Structural basis for recognition of arginine methylated Piwi proteins by the extended Tudor domain

Ke Liu^{a,b,1}, Chen Chen^{c,1}, Yahong Guo^{b,1}, Robert Lam^b, Chuanbing Bian^b, Chao Xu^b, Dorothy Y. Zhao^c, Jing Jin^c, Farrell MacKenzie^b, Tony Pawson^{c,d,2}, and Jinrong Min^{a,b,e,2}

^aHubei Key Laboratory of Genetic Regulation and Integrative Biology, College of Life Science, Huazhong Normal University, Wuhan 430079, People's Republic of China; ^bStructural Genomics Consortium, University of Toronto, 101 College Street, Toronto, ON, M5G 1L7, Canada; ^cSamuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, ON, Canada M5G 1X5; ^dDepartment of Molecular Genetics, University of Toronto, Toronto, ON, Canada M5S 1A8; and ^eDepartment of Physiology, University of Toronto, Toronto, ON, Canada M5S 1A8

Contributed by Tony Pawson, September 2, 2010 (sent for review August 23, 2010)

Arginine methylation modulates diverse cellular processes and represents a molecular signature of germ-line-specific Piwi family proteins. A subset of Tudor domains recognize arginine methylation modifications, but the binding mechanism has been lacking. Here we establish that, like other germ-line Tudor proteins, the ancestral staphylococcal nuclease domain-containing 1 (SND1) polypeptide is expressed and associates with PIWIL1/Miwi in germ cells. We find that human SND1 binds PIWIL1 in an arginine methylation-dependent manner with a preference for symmetrically dimethylated arginine. The entire Tudor domain and a bifurcated SN domain are required for this binding activity, whereas the canonical Tudor domain alone is insufficient for methylarginine ligand binding. Crystal structures show that the intact SND1 extended Tudor domain forms a wide and negatively charged binding groove, which can accommodate distinct symmetrically dimethylated arginine peptides from PIWIL1 in different orientations. This analysis explains how SND1 preferentially recognizes symmetrical dimethylarginine via an aromatic cage and conserved hydrogen bonds, and provides a general paradigm for the binding mechanisms of methylarginine-containing peptides by extended Tudor domains.

Piwi-interacting RNA | Tudor domain-containing proteins

Arginine methylation, the covalent addition of methyl group to arginine residues, is a form of protein posttranslational modification that is increasingly recognized for its importance in mediating protein–protein interactions (1, 2). Three types of arginine methylation—namely, monomethylation, asymmetrical dimethylation, and symmetrical dimethylation—are catalyzed by type I or type II protein arginine methyltransferases (PRMTs). Among these, PRMT5 is considered the major PRMT that induces symmetrical dimethylation of arginine residues on target proteins (2).

Piwi proteins, the germ-line-specific clade of Argonaute family proteins, together with their associated Piwi-interacting RNAs (piRNAs) are evolutionarily conserved in animals and play pivotal roles in transposon silencing during gametogenesis (3, 4). Mammalian Piwi proteins, which include PIWIL1/Miwi, PIWIL2/Mili, PIWIL3 and PIWIL4/Miwi2, contain multiple arginine-glycine (RG) and arginine-alanine (RA) repeats at their N termini, and methylation of these arginine sites is important for recruiting the Tudor domain-containing (TDRD) group of germ-line-enriched Tudor domain proteins such as Tdrd1, Tdrd2/Tdrkh, Tdrd4/RNF17, Tdrd5, Tdrd6, Tdrd7, Tdrd8/Stk31, and Tdrd9 (5–12). Arginine methylation is a conserved posttranslational modification of Piwi family proteins of various species including *Drosophila*, *Xenopus*, and mammals (7, 13). Its importance has been demonstrated by null mutants of *drosophila* PRMT5 (*dart5*, *csul*), in which arginine methylation of Piwi family members is compromised, leading to retrotransposon activation and defects in germ-line development (7, 14, 15).

