a biological role of Tudor-aDMA interactions. Our structural analysis of the Tudor-aDMA interaction may serve as a model for understanding the interactions of Tudor-domain proteins with aDMA modifications in cognate ligands.

The tandem Tudor domains of 53BP1 (ref. 24) and the hybrid Tudor domains of JMJD2A (ref. 25) also recognize methyllysine residues through aromatic caging. However, the electrostatic interactions involving the positively charged methyllysine are weaker compared with dimethylarginine because of the longer distances to the rings of the aromatic cage (**Fig. 2d**). This is consistent with the observation that none of the histone methyllysine readers binds appreciably to the isolated methyllysine amino acid (K_d greater than millimolar)[26]. The same holds true for the extended Tudor domains of TUDOR *Drosophila* and SND1 proteins that bind sDMA in the *anti-syn* conformation[22,23]. The crystal structures indicate that the cation-π geometry is not optimal and the interaction between the cage and the sDMA headgroup may thus require further stabilization by a hydrogen bond (**Supplementary Fig. 12**). Considering the low affinities for individual methyllysine and methylarginine amino acids, different mechanisms have been devised to interpret these modifications. The Tudor domains that are fused to other domains—for example, to another Tudor (53BP1 and JMJD2A)

or to the OB fold (TUDOR *Drosophila* and SND1)—recognize methyllysines or methylarginines as part of a longer peptide sequence, where both the modification and the flanking residues contribute to the binding affinity and specificity[22–26]. In contrast, the basic Tudor domains of SMN and SPF30 achieve increased affinities for arginine- and glycine-rich peptides by external binding cooperativity linked to the presence of multiple DMA modifications (**Fig. 1c** and **Supplementary Table 1a**) and may represent the prototypic mode of Tudor-DMA recognition that acquired new intrinsic properties in the extended architecture of the Tudor fold.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/nsmb/.

Accession codes. The atomic coordinates for the NMR ensembles have been deposited in the Protein Data Bank under accession numbers 4A4E (SMN–sDMA), 4A4F (SPF30–sDMA), 4A4G (SMN–aDMA) and 4A4H (SPF30–aDMA). The chemical shift assignments have been deposited in the Biological Magnetic Resonance Data Bank under accession numbers 18005 (SMN–sDMA), 18006 (SPF30–sDMA), 18007 (SMN–aDMA) and 18008 (SPF30–aDMA).

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

The authors are grateful to G. Demiraslan (Helmholtz Zentrum München) for protein production. We thank I. Poser and T. Hyman (Max Planck Institute of Molecular Cell Biology and Genetics) for the gift of the SMN-GFP stable cell line, J. Brennecke for helpful discussion and the Bavarian NMR Centre for NMR time. K.T. acknowledges support by the Alexander von Humboldt foundation; T.M. acknowledges support by a European Molecular Biology Organization Long Term Fellowship and a Schrödinger Fellowship from the Austrian Science Fund. This work was supported by the Deutsche Forschungsgemeinschaft (M.S. and K.M.N.).

