doi:10.1016/j.jmb.2011.03.002 J. Mol. Biol. (2011) 408, 697–713

Contents lists available at www.sciencedirect.com

Journal of Molecular Biology

journal homepage: http://ees.elsevier.com.jmb

Crystal Structures of the *S. cerevisiae* Spt6 Core and C-Terminal Tandem SH2 Domain

Devin Close[1], Sean J. Johnson[2], Matthew A. Sdano[1], Seth M. McDonald[1], Howard Robinson[3], Tim Formosa[1] and Christopher P. Hill[1]*

[1]*Department of Biochemistry, University of Utah, Salt Lake City, UT 84112-5650, USA*
[2]*Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, USA*
[3]*Department of Biology, Brookhaven National Laboratory, Upton, NY 11973, USA*

Received 2 December 2010; received in revised form 24 February 2011; accepted 1 March 2011
Available online 17 March 2011

Edited by K. Morikawa

Keywords:
protein structure; protein function; gene expression; crystallography

The conserved and essential eukaryotic protein Spt6 functions in transcription elongation, chromatin maintenance, and RNA processing. Spt6 has three characterized functions. It is a histone chaperone capable of reassembling nucleosomes, a central component of transcription elongation complexes, and is required for recruitment of RNA processing factors to elongating RNA polymerase II (RNAPII). Here, we report multiple crystal structures of the 168-kDa Spt6 protein from *Saccharomyces cerevisiae* that together represent essentially all of the ordered sequence. Our two structures of the ~900-residue core region reveal a series of putative nucleic acid and protein–protein interaction domains that fold into an elongated form that resembles the bacterial protein Tex. The similarity to a bacterial transcription factor suggests that the core domain performs nucleosome-independent activities, and as with Tex, we find that Spt6 binds DNA. Unlike Tex, however, the Spt6 S1 domain does not contribute to this activity. Crystal structures of the Spt6 C-terminal region reveal a tandem SH2 domain structure composed of two closely associated SH2 folds. One of these SH2 folds is cryptic, while the other shares striking structural similarity with metazoan SH2 domains and possesses structural features associated with the ability to bind phosphorylated substrates including phosphotyrosine. Binding studies with phosphopeptides that mimic the RNAPII C-terminal domain revealed affinities typical of other RNAPII C-terminal domain-binding proteins but did not indicate a specific interaction. Overall, these findings provide a structural foundation for understanding how Spt6 encodes several distinct functions within a single polypeptide chain.

*Corresponding author. E-mail address: chris@biochem.utah.edu.
Present address: D. Close, Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA.

Abbreviations used: CTD, C-terminal domain; DLD, death-like domain; dsDNA, double-stranded DNA; FA, fluorescence anisotropy; HhH, helix–hairpin–helix; HtH, helix–turn–helix; NSLS, National Synchrotron Light Source; OB, oligonucleotide–oligosaccharide binding; PDB, Protein Data Bank; pSer, phosphoserine; pTyr, phosphotyrosine; RNAPII, RNA polymerase II; SH2, Src homology 2; SSRL, Stanford Synchrotron Radiation Laboratory; WT, wild type; SeMet, selenomethionine substituted; NIH, National Institutes of Health.

Introduction

Gene expression in eukaryotes relies on a synergistic relationship between transcription, RNA processing, and chromatin structure.[1,2] The specific positioning, composition, and posttranslational modification of nucleosomes defines a code for chromatin-templated transcriptional regulation. Moreover, transcription is intimately tied to mRNA processing, surveillance, and export from the nucleus. This coordination relies on precise cooperation among many proteins, with Spt6 being remarkable for playing multifaceted roles in several distinct processes.

Spt6 (suppressor of Ty 6) was originally discovered in *Saccharomyces cerevisiae* as a gene that influences general transcription through manipulation of chromatin structure at upstream promoter elements.[3] Subsequently, Spt6 has been implicated in a variety of biological processes in organisms ranging from yeasts to human, including embryogenesis in zebrafish,[4] multiple stages of development in *Drosophila*,[5] gut morphogenesis in *Caenorhabditis elegans*,[6] signal transduction in mammals,[7,8] and pathogenesis of human immunodeficiency virus.[9,10] The broad utility of Spt6 stems from its ability to perform multiple functions as a histone chaperone, a transcription elongation factor, and a modulator of RNA transcript processing.

Spt6 is required for reassembly of nucleosomes in the wake of an elongating RNA polymerase II (RNAPII), a function that has profound regulatory effects at both intergenic and intragenic start sites.[11,12] Spt6 binds directly to histones and nucleosomes *in vitro*,[13,14] and these activities may contribute to the nucleosome reassembly function. In addition, Spt6 recruits the H3K36 methyltransferase Set2 to the transcription complex,[15] providing a link between the processes of transcription and histone modification. While its roles in modifying and reassembling nucleosomes indirectly influence the elongation rate, Spt6 also directly affects RNAPII, as it stimulates elongation on nucleosome-free DNA templates *in vitro*.[9,16] This role as an elongation factor independent of its effects on chromatin may also be significant *in vivo*, as knocking down Spt6 caused a decrease in the RNAPII elongation rate even in regions where the chromatin was considered to be permissive to transcription.[16] Yet another role as a modulator of transcript processing is indicated by the association of Spt6 with the Rrp6 subunit of the *Drosophila* exosome RNA processing complex[17] and by the requirement for Spt6 to prevent premature 3′ processing at cryptic polyadenylation signals upstream of the appropriate sites.[18] It has also been demonstrated that mammalian Spt6 can bind RNAPII C-terminal domain (CTD) phosphorylated at Ser2 by the P-TEFb kinase and that this interaction can subsequently promote recruitment of RNA processing/export factors such as REF1/Aly.[9,15] Binding to the phosphorylated RNAPII CTD is mediated by a Src homology 2 (SH2) domain that is located near the C-terminus of Spt6 and is conserved from yeast to human.[9] SH2 domains typically recognize phosphorylated tyrosine residues, are ubiquitous in metazoans, and are the primary recognition motif in phosphorylation-mediated signal transduction cascades.[19] Strikingly, the Spt6 SH2 domain is the only SH2 domain predicted to occur in the yeast

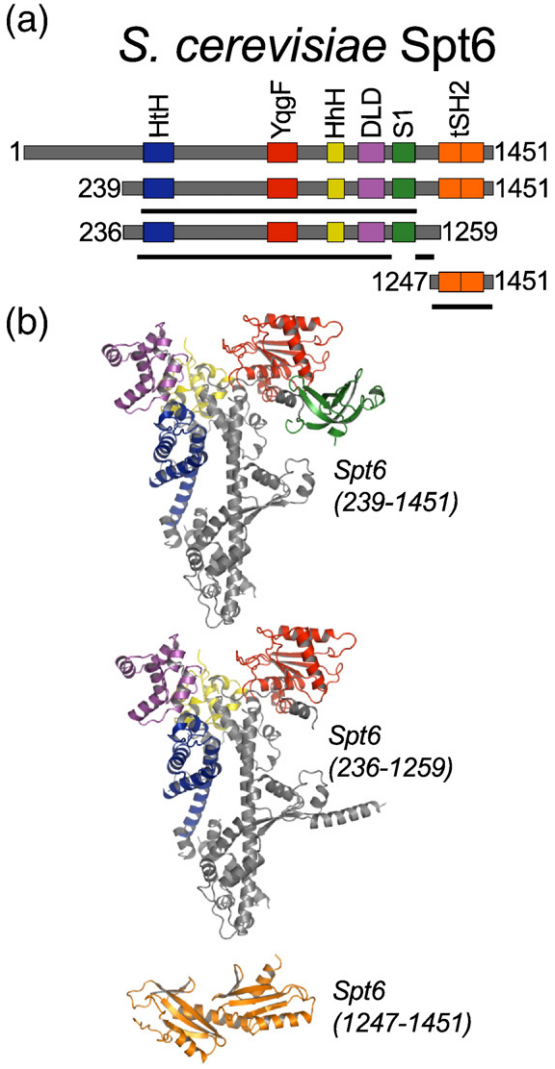


Fig. 1. Spt6 structures. (a) Schematic representation of full-length (top) Spt6 and constructs used for crystallization. The three crystallized constructs (239–1451, 236–1259, and 1247–1451) are shown below, with black continuous lines indicating regions of the protein constructs visible in each of the crystal structures. (b) Three independently determined Spt6 crystal structures colored by domain as in (a).

proteome.[20] Spt6 therefore participates in a wide range of functions affecting transcription, with each activity requiring different subsets of its multiple distinct functional domains.

