Structures and functions of the multiple KOW domains of transcription elongation factor Spt5

Peter A. Meyer[1], Sheng Li[2], Mincheng Zhang[1], Kentaro Yamada[3], Yuichiro Takagi[3], Grant A. Hartzog[4] and Jianhua Fu[1,#]

[1] Department of Biochemistry, Medical College of Wisconsin, Milwaukee, WI 53226, USA;

[2] Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, 201210, China;

[3] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202, USA;

[4] Department of Molecular, Cell & Developmental Biology, University of California, Santa Cruz, CA 95064, USA.

Corresponding author: jfu@mcw.edu; Tel: 414-955-5849; Fax: 414-955-6510

Supplementary Results

Folding units in the KOW-encompassing region of Spt5 and crystallography

Comparison of eukaryotic Spt5 sequences demonstrated strong conservation across species and aided recognition of domains (Supplementary Fig. 1). Comparison of consensus sequences of the KOW/Tudor domains of Spt5, however, revealed a lack of conservation among the individual domains; only the KOW signature motif could be discerned as common to all KOWs within Spt5 and NusG (Supplementary Fig. 2B). This observation suggests that the individual KOWs, although likely following a common Tudor fold, differ in the chemical properties of their tertiary structures. In order to structurally characterize the KOW-encompassing region centrally located between the NGN and CTR domains, we first subjected its sub-regions of various lengths to disorder as well as fold prediction (see Materials and Methods). Rosetta runs returned folds close to Tudor domains for the linker-free KOW regions as had been expected based on their KOW motif. The linker regions were predicted to be partially unstructured by the disorder algorithm used, but suggested to be folded in Rosetta results. To understand this property better, we conducted recombinant protein expression and solubility experiments for the KOWs with and without linkers. From a total of eight constructs

covering KOW1 through KOW5 (Supplementary Fig. 2C), five were found to be soluble: K1 (K: KOW), K1L1 (L: linker), K2K3, L2K4 and L3K5, whereas K4 alone or K4L3 were insoluble and K5 alone is partially soluble. As such, four soluble domains—K1L1, K2K3, L2K4 and L3K5—are considered as putative folding units of the entire central region. The soluble and partially soluble constructs were purified using their C-terminal His6 tags and additional chromatographic steps (Materials and Methods), and subjected to crystallographic and biochemical characterizations.

Crystallization screening and optimization yielded single crystals for the K1L1 and K2K3 constructs. Synchrotron diffraction and data collection allowed structure determinations to 1.09 Å and 1.6 Å for K1L1 and K2K3, respectively (Supplementary Table-1). Both structures were solved through experimental phasing, as molecular replacement with various published KOW and Tudor domain structures failed (Materials and Methods). The high resolutions and data strength produced robust structural refinement results (Supplementary Table-1). Water molecules in mainly the first hydration shell were visible in difference Fourier maps throughout the refinement process and incorporated into the coordinates. The electron density map of K2K3 also revealed locations of larger solutes that had been included in the crystallization mother liquor and crystal cryo-soaking solution (Supplementary Table-1).

Crystal structure of the KOW1-Linker1 domain

The high-resolution experimental electron density map of K1L1 was readily traceable and demonstrated a folded structure twice as large as single Tudor domain. Model building and crystallographic refinement established a high-quality structure consisting of two lobes connected by two coil-like strands (Fig. 1A). A search for homologous structures in the DALI database found high Z-score matches that included proteins with typical Tudor domain fold, e.g. the NusG family proteins, PHD finger protein-1 Tudor domain, human KIN17 C-terminal KOW domain, SH3 domains of SRC kinases and certain ribosomal proteins.

The K1L1 tertiary structure is not physically divisible between K1 and L1 due to the extensive interdomain interface that buries a surface area of 1145 $Å^2$. The interface involves a minimum of five hydrogen bonds and a cation:π electron interaction in addition to van der Waals interactions (Fig. 2A). The interface H-bonding pairs are without exception strictly conserved among all eukaryotic Spt5 proteins (Fig. 2B), indicating a conserved interdomain organization.