AUTHOR CONTRIBUTIONS

K.T. carried out molecular biology experiments, protein purification, NMR analysis and structure calculations. T.M. did quantum chemical calculations and contributed to NMR experiments. D.F. analyzed the ITC experiments. K.M.N. and U.F. designed, and M.M. and C.E. conducted, immunoprecipitation experiments. K.T. and M.S. conceived and designed the project and wrote the paper. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Bedford, M.T. & Clarke, S.G. Protein arginine methylation in mammals: who, what, and why. *Mol. Cell* **33**, 1–13 (2009).
2. Siomi, M.C., Mannen, T. & Siomi, H. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev.* **24**, 636–646 (2010).
3. Boisvert, F.M., Cote, J., Boulanger, M.C. & Richard, S. A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2**, 1319–1330 (2003).
4. Brahms, H., Meheus, L., de Brabandere, V., Fischer, U. & Luhrmann, R. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B′ and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**, 1531–1542 (2001).
5. Friesen, W.J., Massenet, S., Paushkin, S., Wyce, A. & Dreyfuss, G. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol. Cell* **7**, 1111–1117 (2001).
6. Côté, J. & Richard, S. Tudor domains bind symmetrical dimethylated arginines. *J. Biol. Chem.* **280**, 28476–28483 (2005).
7. Friberg, A., Corsini, L., Mourao, A. & Sattler, M. Structure and ligand binding of the extended Tudor domain of *D. melanogaster* Tudor-SN. *J. Mol. Biol.* **387**, 921–934 (2009).
8. Bühler, D., Raker, V., Lührmann, R. & Fischer, U. Essential role for the Tudor domain of SMN in spliceosomal U snRNP assembly: implications for spinal muscular atrophy. *Hum. Mol. Genet.* **8**, 2351–2357 (1999).
9. Pellizzoni, L., Kataoka, N., Charroux, B. & Dreyfuss, G. A novel function for SMN, the spinal muscular atrophy disease gene product, in pre-mRNA splicing. *Cell* **95**, 615–624 (1998).
10. Pellizzoni, L., Yong, J. & Dreyfuss, G. Essential role for the SMN complex in the specificity of snRNP assembly. *Science* **298**, 1775–1779 (2002).
11. Meister, G. *et al.* SMNrp is an essential pre-mRNA splicing factor required for the formation of the mature spliceosome. *EMBO J.* **20**, 2304–2314 (2001).
12. Rappsilber, J., Ajuh, P., Lamond, A.I. & Mann, M. SPF30 is an essential human splicing factor required for assembly of the U4/U5/U6 tri-small nuclear ribonucleoprotein into the spliceosome. *J. Biol. Chem.* **276**, 31142–31150 (2001).
13. Paushkin, S., Gubitz, A.K., Massenet, S. & Dreyfuss, G. The SMN complex, an assemblyosome of ribonucleoproteins. *Curr. Opin. Cell Biol.* **14**, 305–312 (2002).
14. Meister, G. & Fischer, U. Assisted RNP assembly: SMN and PRMT5 complexes cooperate in the formation of spliceosomal UsnRNPs. *EMBO J.* **21**, 5853–5863 (2002).
15. Selenko, P. *et al.* SMN Tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.* **8**, 27–31 (2001).
16. Sprangers, R., Groves, M.R., Sinning, I. & Sattler, M. High-resolution X-ray and NMR structures of the SMN Tudor domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J. Mol. Biol.* **327**, 507–520 (2003).
17. Hebert, M.D., Shpargel, K.B., Ospina, J.K., Tucker, K.E. & Matera, A.G. Coilin methylation regulates nuclear body formation. *Dev. Cell* **3**, 329–337 (2002).
18. Renvoisé, B. *et al.* Distinct domains of the spinal muscular atrophy protein SMN are required for targeting to Cajal bodies in mammalian cells. *J. Cell Sci.* **119**, 680–692 (2006).
19. Vagin, V.V. *et al.* Proteomic analysis of murine Piwi proteins reveals a role for arginine methylation in specifying interaction with Tudor family members. *Genes Dev.* **23**, 1749–1762 (2009).
20. Kirino, Y. *et al.* Arginine methylation of Piwi proteins catalysed by dPRMT5 is required for Ago3 and Aub stability. *Nat. Cell Biol.* **11**, 652–658 (2009).
21. Nishida, K.M. *et al.* Functional involvement of Tudor and dPRMT5 in the piRNA processing pathway in *Drosophila* germlines. *EMBO J.* **28**, 3820–3831 (2009).
22. Liu, H. *et al.* Structural basis for methylarginine-dependent recognition of Aubergine by Tudor. *Genes Dev.* **24**, 1876–1881 (2010).
23. Liu, K. *et al.* Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain. *Proc. Natl. Acad. Sci. USA* **107**, 18398–18403 (2010).
24. Botuyan, M.V. *et al.* Structural basis for the methylation state–specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
25. Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. & Xu, R.M. Recognition of histone H3 lysine-4 methylation by the double Tudor domain of JMJD2A. *Science* **312**, 748–751 (2006).
26. Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D. & Patel, D.J. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040 (2007).
27. Mecozzi, S., West, A.P. Jr. & Dougherty, D.A. Cation-π interactions in aromatics of biological and medicinal interest: electrostatic potential surfaces as a useful qualitative guide. *Proc. Natl. Acad. Sci. USA* **93**, 10566–10571 (1996).
28. Gallivan, J.P. & Dougherty, D.A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. USA* **96**, 9459–9464 (1999).
29. Chen, C. *et al.* Mouse Piwi interactome identifies binding mechanism of Tdrkh Tudor domain to arginine methylated Miwi. *Proc. Natl. Acad. Sci. USA* **106**, 20336–20341 (2009).
30. Chari, A., Paknia, E. & Fischer, U. The role of RNP biogenesis in spinal muscular atrophy. *Curr. Opin. Cell Biol.* **21**, 387–393 (2009).
31. Pellizzoni, L., Charroux, B. & Dreyfuss, G. SMN mutants of spinal muscular atrophy patients are defective in binding to snRNP proteins. *Proc. Natl. Acad. Sci. USA* **96**, 11167–11172 (1999).
32. Eggert, C., Chari, A., Laggerbauer, B. & Fischer, U. Spinal muscular atrophy: the RNP connection. *Trends Mol. Med.* **12**, 113–121 (2006).
33. Raker, V.A., Plessel, G. & Lührmann, R. The snRNP core assembly pathway: identification of stable core protein heteromeric complexes and an snRNP subcore particle *in vitro*. *EMBO J.* **15**, 2256–2269 (1996).
34. Chari, A. *et al.* An assembly chaperone collaborates with the SMN complex to generate spliceosomal SnRNPs. *Cell* **135**, 497–509 (2008).
35. Zhang, R. *et al.* Structure of a key intermediate of the SMN complex reveals gemin2's crucial function in snRNP assembly. *Cell* **146**, 384–395 (2011).
36. Friesen, W.J. *et al.* The methylosome, a 20S complex containing JBP1 and pICln, produces dimethylarginine-modified Sm proteins. *Mol. Cell. Biol.* **21**, 8289–8300 (2001).
37. Kessler, H. Detection of hindered rotation and inversion by NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* **9**, 219–235 (1970).
38. Cheng, D., Cote, J., Shaaban, S. & Bedford, M.T. The arginine methyltransferase CARM1 regulates the coupling of transcription and mRNA processing. *Mol. Cell* **25**, 71–83 (2007).
39. Sims, R.J. *et al.* The C-terminal domain of RNA polymerase IIis modified by site-specific methylation. *Science* **332**, 99–103 (2011).
40. Kirino, Y. *et al.* Arginine methylation of Vasa protein is conserved across phyla. *J. Biol. Chem.* **285**, 8148–8154 (2010).