We have determined multiple crystal structures of Spt6 from *S. cerevisiae* and find that, consistent with the range of functional domains inferred from previous studies, it comprises a series of structural domains whose homologs are known to function in nucleic acid binding and/or protein–protein interactions. The core of the structure comprises several recognizable structural motifs and, in composite, resembles the bacterial transcription factor Tex.[21] A C-terminal region that is tethered to the core by a flexible linker adopts a novel tandem SH2 domain comprising two closely associated SH2 folds, one of which corresponds to the previously predicted SH2 domain of Spt6 and contains many of the standard binding determinants characteristic of this family, while the other lacks these features but contributes to a putative specificity pocket of the more canonical SH2 domain. Our structure of the Spt6 tandem SH2 domain resembles two recently reported homologous Spt6 structures.[22,23] We also show that the Spt6 core domain has DNA-binding activity, and we examine the interaction between the Spt6 tandem SH2 domain and RNAPII-derived peptides for evidence of a phosphorylation-dependent interaction with the CTD.

Results and Discussion

Crystal structures of Spt6(236–1259), Spt6(239–1451), and Spt6(1247–1451)

We have determined three crystal structures that together comprise the entire ordered region of the 1451-residue *S. cerevisiae* Spt6 protein (Figs. 1–3). Based on these structures and sequence analysis, Spt6 residues 1–297, 456–464, 485–500, 562–566, 1003–1008, 1211–1217, and 1441–1451 are likely to

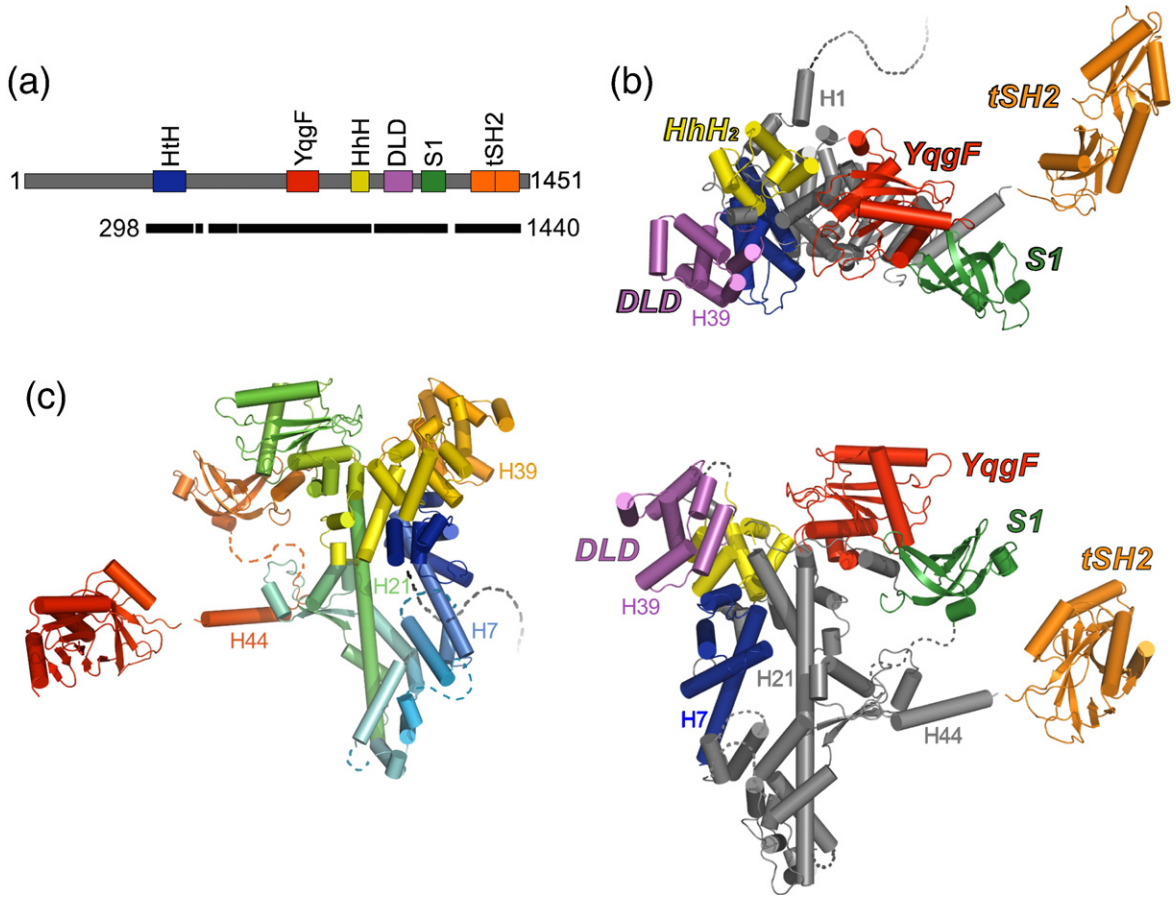


Fig. 2. Composite model of Spt6. (a) Schematic model of the Spt6 protein as in Fig. 1a. Black bar indicates segments of the protein represented by the composite model. (b) Two views of the composite model of Spt6 colored by domain. Approximately two hundred C-terminal residues (tSH2 domain) are expected to be highly mobile with respect to the core. Secondary structure elements mentioned in the text are labeled. Broken lines represent regions of the Spt6 protein that are not visible in our structures and are likely to be disordered, including the first 297 residues. (c) Composite model of Spt6 colored from the N- to C-terminus (blue to red). View orientation is rotated 180° from that of the lower image in (b).