This observation, together with a strong overall sequence conservation, suggests that the yeast structure determined here is a universal model for all K1L1 domains in eukaryotic Spt5.

The ultra-high resolution of K1L1 diffraction data revealed great details for most residues, except for the segment of I428-R438 and last three residues at the C-terminus that did not show electron density. The lack of observable density is interpreted as structural flexibilities in the two regions. Amino acid sidechains of most residues were precisely defined in the 2Fo-Fc density map during the course of structural refinement (Fig. 2C). Dual side-chain conformations (rotamers) exist for M401, S406, N409 and I498, all belong to the K1 Tudor barrel, and the residues are solvent exposed. Occupancies for the minor rotamers vary from 0.28 to 0.46. Among these, M401 is located toward the N-terminus of the loop structure between β2 and β3 of the barrel. Moving toward the C-terminus of this loop, S406 and N409 are found situated at the entry and exit points of a β-turn, respectively, which allows the β3 strand to run in an anti-parallel direction to the loop. As such, their rotamer conformations and occupancies seem to be coupled: conformation A for S406 and N409 is at 0.67 and 0.66, respectively; while conformation B is at 0.33 and 0.34, respectively. This observation indicates that the pair in conformation A is somewhat energetically more favorable than in conformation B. Nevertheless, the main-chain trajectory of β2-loop-β3 remains uniquely defined. N409 is located in the loop connecting β7 to the C-terminal α2 helix and faces inward against the hydrophobic wall formed between V388, the alkyl carbons of R389 (both of β1), and L399 (β2); it apparently is involved in the packing of the C-terminal structure including the α2 helix. Since the average thermal factor for the C-terminal structure is quite low (< 20), the dual side-chain conformation of N409 does not necessarily suggest conformational disorder in the C-terminal structure up to the last observed residue (F508). Side-chains were incompletely modeled for E383, E407, R439, Q461, K496, E503 and R507 due to either missing or weak density, although the main-chain was continuous through these residues.