The Tudor domain, together with Chromo, MBT, PWWP, and Agenet-like domains, comprise the “royal family” of protein domains that engage in protein–protein interactions (16, 17). The Tudor domain is characterized by a β -barrel core and in many cases an aromatic cage suitable for docking methylated lysine or arginine (16, 18). The Tudor domain family is largely divided into two groups: a methyllysine binding group (i.e., JMJD2A, 53BP1) and a methylarginine binding group (SMN, TDRD group [Tdrd1–Tdrd12]) (19). Crystal structures of 53BP1 tandem Tudor domains in complex with histone H4K20me2 and a JMJD2A hybrid Tudor domain with histone H3K4me3 have provided the structural basis for Tudor domain recognition of methylated lysine residues through aromatic cages (20, 21). On the contrary, much less is known about the binding mechanism of methylarginine-specific Tudor domains, although several lines of biochemical evidence suggest they have a preference for binding proteins with symmetrically dimethylated arginine (sDMA) (22). A unique structural feature of Tudor domains from the TDRD group is an additional α -helix and two β -strands N-terminal to the canonical Tudor core domain (19), which may assist this subfamily of Tudor domain in binding to its methylated arginine ligands, and are absent from methyllysine-binding Tudor domains. This extended sequence domain is annotated as a Maternal Tudor domain (23). Among 12 TDRD Tudor proteins, SND1 (TDRD11, staphylococcal nuclease domain-containing 1, Tudor-SN, P100) is the most ancient molecule to possess this N-terminal extension, and we have proposed that this Tudor domain may be a precursor for all germ-line TDRD Tudor domains (19). Therefore, elucidating the binding of the SND1 extended Tudor domain to its ligands is of general importance in understanding the structural determinants underlying specific arginine methylation-dependent interactions and the recognition modes of arginine methylated peptide motifs.

In order to test whether a full extended Tudor domain is required for methylarginine binding, and to establish a prototypic binding mechanism for arginine methylation-dependent Piwi–Tudor interactions, we have investigated the crystal structures of the human SND1 extended Tudor domain in complex with symmetrically dimethylated arginine peptides derived from PIWIL1.

Author contributions: K.L., C.C., Y.G., T.P., and J.M. designed research; K.L., C.C., Y.G., R.L., C.B., C.X., D.Y.Z., and F.M. performed research; K.L., C.C., Y.G., R.L., C.B., C.X., D.Y.Z., J.J., T.P., and J.M. analyzed data; and C.C., T.P., and J.M. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The atomic coordinates, crystallography, and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 3OMC and 3OMG).

¹K.L., C.C., and Y.G. contributed equally to this work.

²To whom correspondence may be addressed. E-mail: pawson@lunenfeld.ca or jr.min@utoronto.ca.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1013106107/-DCSupplemental.

These data delineate the structural basis for preferential binding of symmetrically dimethylated arginine by an extended Tudor domain, which is of general relevance to the family of germ-line mammalian Tudor proteins, and identify distinct mechanisms by which the SND1 extended Tudor domain recognizes arginine methylated peptide motifs.

Results

SND1 Associates with Miwi in Vivo, and Its Extended Tudor Domain Preferentially Binds Piwi Peptides with sDMA. Several TDRD proteins associate with Piwi proteins in mouse testes or in cotransfected cells, and their interactions are most likely arginine methylation-dependent (5). For example, we have established that Tdrkh is a major binding partner of Miwi, and that the Tudor domain is critical in mediating Tdrkh–Miwi interaction (8). Because the crystal structure of the apo-SND1 Tudor domain is highly similar to that of the Miwi-binding germ-line Tudor protein Tdrkh, in terms of overall binding surface and aromatic cage residue position (8, 24), we asked whether SND1 is capable of binding Miwi. Because it is not known whether the SND1 protein is expressed in male germ cells, we performed immunoblotting to examine its tissue distribution in mice. SND1 was highly expressed in pancreas and liver, with moderate to weak levels in other tissues including testis (Fig. S1). We next analyzed SND1 expression dynamics during spermatogenesis (Fig. 1A). SND1 was undetectable or weakly expressed at postnatal week 1, but exhibited elevated protein levels from postnatal week 2 to adult, correlating with the onset of meiosis and Miwi expression (25).

Fig. 1. SND1 associates with Miwi in vivo and its extended Tudor domain preferentially binds Piwi peptides with sDMA. (A) Expression kinetics of SND1 in mouse testes of different developmental ages. ES, mouse embryonic stem cells; DLD1, a human colon carcinoma cell line. (B) Association of SND1 with Miwi in adult testis. Testis lysates was immunoprecipitated using anti-SND1, anti-Miwi, and IgG control antibodies and subject to anti-Miwi antibody immunoblotting. (C) The N terminus sequence of human PIWIL1 or mouse Miwi. The methylation of those arginine residues colored in red is studied in this work. (D) The binding curves of fluorescence polarization measurements of different PIWIL1-R4 peptides to SND1 extended Tudor domain. (E) Binding affinities of SND1 extended Tudor domain to different PIWIL1 peptides.