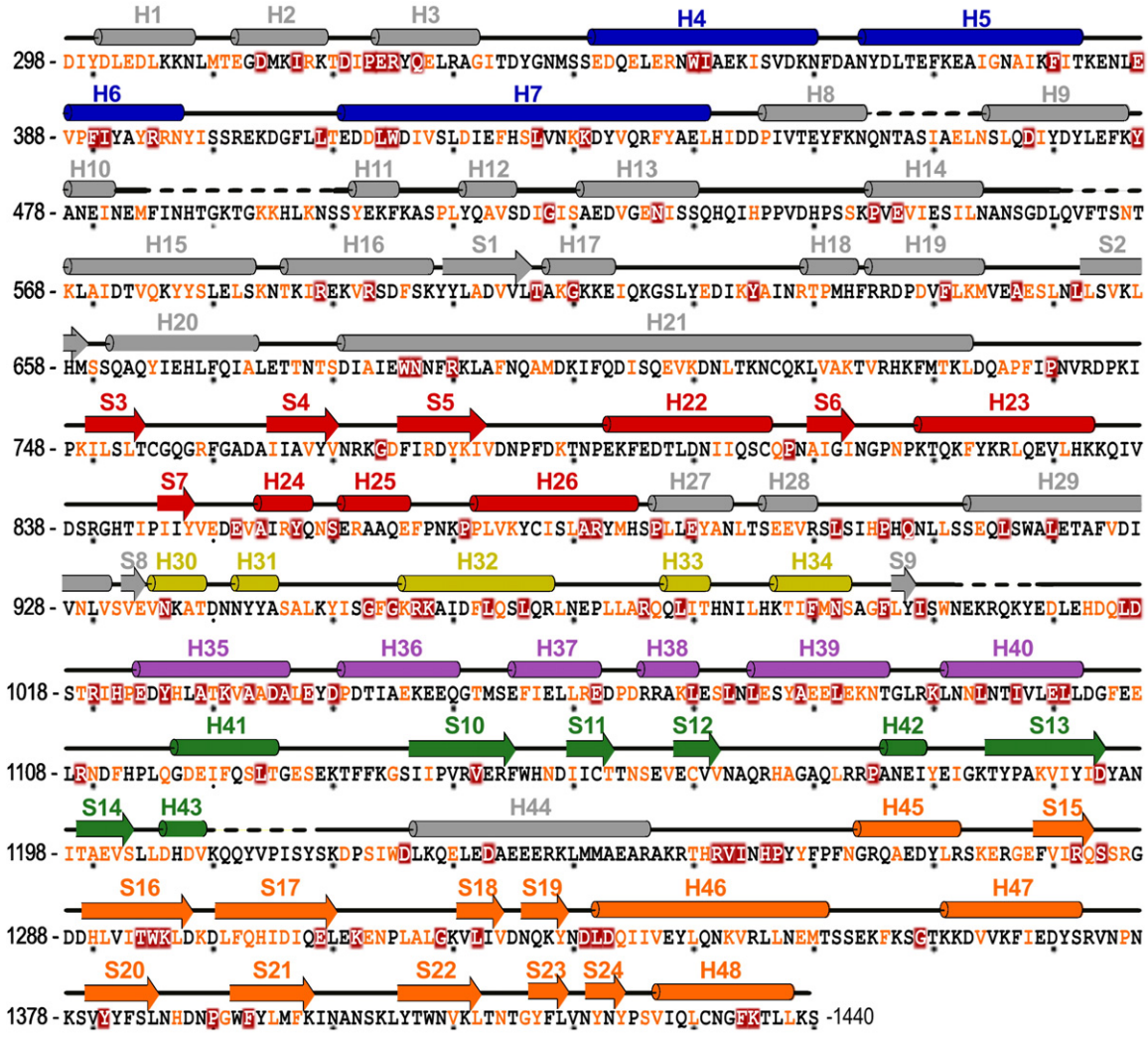


Fig. 3. Spt6 sequence. Spt6 sequence present in the composite model with corresponding secondary structure elements colored by domain as in Figs. 1 and 2. Coloring of sequence represents degree of conservation [dark-red background, invariant; orange font, conserved] in an alignment (see Supplemental Fig. S1) of proteins from *S. cerevisiae*, *Schizosaccharomyces pombe*, *C. elegans*, *Drosophila melanogaster*, *Danio rerio*, and *Homo sapiens*. Alignment was performed using T-coffee.[24] Broken lines indicate regions of disorder that are not included in the model(s).

be disordered in the full-length protein, at least in the absence of binding partners. Spt6 displays multiple recognizable structural domains whose homologs have been implicated in binding of nucleic acids or proteins, although conservation at the sequence level is low, and only three of these domains [$(HhH)_2$, YqgF, and S1] in the Spt6 core were predicted from the sequence.[25]

Full-length Spt6 expressed poorly in *Escherichia coli* and did not yield crystals. In contrast, two different Spt6 constructs that lacked the first ~235 residues and either lacked the C-terminal 192 residues [Spt6 (236–1259)] or extended to the C-terminus [Spt6 (239–1451)] expressed well, and the resulting proteins crystallized in different space groups (Table 1). The Spt6(236–1259) structure was determined by two-wavelength anomalous diffraction using data collected to 2.7 Å from crystals of selenomethionine-substituted (SeMet) protein and was refined against 2.6-Å native data to R/R_{free} values of 22.4%/26.5%. The Spt6(239–1451) structure was determined by molecular replacement using the Spt6(236–1259) structure as a search model and refined against 3.3-Å data to R/R_{free} values of 26.5%/30.8%.

Together, Spt6(236–1259) and Spt6(239–1451) display an ordered structure for the Spt6 core (residues 298–1248), which is primarily helical (54.6% helical, 8.3% strand, 37.1% coil), has overall dimensions of 110 Å × 77 Å × 36 Å, and is quite similar between the two structures (RMSD = 1.2 Å over 747 out of 763 pairs of C^{α} atoms that are ordered in both structures). The core is built around an ~80-Å-long central helix (H21, 680–733), with the rest of the

Table 1. Crystallization conditions

	Spt6(236–1259)	Spt6(236–1259)	Spt6(239–1451)	Spt6(1247–1451)	Spt6(1247–1451)
Crystal	SeMet	Native	Native	SeMet	Native
Protein solution[a]	15 mM Tris, 5% glycerol	15 mM Tris, 5% glycerol	50 mM Tris, 10% glycerol, 5 mM β-mercaptoethanol	15 mM Tris, 5% glycerol	15 mM Tris, 5% glycerol
Protein concentration (mg/ml)	9	8	10	26	26
Reservoir solution	50 mM 4-morpholineethanesulfonic acid (pH 6.0), 13% methyl-2,4-pentanediol, 0.1 M KCl, 5 mM $MgCl_2$	50 mM $Mg(HCO_2)_2$, 10% glycerol	0.1 M Na–4-morpholineethanesulfonic acid (pH 6.5), 13% PEG 4000, 0.2 M $MgCl_2$, 12% glycerol	22% PEG 4000, 0.3 M $(NH_4)_2SO_4$	0.1 M Tris (pH 5.5), 25% PEG 3350, 0.2 M $(NH_4)_2SO_4$
Drop ratio (μl:μl), protein:reservoir	2:2	3:1	3:1.5	2: 2	1:2
Additive	—	0.6 μl 1.0 M NDSB-256	—	—	—
Temperature (°C)	13	13	13	21	21
Cryo solution[b]	30% methyl-2,4-pentanediol	35% glycerol	30% glycerol	30% glycerol	20% glycerol

PEG, polyethylene glycol.
[a] All protein solutions were in 100 mM NaCl and Tris, pH 7.5 (defined at room temperature).
[b] Cryoprotectant solutions were made the same as for the reservoir solution, but including glycerol.

protein chain wrapping around this helix at both ends to give an overall V shape. Additionally, Spt6 (239–1451) displayed interpretable density for the S1 domain (residues 1128–1210), whereas the corresponding density in Spt6(236–1259) was poorly defined, and the S1 domain was not included in this refined structure. As discussed below, we favor a model in which the S1 domain is highly mobile in solution. Interestingly, the 240 C-terminal residues of Spt6(239–1451), encompassing all of the residues absent from the shorter Spt6(236–1259) construct, are not visible in the electron density. The Spt6(239–1451) crystals have an estimated solvent content of 60% with cavities that could accommodate the 240 C-terminal residues, which suggests that this region is tethered by a flexible linker that is mobile both in solution and in the Spt6(239–1451) crystals.

Guided by secondary structure predictions and limited proteolysis experiments (data not shown), we expressed and crystallized the Spt6 C-terminal region, Spt6(1247–1451). This structure was determined by the selenomethionine single-wavelength anomalous diffraction method using 2.7-Å-resolution data and refined to R/R_{free} values of 20.7%/25.4%. A second native crystal form yielded 2.1-Å-resolution data, and this model was refined to R/R_{free} values of 17.9%/21.2%. The SeMet and native proteins crystallized in different space groups with one and four molecules in the asymmetric unit, respectively (Table 2). Surprisingly, we found that the Spt6 C-terminal region comprises not one SH2 domain as anticipated,[20,28] but two SH2 folds that are packed tightly against each other to form a tandem SH2 domain composed of N-terminal (nSH2, residues 1250–1353) and C-terminal (cSH2, residues 1354–1440) folds. While this manuscript was in the final stages of preparation, equivalent tSH2 structures that overlap closely with our refined model were reported for Spt6 homologs from *Candida glabrata*[22] (RMSD = 1.0 Å over 184 C^{α} atoms with 87% sequence identity) and *Antonospora locustae*[23] (RMSD = 1.6 Å over 164 C^{α} atoms with 24% sequence identity).