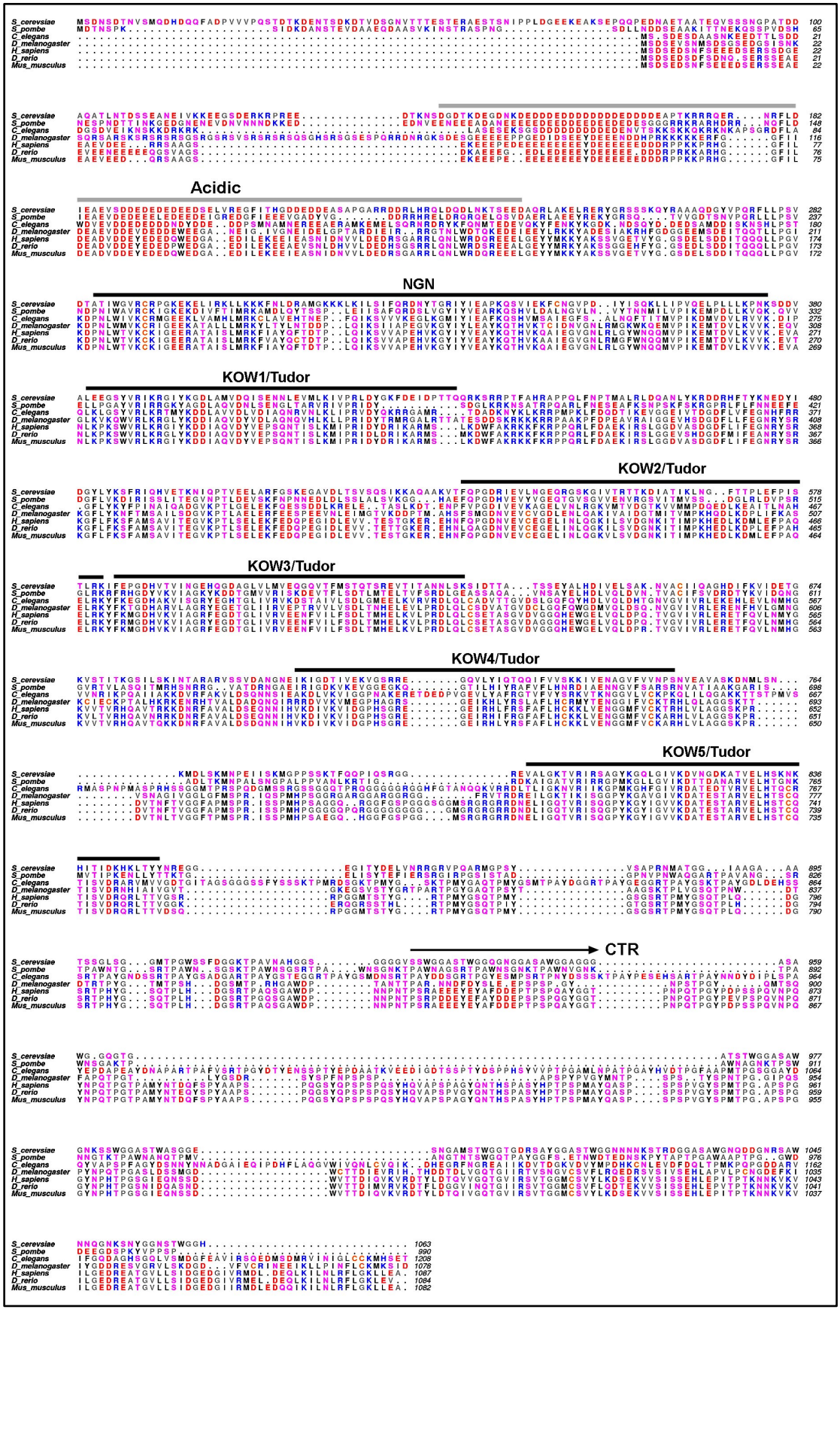
Crystal structure of the KOW2-KOW3 tandem domains