To examine its potential association with Miwi, we immunoprecipitated SND1 from testis lysates and immunoblotted with Miwi antibody and found that SND1 coprecipitated with Miwi as compared with a negative control IgG immunoprecipitation (Fig. 1B). These results indicate a potential unique involvement of the multifunctional SND1 in the Piwi/piRNA pathway in the germ cells.

SND1 consists of four staphylococcal nuclease-like (SN) domains, and a Tudor domain, which is embedded in a fifth SN domain (Fig. 2A). We refer to the Tudor domain and the fifth SN domain together as an extended Tudor domain in this study. To pursue the endogenous association with Piwi identified by immunoprecipitation in the mouse, and to better understand the binding specificity of SND1, we used a fluorescence polarization binding assay to measure the affinity of the human SND1 extended Tudor domain for an N-terminal peptide from human PIWIL1 protein, which is 100% conserved with mouse Miwi. This peptide contains multiple RA/RG repeats, and was synthesized with a symmetrical dimethylation modification at one of three different arginine sites (R4, R10, or R14) (Fig. 1C–E). The SND1 extended Tudor domain bound all three of these sDMA peptides (Fig. 1D and E). Simultaneous methylation of all three arginine residues did not markedly affect binding. SND1 bound to the R4me2s peptide (symmetrical dimethylation at arginine 4) with a K_D of 10 μM. The binding affinity was about twofold weaker for a peptide that was monomethylated at arginine 4 (R4me1), about fourfold weaker for a peptide with asymmetrical dimethylation at arginine 4 (R4me2a), and about ninefold weaker for an unmodified R4 peptide at a salt concentration of 50 mM NaCl (Fig. 1D and E). Hence, the SND1 extended Tudor domain preferentially recognizes symmetrically arginine-dimethylated PIWIL1/Miwi peptides. At a higher salt concentration of 150 mM NaCl, the binding affinities were reduced roughly fourfold for each peptide (Fig. 1D and E), indicating that electrostatic interactions play a role in the SND1 interaction with PIWIL1 N terminus.

Overall Structures of the SND1 Extended Tudor Domain in Complex with Two PIWIL1 Peptides, R4me2s and R14me2s. To gain insight into the molecular mechanism of selective binding of the SND1 extended Tudor domain to symmetrically arginine-dimethylated proteins, we determined the crystal structures of the SND1 extended Tudor domain in complex with R4me2s and R14me2s peptides (Fig. 2 and Fig. S2). The apostructure of the SND1 extended Tudor domain has been determined before (24) and is almost identical to that in our structures of the domain complexed with peptides. Based on our structures of these complexes, we can divide the SND1 extended Tudor domain into three modules: the canonical Tudor domain colored in blue, the fifth SN-like domain colored in cyan and green, and the linker α-helix colored in yellow. The SN-like domain is split into two segments by the Tudor domain: the N-terminal two β-strands colored in cyan and the C-terminal three α-helices and three β-strands colored in green (Fig. 2). The canonical Tudor domain, the linker α-helix, and the N-terminal two β-strands of the SN domain form the maternal Tudor domain sequence.

In order to test the significance of the N- and C-terminal segments of the SN domain in Piwi protein binding, we made two deletion mutants by separately truncating the N- and C-terminal parts of the SN domain. Binding data obtained from these mutants showed that deleting the N-terminal two β-strands of the SN domain (construct covering residues 700–910) totally disrupted binding, and deleting the C-terminal part of the SN domain (residues 650–800) severely diminished binding (K_D = 300 μM) (Table 1). Therefore, both segments of the SN domain are critical for binding of the extended Tudor domain to its ligands. TDRKH is a germ cell-specific Piwi-binding Tudor protein, which has a 25% sequence identity and 43% sequence similarity with the extended Tudor domain of SND1, and is therefore predicted to possess an extended Tudor domain (Fig. S2C). To test

Fig. 2. Crystal structures of human SND1 extended Tudor domain in complex with PIWI1 peptides. (A) Domain structure of SND1. SN, staphylococcal nuclease-like domain. (B) Complex structure of SND1 and R4me2s peptide: The SND1 extended Tudor domain is shown in cartoon, and the R4me2s peptide and residues interacting with the R4me2s peptide from SND are shown in stick models. Hydrogen bonds are shown in red dotted lines. (C) Complex structure of SND1 and R14me2s peptide: The SND1 extended Tudor domain is shown in cartoon, and the R14me2s peptide and residues interacting with the R14me2s peptide from SND are shown in stick models. Hydrogen bonds are shown in red dotted lines.