N-terminal region

The first ~300 amino acids of Spt6 are extremely acidic, with an overall charge of −62 and a predicted p*I* of 4.3, and are also predicted to be disordered.[29] This is consistent with our Spt6(236–1259) and Spt6 (239–1451) structures, which lacked discernible density prior to residue 298. Despite the lack of inherent order, this region of Spt6 is functionally important. Residues 239–268 bind the essential Spn1/Iws1 protein and overlap with residues required for nucleosome binding.[14] Furthermore, the *spt6-1002* allele, a deletion of residues 2–122, displays synthetic lethality with deletion of the gene encoding the transcription factor Paf1.[18]

Helix–turn–helix domain

Residues 336–442 resemble a DNA-binding helix–turn–helix (HtH) motif, as seen in transcription factors such as c-Myb and the bacterial sigma factors.[30] Nevertheless, the structure is not consistent with binding DNA in a canonical manner; binding of Spt6 H7 in the major groove of DNA in the manner predicted for a canonical HtH domain would cause steric clashes between bound DNA and the rest of the structure. It is possible that Spt6 undergoes conformational changes upon binding DNA or that the Spt6 HtH domain serves as a protein–protein interaction motif, as occurs with members of the PWI subgroup of HtH domains.[31] The Spt6 HtH overlaps with the U4/U6 ribonucleoprotein Prp3 PWI domain [RMSD of 2.7 Å, 69 C^{α}, Protein Data Bank (PDB) code 1x4q] and the Nab2 PWI domain (RMSD of 2.9 Å, 73 C^{α}, PDB code 2v75), conserves the eponymous PWI motif as NWI (Asn349-Trp350-Ile351), and could utilize the equivalent protein-binding surface without invoking a conformational change in Spt6.

YqgF homologous domain

The Spt6 YqgF domain (residues 735–887) resembles members of the YqgFc superfamily, such as the *E. coli* protein YqgF and the RuvC class of Holliday junction resolvases.[25] The alignment is especially close with *E. coli* RuvC (RMSD of 2.9 Å, 117 C^{α}, PDB code 1hjr). Despite this similarity, the putative Spt6 YqgF catalytic site lacks carboxylate side chains that are critical for coordinating magnesium ions

Table 2. Data collection and refinement statistics

	Se-Spt6 (236–1259) peak	Se-Spt6 (236–1259) inflection	Spt6 (236–1259)	Spt6 (239–1451)	Se-Spt6 (1247–1451) peak	Spt6 (1247–1451)
Data collection						
Space group	$P2_12_12_1$		$P2_12_12_1$	$P3_121$	$I4_122$	$P2_12_12_1$
Unit cell dimensions (Å)	$a=114.0$		$a=115.1$	$a=118.7$	$a=97.5$	$a=78.8$
	$b=116.4$		$b=116.2$	$b=118.7$	$b=97.5$	$b=105.2$
	$c=122.7$		$c=117.4$	$c=214.4$	$c=132.4$	$c=119.5$
Molecules per asymmetric unit	1		1	1	1	4
Solvent content (%)	63.7		62.4	60.0	63.5	53.1
Beamline[a]	SSRL 11-1	SSRL 11-1	SSRL 7-1	NSLS X29	SSRL 9-1	Home
Wavelength (Å)	0.97886	0.97922	0.97773	1.10000	0.97908	1.54178
Resolution (Å)	50–2.7	50–2.7	32–2.6	46–3.3	35–2.7	30–2.1
High-resolution shell (Å)	(2.8–2.7)	(2.8–2.7)	(2.7–2.6)	(3.42–3.3)	(2.8–2.7)	(2.18–2.1)
No. of unique reflections	83,751	82,946	49,595	27,025	8481	58,489
No. of total reflections	261,886	257,334	434,557	182,336	649,419	258,498
Mean I/σ_I	22.1 (2.4)	22.1 (2.1)	29.8 (4.3)	39.8 (3.3)	25.1 (3.7)	16.8 (3.2)
Completeness (%)	95.8 (75.8)	94.7 (70.1)	99.8 (99.9)	99.5 (99.9)	92.2 (74.5)	99.5 (99.2)
R_{sym} (%)[b]	4.8 (35.7)	4.7 (38.3)	6.5 (43.9)	5.8 (56.9)	7.0 (29.6)	6.6 (47.8)
Refinement						
R_{cryst}[c]/R_{free}[d] (%)			22.4/26.5	26.5/30.8	20.7/25.4	17.9/21.2
No. of non-H atoms						
Protein			6788	6876	1575	6988
Solvent			48	0	19	491
$\langle B \rangle$ (Å^2)			103.7	167.6	74.7	44.2
RMSD bond lengths (Å)			0.006	0.013	0.008	0.004
RMSD bond angles (°)			0.895	1.59	1.12	0.757
Ramachandran outliers (%)			0.0	0.8	0.0	0.40
Ramachandran favored (%)			95.6	93.6	97.9	97.5
Rotamer outliers (%)			0.94	0.3	0.6	1.4

Values in parentheses correspond to the high-resolution shell.
Refinement statistics were determined by PHENIX[26] and MolProbity.[27]
[a] Data were collected at the SSRL, the NSLS, or on a Rigaku MicroMax-007HF rotating anode X-ray generator with a copper anode and VariMax confocal optics and a Rigaku R-AXIS IV image plate detector (home).
[b] $R_{sym}=(\sum|(I-\langle I\rangle)|)/(\sum I)$, where $\langle I\rangle$ is the average intensity of multiple measurements.
[c] $R_{cryst}=(\sum|F_{obs}-F_{calc}|)/(\sum|F_{obs}|)$.
[d] R_{free} is the R_{cryst} based on ~1000 (at least 10%) of the reflections that were excluded from refinement.

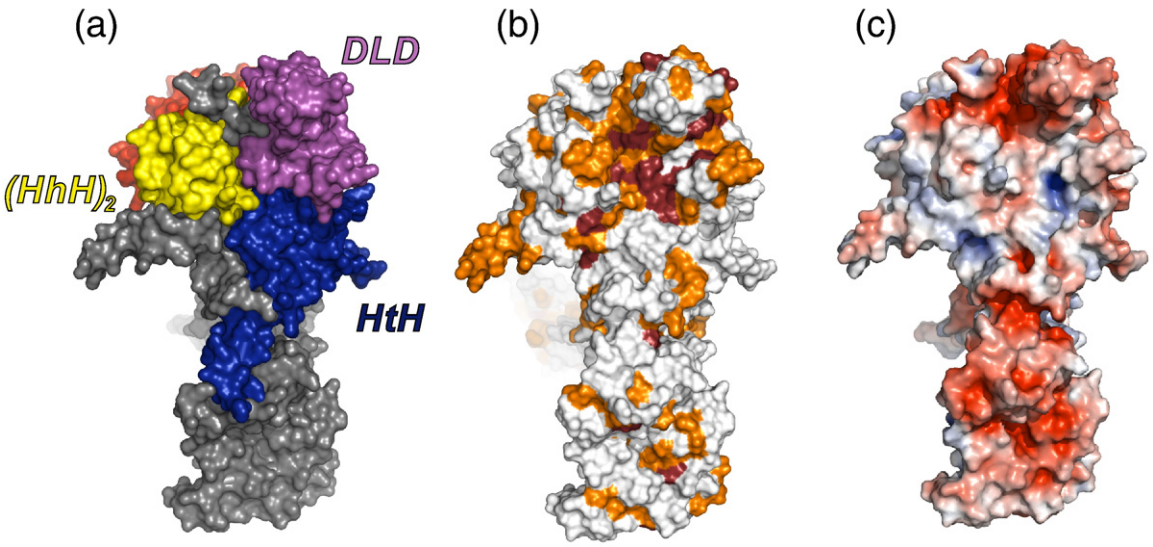


Fig. 4. The most conserved surface of the Spt6 core. (a) View of the Spt6 core showing the interface between HtH, DLD, and $(HhH)_2$ domains. (b) Same orientation as (a) but colored by conservation to illustrate the high level of surface conservation at the intersection of these domains, especially on the DLD. Coloring represents degree of conservation as described in Fig. 3 and Supplemental Fig. S1. (c) Same as (b) but colored by electrostatic potential (−5 to +5 kT/e).

that mediate phosphodiester bond hydrolytic cleavage.[32] Thus, it does not appear that the Spt6 YqgF fold is capable of nuclease activity using a catalytic mechanism similar to that of RuvC or related RNase H-fold nucleases.