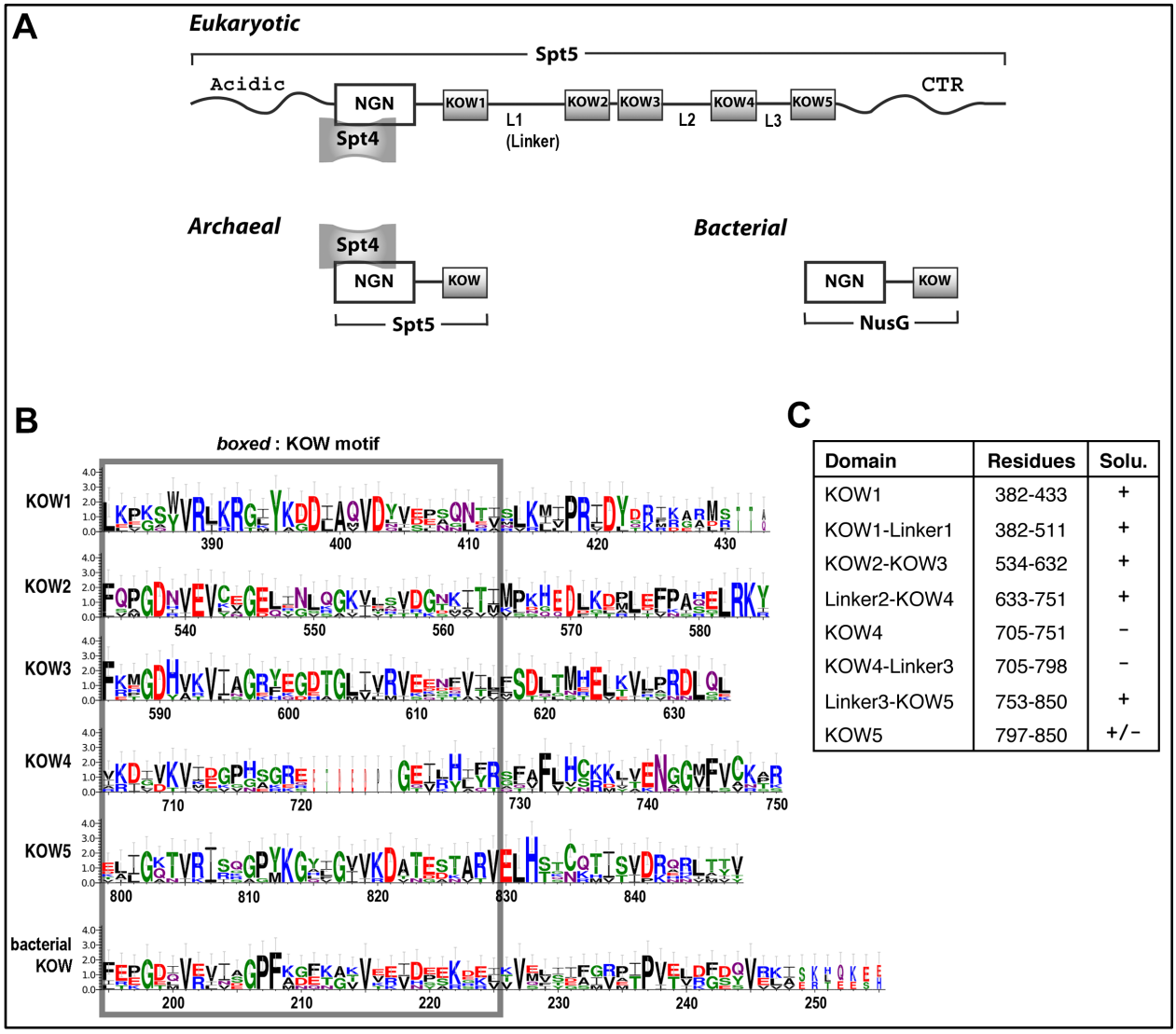
The recombinant protein of K2K3 domains crystallizes with two independent molecules (chain-A and -B) in an asymmetric unit of the unitcell (Supplementary Table-1). The final refined models demonstrate no significant difference: RMS deviation between molecule-A and -B is 0.75 Å for main-chain atoms and 1.07 Å for all atoms. Two residues, R548 of molecule-A

and R621 of molecule-B, showed alternative side-chain conformations according to the high-resolution density map. The K2 structure encompasses residues from F534 to K582, and K3 from F584 to K632, both consisting of 49 residues. Side-chain model could not be established for R539, N594, E596 and E608 of molecule-A, and R584, K559 and E596 of molecule-B, while the main-chain structures were continuous for both molecules. We did not observe a significantly flexible structural element in the K2K3 structure. There are high degrees of similarity between the individual barrels within the K2K3 structures. For molecule-A, the RMS deviations between K2 and K3 are 1.06 Å and 1.28 Å for the main-chain and side-chain atoms, respectively. For molecule-B, the values are 1.18 Å and 1.35Å for the main-chain and side-chain, respectively. The description in the main text is given based on molecule-A.