ONLINE METHODS

Sample preparation. The plasmids expressing $SMN_{84–147}$ and $SPF30_{65–128}$ with N-terminal fused histidine tags were transformed into *Escherichia coli* BL21 (DE3) cells. All mutants were constructed using site-directed mutagenesis, and the clones were confirmed by sequencing. Cells were grown at 20 °C in the presence of kanamycin. ^{13}C- and ^{15}N-labeled samples were prepared by growing cells in minimal medium supplemented with $^{15}NH_4Cl$ (1g liter^{-1}) and $^{13}C_6$ glucose (2g liter^{-1}) as the sole nitrogen and carbon sources, respectively. Protein synthesis was induced by the addition of 0.5 mM of isopropyl-1-thio-D-galactopyranoside (IPTG) at OD_{600} ~0.8, and cells were collected after 20–24 h. Cells were resuspended in lysis buffer containing 50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, 1 mM tris-(2-carboxyethyl)-phosphine (TCEP), 0.1% (v/v) Triton X-100 , lysed by sonication and centrifuged at 15,000*g*. Two purification steps were used for all protein samples. The first one involved Ni-NTA gravity-flow resin (Qiagen) followed by removal of the tag, and the second one used a Superdex-75 size exclusion column (GE Healthcare). MMA, aDMA and sDMA were purchased from Enzo Life Sciences and sDMA-containing peptides from Peptide Specialty Laboratories.