Helix–hairpin–helix domain

Residues 933–1002 form two consecutive helix–hairpin–helix (HhH) motifs that pack together through highly conserved hydrophobic residues at an ~90° angle to form a $(HhH)_2$ domain that resembles known double-stranded DNA (dsDNA) binding domains.[33] These include proteins such as *E. coli* RNA polymerase α CTD (RMSD of 2.3 Å, 56 C^{α}, PDB code 1lb2) and the Holliday junction-binding protein RuvA (RMSD of 2.2 Å, 60 C^{α}, PDB code 1bvs). The first Spt6 HhH represents a characteristic HhH motif in the relative angle of the antiparallel helices and the presence of the Gly-hydrophobic-Gly motif within the hairpin loop.[33] The second HhH motif is more variant, as also observed in other $(HhH)_2$ domains, including DNA polymerase β and 5′ to 3′ exonucleases. Though

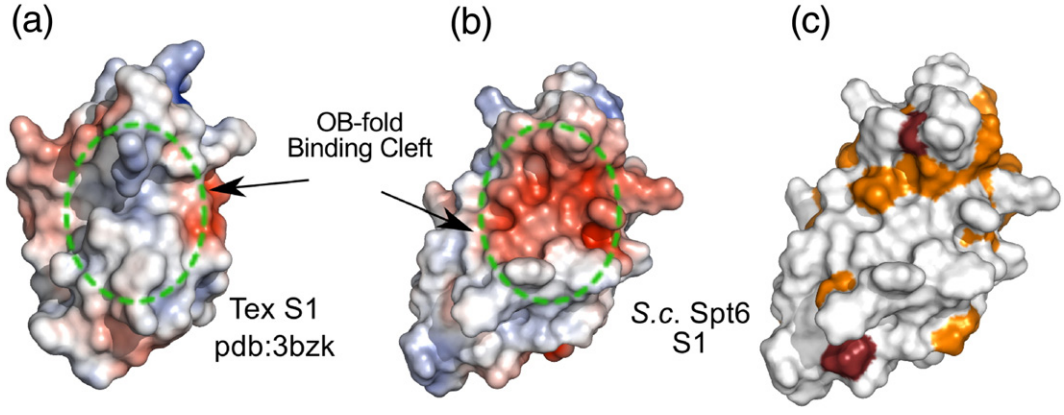


Fig. 5. Surface representations of the Spt6 and Tex S1 domains. (a) Electrostatic surface representation (−5 to +5 kT/e) of the S1 domain from *Pseudomonas aeruginosa* Tex (PDB code 3bzk) with the position of the nucleic acid binding OB-fold cleft approximated by the circle with the dotted green line. (b) Electrostatic surface representation (−5 to +5 kT/e) of the *S. cerevisiae* Spt6 S1 domain in the same orientation as in (a) showing a clustering of negative charge in the putative OB-fold binding cleft. (c) *S. cerevisiae* Spt6 S1 domain in the same orientation as (b) but colored by conservation in the same color scheme as in Fig. 3 to illustrate the low level of conservation within the region equivalent to the binding cleft of canonical S1 domains.

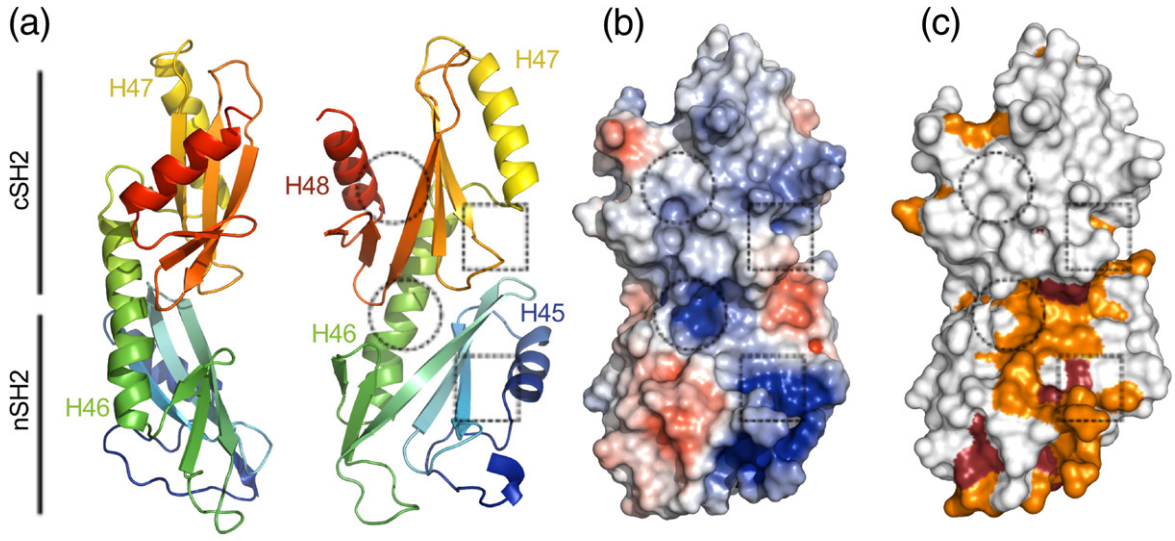


Fig. 6. The *S. cerevisiae* Spt6 tSH2 domain. (a) Two views of a cartoon representation of the tSH2 domain colored from the N- to C-terminus (blue to red; residues 1247–1440 are shown). Dotted squares and circles show the approximate positions of the canonical SH2 domain pTyr and specificity pockets, respectively. (b) Surface representation colored by electrostatic potential surface (−5 to +5 kT/e). (c) Same as (b) but colored by residue conservation in the same color scheme as in Figs. 3 and 4.

$(HhH)_2$ folds are primarily found in proteins that interact with DNA, they also occur in proteins that mediate protein–protein interactions, such as the sterile α motif proteins.[34] Notably, the first 28 ordered residues (~298–325) of Spt6 wrap around the $(HhH)_2$ domain in a fashion that would occlude binding of a canonical $(HhH)_2$ domain to a dsDNA ligand, although this interaction could be transient. The absence of corresponding density for H1 and part of H2 in the initial SeMet Spt6(236–1259) and Spt6(239–1451) maps suggests that these N-terminal residues might adopt a different conformation to allow binding of a physiological partner *in vivo*.

Death-like domain

Residues 1019–1104 form a prominent lobe of the structure that resembles members of the death domain superfamily. Death domains typically serve as recognition modules in proteins that assemble and activate inflammatory and apoptotic complexes.[35] The Spt6 death-like domain (DLD) maintains the characteristic overall topology of death domains, consisting of a six-helix bundle with three stacked antiparallel helices but with an additional helix inserted between the final two helices of the bundle (H39 in Figs. 2 and 3). Spt6 aligns reasonably well with several known death domain superfamily proteins, including the caspase-2-activating PIDDosome PIDD protein subunit component (RMSD of 3.0 Å, 60 C^{α}, PDB code 2of5). Although it is unlikely that the Spt6 DLD functions in an apoptotic process in yeast, its prominent location and the observation that it displays the most highly conserved region of the Spt6 surface suggest that it mediates important intermolecular interactions (Fig. 4).