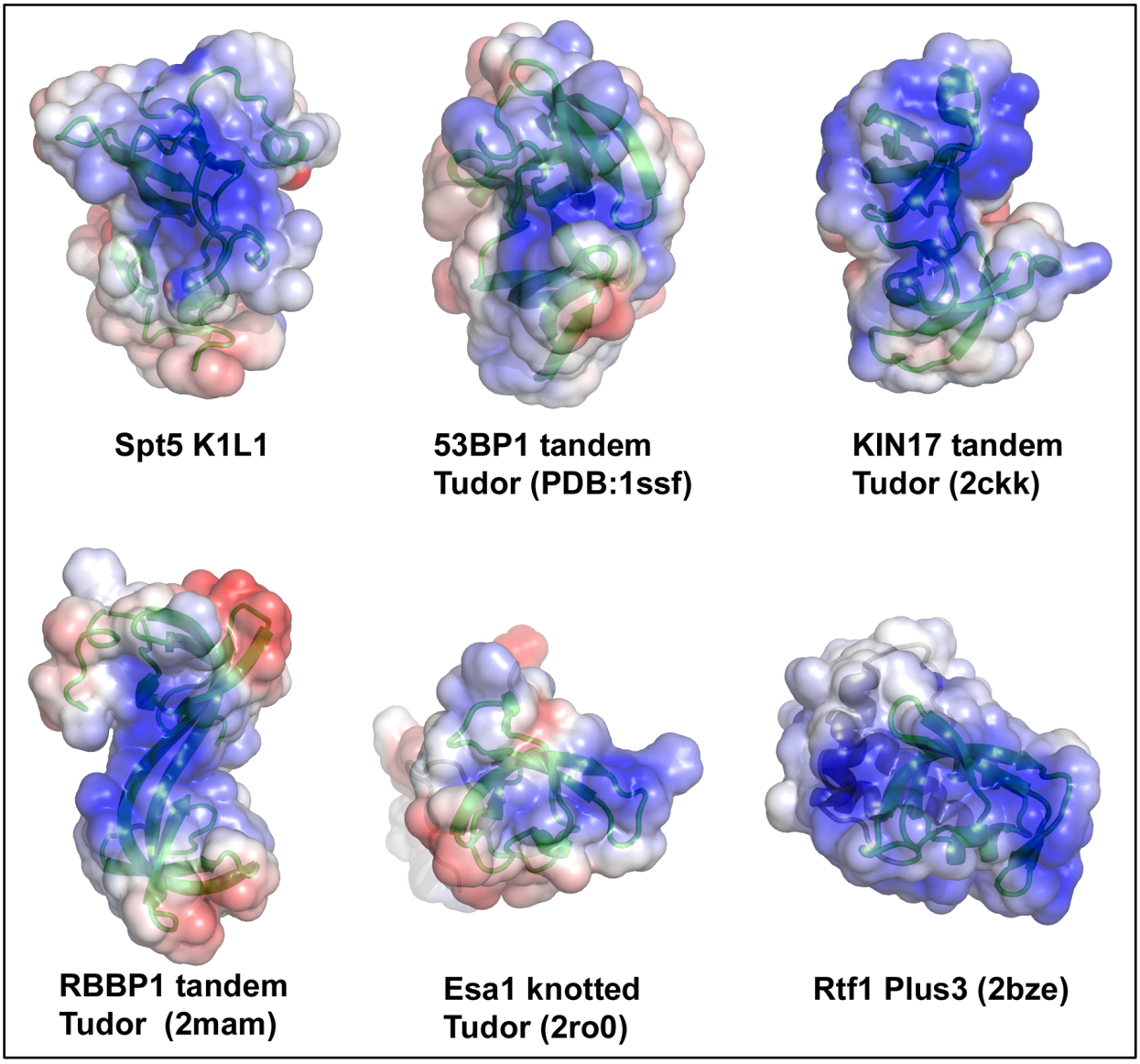
Supplementary Figures

Supplementary Fig. 1 (next page). A multiple sequence alignment for Spt5 proteins from various eukaryotic sources. The bars on top of the sequences indicate major structural domains that have been identified based on sequence homologies with NusG family proteins. Definition for the KOW/Tudor domains was additionally aided by inspection of the available three-dimensional structures of Tudor domains including those of *T. thermophilus* NusG, human Snd1, 53BP1 of human and frog, Hsr9 of *C. elegans*, *S. cerevisiae* Rad9, *S. pombe* Rhp9 and the second, third and fifth KOW domains of human Spt5 (NMR structures).