NMR spectroscopy. The samples used for structure determination contained 4 mM protein (SMN or SPF30) and 18 mM of the respective ligand (aDMA or sDMA) in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl with 7% (v/v) ^2H_2O added for the lock. Spectra were recorded at 298 K using AVIII600, AVIII750 or AV900 Bruker NMR spectrometers equipped with cryogenic triple resonance gradient probes (except AVIII750). Spectra were processed using NMRPipe[41]. Protein backbone and side chain assignments were obtained from standard triple resonance experiments[42]. For every complex structure six NOESY experiments were acquired. NOE distance restraints were obtained from ^{15}N- and ^{13}C-edited 3D NOESY spectra (mixing time 70 ms). Intermolecular NOE restraints were derived from 3D ^{15}N,^{13}C-filtered, ^{15}N- and ^{13}C-edited NOESY spectra[43] (mixing time 70 ms). Stereospecific assignments of the methyl groups of valine and leucine were obtained using 10% ^{13}C-labeled protein samples[44]. Quantitative J correlation spectroscopy measurements to determine the χ_1 rotamer of the asparagine residue have been carried out as described[45,46].

To confirm the bound conformation of the ligands based on NOEs derived from samples with the ligands in excess, we measured intermolecular NOEs using samples with the protein in excess (**Fig. 1f**). In these samples, where the chemical shifts of the ligands represent the genuine bound form, ^{15}N,^{13}C-filtered 2D and 3D ^{15}N- and ^{13}C-edited NOESY spectra were recorded at different temperatures (from 278 to 318 K) and different mixing times (from 10 to 200 ms) to shift the exchange regime in order to obtain a comprehensive qualitative NOE analysis.

Structure determination. Automated NOESY cross-peak assignments and structure calculations with torsion angle dynamics were done by using the program CYANA 3.0 (ref. 47). For every single peak list, the distances between the protons were calibrated using CYANA's iterative calibration algorithm. The algorithm employs a $V = A\,u^{-6}$ relationship between the peak volume, V and the corresponding upper distance bound, u. The constant A is assumed such that the median of all peak volumes in a given NOESY spectrum corresponds to a distance of 4 Å and accordingly calibrates the distances for all the cross peaks (for a range between 2.4 and 6 Å). The set of NOE distance restraints derived from CYANA together with φ and ψ backbone dihedral angle restraints derived from TALOS$^+$ (ref. 48) based on the chemical shifts were used for water refinement[49] using CNS[50]. Ligands have been parametrized by using the Dundee PRODRG2 Server[51], applying the crystal structures from the Heterocompound Information Centre (Uppsala, Sweden)[52]. The compound names are DA2 for aDMA and 2MR for sDMA. All structures were validated by using iCing (http://nmr.cmbi.ru.nl/icing/). Ramachandran plot statistics for residues in most favored regions, additional allowed regions, generously allowed regions, disallowed regions are: SMN–sDMA 89.5%, 10.1%, 0.2%, 0.3%; SPF30–sDMA 96.7%, 2.9%, 0.3%, 0.2%; SMN–aDMA 87.9%, 11.8%, 0.2%, 0.1%; SPF30–aDMA 95.6%, 4.4%, 0%, 0%. Molecular images were generated with PyMol (Schrödinger) and Molekel (http://molekel.cscs.ch/wiki/pmwiki.php).

***Ab initio* ligand calculations.** Arginine, aDMA and sDMA starting geometries were used in the Gaussian 03 program suite (Gaussian), using Becke's three-parameter hybrid Hartree-Fock density functional method[53] in combination with the Lee-Yang-Parr correlation functional (B3LYP)[54] and the 6–31g(d) basis set. All structures were optimized and characterized by frequency calculations as true minima. Atomic charges were computed from the Mulliken population analysis. Molecular electrostatic potentials for the arginine, aDMA and sDMA guanidino moieties in **Figure 2a** are plotted in units of eV on the electron isodensity surface at $\rho = 0.05\ e$ Å^{-3}; "e" is the electron charge.