S1 domain

A mostly unstructured linker of 15 residues leads to the S1 domain (residues 1129–1219), which adopts the canonical S1/oligonucleotide–oligosaccharide binding (OB)-fold of a β-barrel composed of two three-stranded β-sheets where strand 1 (S10) is shared by both sheets.[36] Despite the structural similarity, Spt6 lacks the typical S1 binding cleft residues that are important for binding nucleic acids.[36] In addition, the predicted electrostatic potential surface does not appear conducive to nucleic acid binding, shows a low level of conservation, and, as discussed below, is not required for dsDNA binding (Fig. 5). This is in contrast to the distantly related bacterial Tex protein, which loses its capacity to bind DNA or RNA in the absence of the S1 domain.[21] OB folds are used to bind partners other than nucleic acids, including oligosaccharides and proteins;[36,37] thus, it remains possible that the Spt6 S1 domain is used for an important interaction that does not involve nucleic acids.

Tandem SH2 domain

The S1 domain is followed by an unstructured ~10-residue segment and an ~30-Å helix (H44; 1227–1247) that buries ~440 $Å^2$ of accessible surface area against the core in the Spt6(236–1259) structure (Fig. 2). While some density is present for H44 in the

Spt6(239–1451) maps, this region is too disordered for reliable model building, which may indicate that this interface is not always formed in solution. H44 links the core to a tandem SH2 domain (tSH2; residues 1250–1440) that comprises N-terminal (nSH2; residues 1250–1353) and C-terminal (cSH2; 1353–1440) folds that associate through an ~800-Å^2 interface to form a single structural unit (Fig. 6). Both nSH2 and cSH2 conform to the standard SH2 domain fold (standard SH2 nomenclature in parenthesis) of an N-terminal helix (αA), a central three-stranded β-sheet (βB–βD), a small two-strand extension to the β-sheet (βE–βF), and a second helix (αB).[38] Intervening loops are labeled based on their relative position between these elements (e.g., the BC loop connects the βB and βC strands). The Spt6 nSH2 and cSH2 superimpose well with each other (RMSD of 2.1 Å, 69 C$^\alpha$; Fig. 7) and with the multitude of other characterized SH2 domains, such as those from v-Src kinase (nSH2: RMSD of 2.0 Å, 82 C$^\alpha$; cSH2: RMSD of 2.1 Å, 80 C$^\alpha$, PDB code 1sps; Fig. 7) and Nck2 (nSH2: RMSD of 2.1 Å, 83 C$^\alpha$; cSH2: RMSD of 1.9 Å, 73 C$^\alpha$, PDB code 2cia).

The relative orientation of the two SH2 folds that comprise the Spt6 tSH2 domain is unlike that of previously reported tandem SH2 domains from

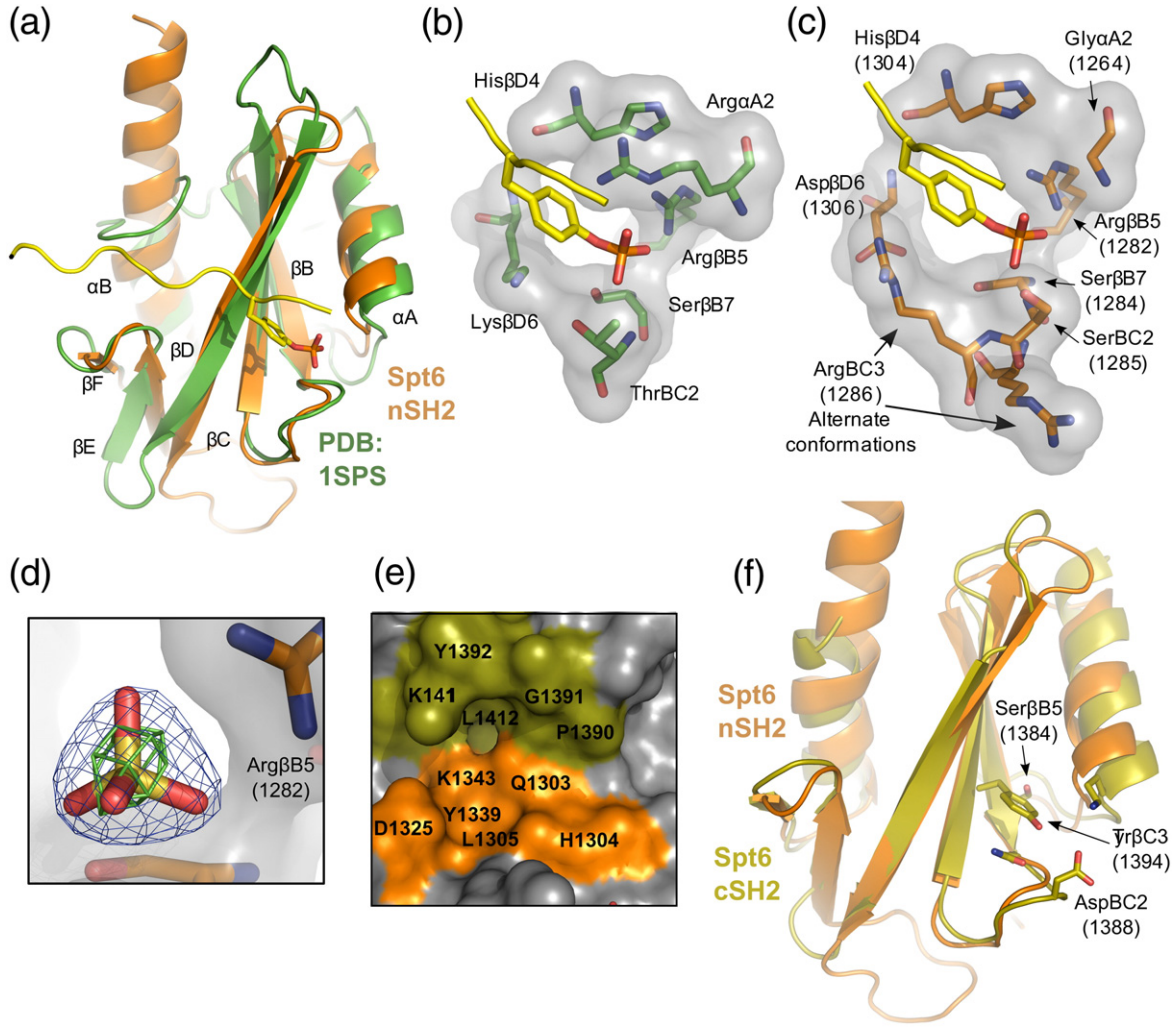


Fig. 7. tSH2 binding pockets. (a) Overlay of Spt6 nSH2 (orange) and the v-Src Kinase SH2 domain (green; PDB code 1sps) bound to a pTyr ligand (yellow). Secondary structure elements are labeled based on the standard SH2 nomenclature.[38] (b) Detailed view of the pTyr binding pocket of 1sps. Residues contributing to the coordination of the pTyr ligand are shown. (c) Same as (b) but for the Spt6 nSH2, with the pTyr peptide from 1sps positioned after the overlap on the SH2 protein domains. (d) Electron density for a sulfate bound in the tSH2 crystal structures. Blue density represents the $2mF_o - DF_c$ map contoured at 2.0 σ, and green density represents an anomalous difference Fourier map contoured at 3.0 σ. (e) The putative nSH2 specificity pocket. Residues from both nSH2 (orange) and cSH2 (olive) line the nSH2 specificity pocket. (f) Alignment of the Spt6 nSH2 (orange) and cSH2 (olive) folds. Residues that protrude into the typical location of the pTyr binding pocket of the cSH2 fold are shown.

other proteins. The αB helix of nSH2 undergoes an ~20° kink where the αB helix would end in a canonical SH2 domain and extends along the backside of the cSH2 fold to form an extensive hydrophobic packing interface with the cSH2 central β-sheet. The relative orientation of nSH2 and cSH2 folds therefore appears to be constrained, consistent with the observation that superposition of the five crystallographically independent tSH2 domains observed in our various crystal forms [one in SeMet Spt6(1247–1451) and four in native Spt6(1247–1451)] indicates a maximum relative rotation of ~8° between nSH2 and cSH2 folds.