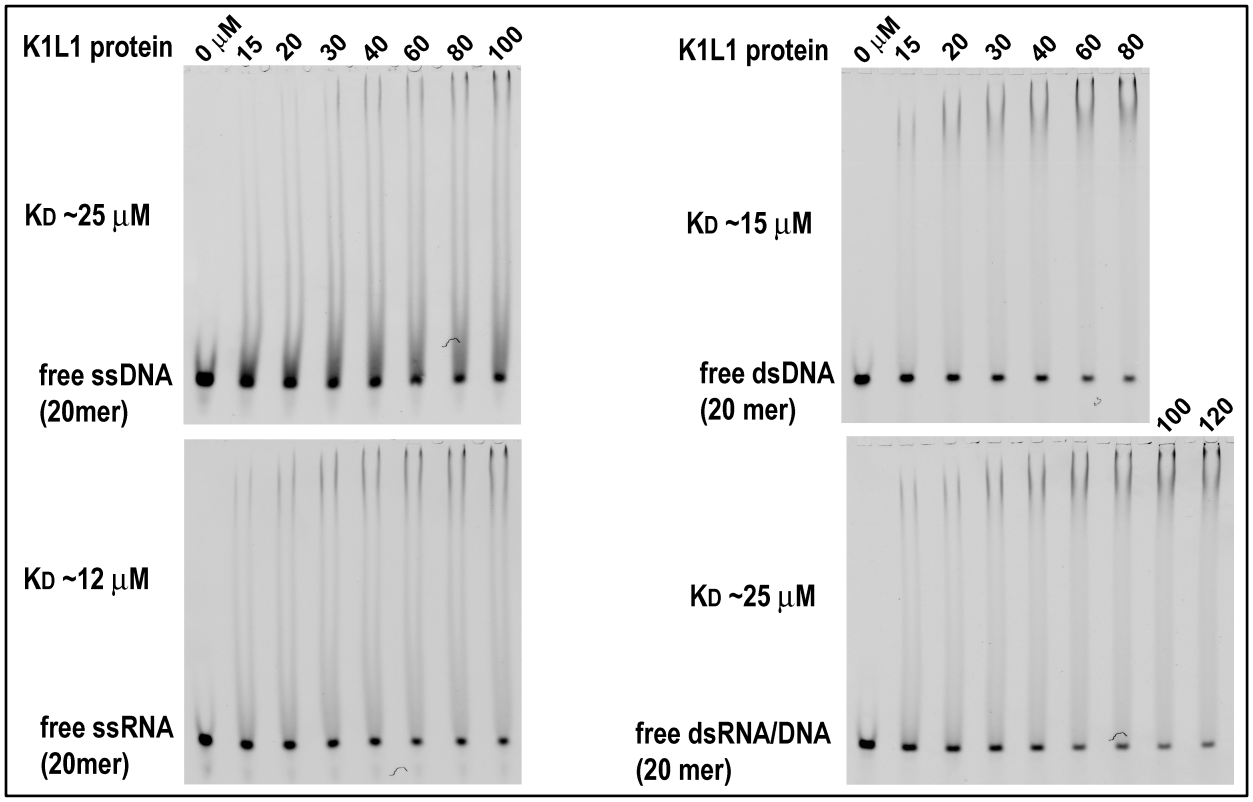

Acidic
NGN
KOW1/Tudor
KOW2/Tudor
KOW3/Tudor
KOW4/Tudor
KOW5/Tudor
CTR
S_cerevsiae
S_pombe
C_elegans
D_melanogaster
H_sapiens
D_rerio
Mus_musculus