ITC experiments. Calorimetric titrations were carried out on an iTC200 microcalorimeter (MicroCal) at 25 °C. Protein and ligand samples were dissolved in 20 mM sodium phosphate (pH 6.5), 50 mM NaCl. The 200-µl sample cell was filled with a 0.5 or 1 mM solution of protein, and the 40-µl injection syringe with 20 or 40 mM of the titrating ligand. Each titration consisted of a preliminary 0.2-µl injection followed by 40 subsequent 1-µl injections. The heat of the injections was corrected for the heat of dilution of every ligand into the buffer. At least two replicas were done for each experiment. Binding thermodynamic models[55,56] were tested to interpret the calorimetric data. The fit attempts were accomplished using the nonlinear Levenberg-Marquardt method[57]. The errors of each fitting parameter were calculated with a 95.4% confidence limit by the Monte Carlo simulation method. An eventual error of 10% in the protein and/or ligand concentration was also taken into account. In all cases errors were within 10% of the reported values.

Immunoprecipitation. SMN complexes were immunoprecipitated with anti-GFP antibody from HeLa cells stably expressing SMN-GFP from a bacterial artificial chromosome and washed with NET-2 buffer (50 mM Tris pH 7.5, 150 mM NaCl, 0.05% (v/v) NP-40). Complexes were then incubated twice for 15 min in NET-2 buffer containing Sm D3 peptide (Peptide Specialty Laboratories), with or without 4 sDMA modifications, and washed after each incubation. Remaining complexes were eluted by boiling in Laemmli buffer, followed by western blotting with Y12 antibody.

41. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277–293 (1995).
42. Sattler, M., Schleucher, J.R. & Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog. Nucl. Magn. Reson. Spectrosc.* **34**, 93–158 (1999).
43. Breeze, A.L. Isotope-filtered NMR methods for the study of biomolecular structure and interactions. *Prog. Nucl. Magn. Reson. Spectrosc.* **36**, 323–372 (2000).
44. Neri, D., Szyperski, T., Otting, G., Senn, H. & Wuthrich, K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* **28**, 7510–7516 (1989).
45. Grzesiek, S. & Bax, A. A three-dimensional NMR experiment with improved sensitivity for carbonyl-carbonyl J correlation in proteins. *J. Biomol. NMR* **9**, 207–211 (1997).
46. Hu, J.S. & Bax, A. χ_1 angle information from a simple two-dimensional NMR experiment that identifies trans $^3J_{NC\gamma}$ couplings in isotopically enriched proteins. *J. Biomol. NMR* **9**, 323–328 (1997).
47. Güntert, P. Automated structure determination from NMR spectra. *Eur. Biophys. J.* **38**, 129–143 (2009).
48. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS$^+$: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44**, 213–223 (2009).
49. Linge, J.P., Williams, M.A., Spronk, C.A., Bonvin, A.M. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **50**, 496–506 (2003).
50. Brünger, A.T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**, 905–921 (1998).
51. Schüttelkopf, A.W. & van Aalten, D.M. PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 1355–1363 (2004).
52. Kleywegt, G.J. Crystallographic refinement of ligand complexes. *Acta Crystallogr. D Biol. Crystallogr.* **63**, 94–100 (2007).
53. Becke, A. Density-functional thermochemistry. III. the role of exact exchange. *J. Chem. Phys.* **98**, 5648–5652 (1993).
54. Lee, C., Yang, W. & Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B Condens. Matter.* **37**, 785–789 (1988).
55. Gill, S.J. Thermodynamics of ligand-binding to proteins. *Pure Appl. Chem.* **61**, 1009–1020 (1989).
56. Capaldi, S. *et al.* The X-ray structure of zebrafish (*Danio rerio*) ileal bile acid-binding protein reveals the presence of binding sites on the surface of the protein molecule. *J. Mol. Biol.* **385**, 99–116 (2009).
57. Press, W., Teukolsky, S., Vetterling, W. & Flannery, B. *Numerical Recipes 3rd Edition: The Art of Scientific Computing* (Cambridge University Press, 2007).