if the binding characteristics of the extended Tudor domain are conserved, we measured the binding affinities of TDRKH Tudor domain with different boundaries to the R4me2s peptide and found that TDRKH (residues 290–561) binds R4me2s with a K_D of 4 μ M. As predicted, binding was severely diminished ($K_D = 310 \mu$ M) following C-terminal truncation of the TDRKH Tudor domain (residues 290–430), and was undetectable for a construct covering just the canonical Tudor domain (residues 327–420) (Table 1). These truncation mutants are well behaved during purification and one truncation mutant was crystallized previously (TDRKH, residues 327–420) (8). Therefore, both the N- and C-terminal parts of the split SN domain of SND1, in addition to the Tudor core domain, are essential for ligand binding, and this arrangement is likely conserved in the extended Tudor domains of germ-line-specific Tudor proteins such as TDRKH and Tudor, a notion that is supported by a recent study published during the preparation of our manuscript (26).

Because the SN domain contains five β -strands, which resembles a Tudor domain, we overlaid the fifth SN domain and the canonical Tudor domain of SND1, and found that the β -barrel core of the SN domain can be superimposed with the Tudor domain (Fig. S2B), although the Tudor-like β -barrel core of the SN domain lacks an aromatic cage. Therefore we propose that the extended Tudor domain harbors unique interdigitated tandem Tudor domain folds.

The mechanistic basis for the importance of the SN domain in ligand binding was revealed by the structures of the extended SND1 Tudor domain with complexed peptides. These showed

that all three modules (the N- and C-terminal SN segments and the Tudor core) are involved in binding the methylarginine peptides (Fig. 2). Together the three modules form a wide negatively charged binding groove that accommodates the methylarginine containing peptides. This finding is consistent with the loss of binding displayed by the deletion mutants, because the structures predict that deleting any part of the extended Tudor domain would severely diminish peptide binding.

PIWI1 R4me2s and R14me2s Peptides Bind the Extended Tudor Domain of SND1 in Reverse Directions. Surprisingly, the R4me2s and R14me2s peptides bound the SND1 extended Tudor domain in reverse directions, although the methyl-guanidinium group of the symmetrical dimethylarginine could be well superimposed (Fig. 3A). Both R4me2s and R14me2s peptides form extensive interactions with SND1 extended Tudor domain (Fig. 2). In the crystal structure of the SND1-R4me2s complex, the Ala7 of the peptide is inserted into a shallow pocket and Ala5 is pointed to the sidewall of the binding groove (Fig. 2B and Fig. S3). Arg6 forms electrostatic and water-mediated hydrogen bonding interactions with SND1 (Fig. 2B and Fig. S3). This conformation can explain why the RA/RG motif is preferred by SND1. However, in the case of the SND1-R14me2s complex, if the R14me2s still adopted the same orientation as the R4me2s peptide, the Glu17 residue would occupy the pocket filled by Ala7 in the case of the SND1-R4me2s structure, which would be too shallow for Glu17, and more importantly would cause charge expulsion (Fig. 2 and Fig. S3). Therefore, from an energetic point of view,

SND1 mutant name	K_D , μM
SND1-deltaN	No binding
SND1-deltaC	300 ± 50
TDRKH (290–530)	4 ± 1
TDRKH (290–430)	310 ± 50
TDRKH (327–420)	No binding

By superimposing the R4me2s and R14me2s structures, we note that the peptides do not overlay well except for the methylarginine group (Fig. 34). It is conceivable that the wide binding groove of the SND1 extended Tudor domain provides plasticity to accommodate methylarginine peptides in different binding modes (Fig. 34). In addition to the first cluster of RA/RG repeats in PIWIL1, Piwi proteins harbor methylarginines in other sequence contexts (5). The wide binding groove of the extended Tudor domain in SND1 potentially enables it to recognize multiple different methylarginine motifs, a prediction that warrants further investigation.

A series of mutagenesis experiments were performed to verify the importance of these residues in SND1 for methylarginine binding. Mutating F740, Y746, Y763, or Y766 to leucine in SND1 reduced the binding to the R4me2s peptide by 11–22-fold (Table S1). Mutating N768 to aspartic acid increased the binding affinity twofold, presumably by introducing extra electrostatic interactions with the methylarginine; in contrast, mutating N768 to glutamine or arginine disrupted binding (Table S1). N768R introduces an unfavorable positive charge and the N768Q mutation possibly disrupts the hydrogen bond with the methylarginine. These data show that the methylarginine binding pocket observed in the crystal structures is essential in binding a methylated arginine peptide.