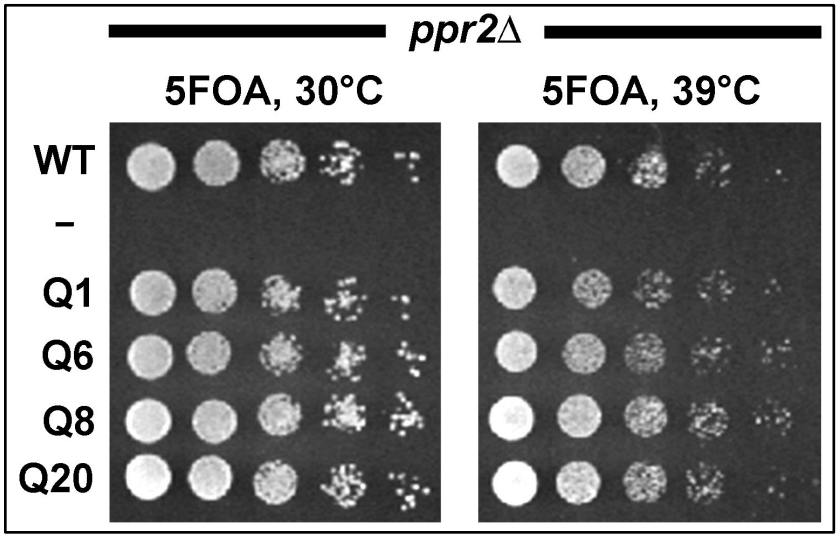
Supplementary Fig. 2. Spt5 contains multiple KOW/Tudor domains in its central region. (**A**) Domain compositions of Spt5 and its homologs in bacteria and archaea. (**B**) Amino acid consensus sequence for eukaryotic KOW1 through KOW5 with the numbering in *Saccharomyces cerevisiae,* and consensus for the single KOW of bacteria with the numbering in *A. aeolicus*. The *grey* box delimits the canonical KOW motif encompassed within the KOW domains. (**C**) Domain definition and protein solubility for recombinant yeast KOWs expressed in *E. coli*.



Supplementary Fig. 3. Surface electrostatic potentials of Spt5 K1L1 domain and Tudor domains implicated for interactions with nucleic acids. Scale of the charges ranges from +10 (*blue*) to -10 (*red*). A probe with a radius of 1.4 Å was used in calculating the solvent accessible surfaces.



Supplementary Fig. 4. Typical gel-shift runs to assess K1L1 binding constants toward different forms of nucleic acids of 20 nucleotides long (5'-labeled with TAMRA). All the probes were set to 5 μM in a binding mix, and the K1L1 concentration incremented as indicated on the top of each gel. The gels were imaged on a Typhoon scanner and quantified with ImageQuant (Materials and Methods). Four to six independent experiments were performed to obtain a mean binding curve for each form of the nucleic acid, and K_D values were derived from the 50% saturation point of a binding curve. The smeared appearance of bound complex bands is explained in Fig. 5 legend.



Supplementary Fig. 5. The non-lethal *spt5-PCP* mutations do not have Ts growth defects in the *ppr2Δ* (TFIIS-delete) background. The *spt5-PCP* alleles were shuffled into the GHY2731 (*ppr2Δ*) strain, and growths of the resulting strains were assayed on 5FOA-containing plates using serial dilutions and incubation at 30° and 39°C.

Supplementary Tables

Supplementary Table-1: Crystallographic Data Statistics

I. X-ray diffraction data

Crystal: K1L1

Dataset Name	Native 1 – Sulfur edge	Native 2 – higher resolution
Wavelength (Å)	1.5484	0.9793
Unitcell edges (Å)	a = 42.75, b = 44.43, c = 64.29	a = 42.69, b = 44.39, c = 64.07
Space group	$P2_12_12_1$	$P2_12_12_1$
Asymmetric unit	1 molecule	1 molecule
Resolution (Å)	50.0 – 1.55 (1.64 – 1.55)	50.0 – 1.09 (1.11 – 1.09)
Completeness (%)	93.1 (58.5)	93.9 (52.1)
R_{merge} (%)	4.8 (60.3)	11.5 (33.7)
I/σ<I>	48.3 (2.6)	14.0 (2.6)
Redundancy	22.2 (5.0)	6.0 (2.3)