Implications for ligand binding by the tSH2 domain

The extent to which Spt6 nSH2 corresponds to a prototypic SH2 domain is seen in a comparison with v-Src (PDB code 1sps) (Fig. 7a–c). The primary determinants of phosphate binding are preserved in nSH2, whereas the positively charged Arg/Lys side chains of classical SH2 domains that flank the aromatic ring of phosphotyrosine (pTyr) ligands are absent from their usual positions. Conserved residues include the consensus FLVRES (FVIRQS, 1279–1284 in Spt6) sequence motif that contributes the phosphate-coordinating Arg1282 and Ser1284 side chains (Spt6 numbering). Moreover, the following Ser1285 Spt6 side chain is also well positioned to hydrogen bond the phosphate and functionally substitute the ThrBC2 of the Src SH2 domain. Other positions within SH2 domains that are important for binding phosphate include HisβD4 (Spt6 H1304) and Ser/ThrβC4 (Spt6 Thr1294), whose side chains hydrogen bond the Arg1282 side chain in an optimum orientation for phosphate binding and, in cognate SH2–ligand complexes, also form a main chain to main chain hydrogen bond with the ligand residue following the pTyr.

Classical SH2 domains typically have a basic residue at the αA2 position that binds against one side of the tyrosine ligand aromatic ring where it forms an amino–aromatic interaction with the π ring and also hydrogen bonds with both the phosphate and the tyrosine main-chain carbonyl. In contrast, the Spt6 nSH2 has a glycine at this position, G1264, which can make none of the same ligand interactions. This does not argue strongly against binding of pTyr by Spt6, however, because a number of other SH2 domains that bind pTyr ligands have a variety of substitutions at this position, including the PTPN11/SHPTP2/Syp phosphatase (PDB code 1ayc) that, like Spt6, has a Gly at αA2 and is known to bind a pTyr-containing peptide.[39] On the other side of a canonical pTyr ligand side chain, classical SH2 domains typically have a basic residue in the βD6 position. This residue is a lysine in Src but is an aspartate (D1306) in Spt6. Interestingly, in three of the five crystallographically independent Spt6 tSH2 molecules in our structures, the space typically occupied by the basic βD6 side chain is filled by R1286 in the BC3 position (Fig. 7c). Thus, Spt6 retains the ability to provide a positively charged basic group in this position, consistent with the potential to bind pTyr.

Ammonium sulfate was present in the crystallization solutions for both native and SeMet tSH2 domain structures, and all five crystallographically independent molecules displayed a sulfate ion at the putative nSH2 phosphate binding site, where it forms hydrogen bonds with R1282, S1284, and S1285 in the same manner as the phosphate of pTyr–SH2 complexes (Fig. 7d). Assignment of the density as sulfate was confirmed in anomalous difference Fourier maps for the native data, which showed peaks that were similar in size to those of cysteine and methionine sulfur atoms for some of the sulfates. This further suggests that nSH2 binds a phosphorylated ligand and that it might accommodate a pTyr side chain in a suboptimal binding pocket.

Typical SH2 domain ligands bind through a two-prong mechanism that, in addition to binding pTyr, also involves binding of the three side-chain residues C-terminal to the pTyr into the "specificity pocket."[38] Binding partner preference is typically defined in the specificity pocket by the BG and EF loops and the βD3 and βD5 residues, which usually favor binding of hydrophobic residues. In contrast, the Spt6 nSH2 fold predominantly displays charged and polar residues in this site, and there is no BG loop due to the extension of nSH2 αB to the cSH2 fold. Instead, cSH2 residues such as the DE loop and a βD side chain (K1411) protrude into the pocket, forming the top portion of the nSH2 specificity pocket (Fig. 7e). This indicates that, if the nSH2 fold binds substrate in the typical "two-pronged" manner common to SH2 domains, the cSH2 fold would make significant contributions to binding.

In contrast to nSH2, the cSH2 fold appears to be cryptic and unlikely to bind a phosphorylated ligand because residues critical for phosphate binding are substituted to display a very different chemical environment, and the Y1394 side chain fills the space where a phosphate would typically bind (Fig. 7f). Moreover, the region of the specificity pocket lacks even a shallow depression, as it is filled by bulky, aromatic side chains (F1397, Y1406, W1408, and F1434). Thus, in contrast to nSH2 and consistent with the lack of sequence conservation at cSH2 (Fig. 6c), it seems unlikely that cSH2 binds ligands in a manner reminiscent of SH2 domains.

Binding of Spt6 tSH2 domain with phosphorylated peptides

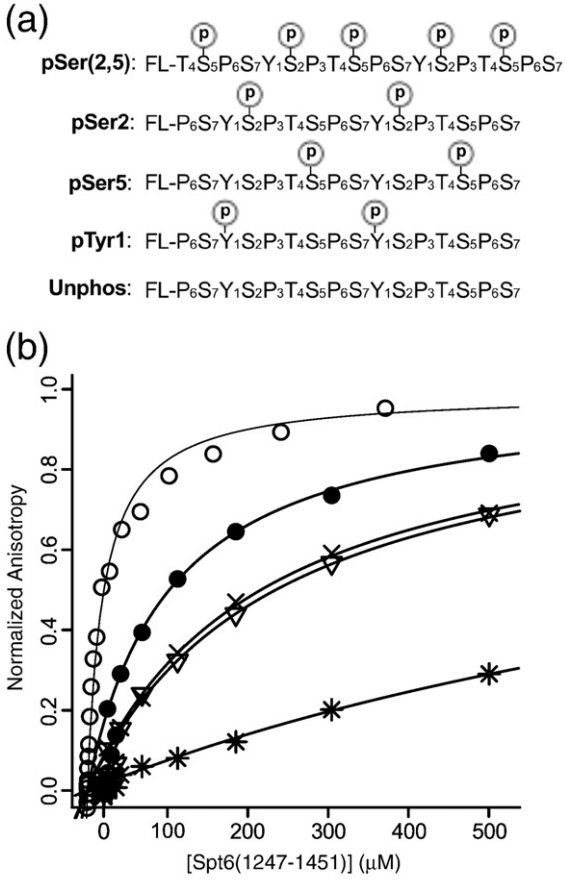
The human Spt6 CTD has been reported to bind the heptad repeat sequences of the mammalian

RNAPII large subunit when the RNAPII CTD is treated with P-TEFb, a kinase that phosphorylates Ser2 during transcriptional elongation.[9] In order to investigate this interaction more quantitatively, we used fluorescence anisotropy (FA) to measure binding of *S. cerevisiae* Spt6(1247–1451) to di-heptad repeat peptides representing various phosphoisoforms of the RNAPII CTD (Fig. 8). Peptides tested include sequences representing phosphoserine (pSer) 2, pSer5, pTyr1, and pSer(2,5). pSer5 and pSer(2,5) peptides were included to test specificity and because these modifications also occur on the RNAPII heptad repeats. A pTyr1 peptide was included because SH2 domains typically bind pTyr peptides and this modification occurs in mammals,[40] although it has not been reported to occur in yeast.

All of the peptides assayed bound with affinities in the range of ~20–250 μM, which is similar to the affinity of other RNAPII CTD interactions with isolated binding domains,[41] but is ~10- to 100-fold weaker than is typically found for the interaction of SH2 domains with pTyr ligands.[42] Of the ligands we assayed, the pSer(2,5) peptide bound Spt6(1247–1451) with the highest affinity (23 μM). This may indicate that RNAPII CTD sequences phosphorylated on both Ser2 and Ser5 are the authentic *in vivo* ligands for the yeast Spt6 tSH2 domain, which would be consistent with the reports that Ser2-phosphorylated sequences are preferred[9] and that localization of Ser2 and Ser2/5 phosphorylation overlaps substantially in average transcription units.[43,44] On the other hand, interpretation is complicated by the fact that our pSer2/5 peptide has considerably more negative charge than the other peptides assayed and thus may be more prone to nonspecific effects. Interestingly, the pTyr1 peptide binds with a K_d of 110 μM, which is tighter than that of the pSer2 (197 μM) and pSer5 (245 μM) peptides bearing an equivalent number of phosphate groups. A higher affinity for pTyr1 peptide with the same overall charge as pSer2 or pSer5 peptide is consistent with the similarity between nSH2 and well-characterized pTyr binding SH2 domains, but the physiological relevance of this result is not clear given the lack of observed pTyr modification of the RNAPII CTD in yeast.