The extended Tudor domain has a similar aromatic cage to that observed in the 53BP1 and L3MBTL1 structures (20, 29). How-

Arginine methylation sites on the N termini of Piwi proteins are evolutionarily conserved. The RG/RA-rich clusters provide docking sites for proteins with Tudor domains (8), resembling the recognition of histone modifications by methyllysine-binding domains. Given that there are six arginine sites in the first RG/RA-rich cluster on the PIWIL1/Miwi N terminus, multiple arginine residues may be methylated simultaneously. Our binding data showed that a peptide with three symmetrically dimethylated arginines (R4, R10, and R14) only showed slightly increased affinity toward the extended Tudor domain of SND1, indicating that one methylarginine is adequate for binding a single extended Tudor, and that there is no obvious requirement for multivalent interactions. However, it is conceivable that multiple methylation sites may increase the local concentration of available methylated ligands and enhance the possibility of initial recruitment of the extended Tudor domain. On the other hand, some germ-line Tudor domain proteins contain multiple Tudor domains. For instance, TDRD6 has eight different Tudor domains, which potentially could bind multiple distinct arginine methylation marks at the same time (30). Regardless, our finding that the binding surface of the SND1 extended Tudor domain can bind methylated peptides in opposite orientations suggests a surprising plasticity in its ability to accommodate arginine methylated sites. During the preparation of our manuscript, a significant study has shown the crystal structure of extended Tudor domain of *Drosophila* Tudor with arginine methylated peptides of *Drosophila* Aubergine (26).

Fig. 3. Methylarginine binding pockets in SND1. (A) Superposition of SND1-R4me2s and SND1-R14me2s complexes. R4me2s peptide is colored in yellow, and R14me2s peptide is colored in aquamarine. Symmetrically dimethylated arginines from both complexes are shown in stick models. (B) Methylarginine binding pocket residues and R4me2s peptide are shown in stick models. (C) Methylarginine binding pocket residues and R14me2s peptide are shown in stick models. (D) Model of asymmetrical dimethylated arginine bound by SND1. Note the absence of a hydrogen bond between Rme2s and N768. (E) Methyllysine-binding pocket in the 53BP1-H4K20me2 complex.

As predicted (19), the extended Tudor domain of the *Drosophila* Tudor protein displays similar domain folds and conserved aromatic cage residues for binding sDMA to the ancestral SND1 polypeptide. Together, these findings underscore a previously underappreciated requirement of the flanking sequences of canonical Tudor core domain in assisting the binding of methylarginine-containing peptides, a feature that is likely conserved across germ-line Tudor proteins (19).

A role for SND1 in small RNA pathways was first suggested by the discovery of SND1 as a component of RNA induced silencing complex of the siRNA pathway. Subsequently, SND1 has been shown to bind and cleave inosine-containing hyperedited micro-RNAs through its SN domains (31). However, the binding substrates for SND1 Tudor domain in these physiological contexts are still unclear. In this study, we showed the coexpression and interaction of PIWIL1/Miwi and SND1, suggesting a previously undescribed function for SND1 in regulating piRNA pathways. It will be of considerable interest to investigate whether SND1 may affect piRNA biogenesis, processing, or stability. A mouse knockout model is necessary to further understand the biological role of SND1 in these processes.

Experimental Procedures

Protein Expression and Purification. The extended Tudor domain of human SND1 (residues 650–910) was cloned into a modified

pET28-MHL vector to generate N-terminal His-tagged fusion protein. The recombinant protein was overexpressed in *Escherichia coli* BL21 (DE3)-V2R-pRARE2 at 16°C and purified by affinity chromatography on Ni-nitrilotriacetate resin (Qiagen) followed by TEV protease treatment to remove the tag. For crystallization experiments, purified protein was further treated with Endoproteinase Glu-C. All of them were further purified by anion-exchange column. Finally, the protein was concentrated to 15 mg/mL in a buffer containing 20 mM Tris • HCl, pH 7.5, 0.2 M NaCl, and 1 mM DTT.

Mutant proteins were also overexpressed in *Escherichia coli* BL21 (DE3)-V2R-pRARE2 and purified using the same procedure as described above, except further purification using size-exclusion chromatography (Hi Load 16/60 Superdex 75, GE Healthcare Biosciences). The molecular weight of all protein samples was checked by mass spectrometry.

Binding Assays. All peptides used for fluorescence polarization and isothermal titration calorimetry (ITC) measurements were synthesized by Tufts University Core Services. Fluorescence polarization binding assay was performed in 10 μ L at a constant fluorescence labeled-peptide concentration of 40 nM and increasing amounts of SND1 (residues 650–910) at concentrations ranging from low to high micromolar in a buffer of 20 mM Tris • HCl, pH 7.5, 50 mM NaCl or 150 mM NaCl, 1 mM DTT, and 0.01%