Crystal: K2K3

Dataset Name	Se-methionine (Se edge)
Wavelength (Å)	0.9793
Unitcell edges (Å)	a = 44.25, b = 54.24, c = 95,15
Space group	$P2_12_12_1$
Asymmetric unit	2 molecules
Resolution (Å)	50.0 – 1.60 (1.63 – 1.60)
Completeness (%)	89.5 (52.5)
R_{merge} (%)	8.9 (18.7)
I/σ<I>	82.2 (12.2)
Redundancy	12.3 (5.1)

II. Structural refinement

	K1L1	K2K3
Resolution range (Å)	36.5 – 1.09	34.3 – 1.60
Number of reflections (free set)	48357 (2478)	27657 (1391)
Amino acid residues	116	198
Water/solute molecules	161/0	241/6
R-work	0.182	0.185
R-free	0.199	0.203
Standard stereochemistry		
RMS bond deviation (Å)	0.009	0.019
RMS angle deviation (°)	1.321	1.655
RMS chiral deviation (°)	0.091	0.135
Clash score	3.8	6.9
Ramachandran plot		
favored/allowed/outlier (%)	98.3/1.7/0.0	98.9/0.5/0.5
Side-chain rotamer outliers	0	0
Average B-factors ($Å^2$)		
all protein	13.9	15.4
main-chain	12.7	12.4
side-chain	15.0	18.6
water	27.7	31.3
solute (sulfate & glycerol)	N/A	40.5

Supplementary Table-2:

Tudor domain proteins with reported nucleic acid-binding activities

Protein	PDB Code	Domain Type	Binding	Assay Method	K_d (μM)	Ref
53BP1	1ssf	tandem Tudor	dsDNA	NMR	460	(1)
KIN17	2ckk	tandem Tudor	RNA	blotting, pulldown	UD[a]	(2, 3)
RBBP1	2mam	tandem Tudor	ss & dsDNA	NMR, ITC	16.0--99	(4)
Esa1	2ro0	knotted Tudor	poly-U, -C	pulldown, NMR	28.0, 21.0	(5)
Rtf1	2bze	Plus3/Tudor	ss & bubble DNA	Gel shift	3.0 (bubble)	(6)

[a] un-determined.

References to Supplementary Table-2:

1. **Charier G, Couprie J, Alpha-Bazin B, Meyer V, Quemeneur E, Guerois R, Callebaut I, Gilquin B, Zinn-Justin S.** 2004. The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. Structure **12:**1551-1562. http://www.ncbi.nlm.nih.gov/pubmed/15341721

2. **le Maire A, Schiltz M, Stura EA, Pinon-Lataillade G, Couprie J, Moutiez M, Gondry M, Angulo JF, Zinn-Justin S.** 2006. A tandem of SH3-like domains participates in RNA binding in KIN17, a human protein activated in response to genotoxics. J Mol Biol **364:**764-776. http://www.ncbi.nlm.nih.gov/pubmed/17045609

3. **Pinon-Lataillade G, Masson C, Bernardino-Sgherri J, Henriot V, Mauffrey P, Frobert Y, Araneda S, Angulo JF.** 2004. KIN17 encodes an RNA-binding protein and is expressed during mouse spermatogenesis. J Cell Sci **117:**3691-3702. http://www.ncbi.nlm.nih.gov/pubmed/15252136

4. **Gong W, Wang J, Perrett S, Feng Y.** 2014. Retinoblastoma-binding protein 1 has an interdigitated double Tudor domain with DNA binding activity. J Biol Chem **289:**4882-4895. http://www.ncbi.nlm.nih.gov/pubmed/24379399

5. **Shimojo H, Sano N, Moriwaki Y, Okuda M, Horikoshi M, Nishimura Y.** 2008. Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J Mol Biol **378:**987-1001. http://www.ncbi.nlm.nih.gov/pubmed/18407291

6. **de Jong RN, Truffault V, Diercks T, Ab E, Daniels MA, Kaptein R, Folkers GE.** 2008. Structure and DNA binding of the human Rtf1 Plus3 domain. Structure **16:**149-159. http://www.ncbi.nlm.nih.gov/pubmed/18184592