The conclusion that the Spt6 tSH2 domain interacts specifically with phospho-CTD peptides is reinforced by our observations that an unphosphorylated form of the RNAPII CTD showed negligible (K_d of >1000 μM) binding and that similar results were obtained when the assays were performed under a phosphate-buffered saline condition (data not shown). As a further test of specific interactions, we measured binding to a mutant form of Spt6(1247–1451) in which residues R1282 and S1284, which are important for phosphate binding in typical SH2 domains, were both substituted with alanine. Binding to the pSer(2,5) peptide was decreased by ~4-fold, while binding to the pTyr1, pSer2, and pSer5 peptides was decreased by 2.4-fold, 1.2-fold, and 1.4-fold, respectively. These modest effects are consistent with the putative nSH2 phosphate binding site contributing to the binding interaction but not performing a dominant role for interaction with the peptides assayed.

Deletion of the entire tSH2 region by truncation of the *SPT6* gene leads to defects in growth attributed



(c)

CTD peptide	K_D (μM)† WT	R1282A/S1284A
○ pSer(2,5)	23.0±2.2	90.9±14.2
● pTyr1	110.2±9.1	261.6±32.8
× pSer2	197.1±39.69	228.4±17.8
▽ pSer5	245.4±29.4	348.1±56.7
✳ Unphos	>1000	>1000

†Average K_D ± standard error for at least 3 independent measurements

Fig. 8. Spt6 tSH2 binds RNAPII CTD phosphopeptides. (a) Peptides used in binding studies with positions of pSer and pTyr residues indicated. (b) Representative FA binding isotherms for Spt6(1247–1451) binding to various peptides with symbols defined in (c). (c) Binding affinities for WT and R1282A/S1284A Spt6(1247–1451) proteins based on FA experiments.

to suboptimal transcription elongation.[22,23] To examine the importance of tSH2 residues implicated in pTyr binding *in vivo* more carefully, we mutated the single genomic copy of *SPT6* to produce proteins with R1282H, S1284D, R1286A, Q1303E, EN1313/1314AA, or K1343E mutation (nSH2 domain) or P1390A or K1411E mutation (cSH2 domain). The effects of these mutations were quite mild, failing to recapitulate the severe defects caused by truncation of the gene (data not shown; see Materials and Methods for a list of phenotypes screened). The C-terminal region of Spt6 therefore appears to have some activity that is not interrupted when residues within the tSH2 domain expected to be important for binding phosphorylated substrates are mutated.

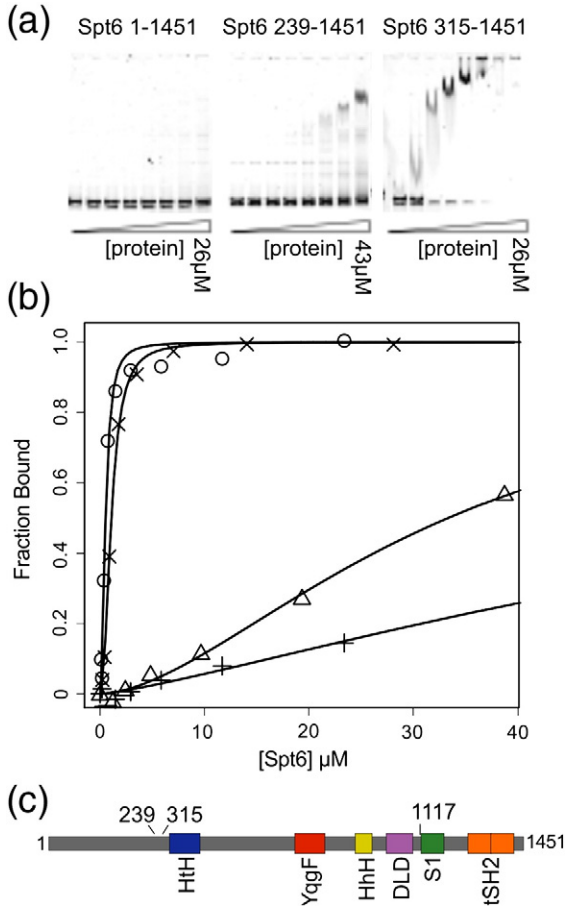


Fig. 9. dsDNA binding studies. (a) Representative gel shift assays for three different constructs of Spt6. (b) Representative binding isotherms used to calculate dissociation constants for various Spt6 constructs binding to the 177-bp Widom 601 dsDNA.[45] Symbols used to indicate isotherms for different constructs are as follows: (+) Spt6 (1–1451), $K_d = 106.7 \pm 38.9$ μM; (Δ) Spt6(239–1451), $K_d = 33.7 \pm 3.8$ μM; (×) Spt6(315–1117), $K_d = 1.08 \pm 0.06$ μM; (○) Spt6(315–1451), $K_d = 0.53 \pm 0.07$ μM. (c) Schematic diagram indicating endpoints for constructs used in DNA binding studies.

Our data are consistent with a recent report that concluded that pSer2 RNAPII CTD peptides bound the tSH2 domain of the *C. glabrata* Spt6 homolog with an affinity of 10 μM.[22] The tighter affinity observed in that study could reflect differences between the proteins but is more likely to be due to the very low (10 mM) concentration of NaCl used in the binding assays. Our data also extend the earlier work by showing that the *S. cerevisiae* Spt6 tSH2 domain displays little discrimination for binding CTD peptides with different modifications but does show a small preference for a peptide with a single phosphorylated tyrosine (pTyr1). Our findings suggest that Spt6 activities *in vivo* may be modulated by phosphorylation of binding partners and that the ligands of the tSH2 domain may include phosphorylated tyrosine residues.

Binding to dsDNA

To test whether Spt6 is capable of binding dsDNA, we performed electrophoretic mobility gel shift assays using a 177-bp dsDNA fragment. We tested several different Spt6 constructs and found that binding was tighter ($K_d \pm$ standard deviation) for Spt6(315–1451) (K_d of 0.53 ± 0.07 μM) than for Spt6 (239–1451) (K_d of 33.7 ± 3.8 μM) or Spt6(1–1451) (K_d of 106.7 ± 38.9 μM) (Fig. 9), demonstrating that the disordered and negatively charged N-terminal residues diminish dsDNA binding. Unlike Tex, which requires the S1 domain for nucleic acid binding,[21] an Spt6 construct lacking the S1 domain, Spt6(315–1117), retains the ability to bind dsDNA with a K_d of 1.08 ± 0.06 μM (Fig. 9b).

Overall structure and functional implications

Our composite model of Spt6 (Fig. 10) features a core region (residues ~298–1117) that has multiple recognizable domains whose packing in the crystal likely reflects, to a large extent, their organization in solution, at least in the absence of binding partners. The N-terminal residues 1–297 display considerable overall negative charge and are expected to be highly mobile, while residues 239–263 also comprise the Spn1/Iws1-binding determinant and overlap with the nucleosome binding site.[14] This high degree of mobility may provide a flexible tether for bridging binding partners, such as Spn1/IwsI, RNAPII, and nucleosomes,[9] and the negative charge may modulate histone and DNA interactions.

Inherent flexibility is also a feature of the C-terminal S1 domain, H44, and tSH2 domain. The S1 domain is loosely associated with the core, lacks density in the Spt6(236–1259) structure, and may be visible in the Spt6(239–1451) structure only because of ordering by a crystal lattice contact. Whereas the core of the distantly related Tex protein clearly resembles that of Spt6, the Tex and Spt6 S1 domains are displaced

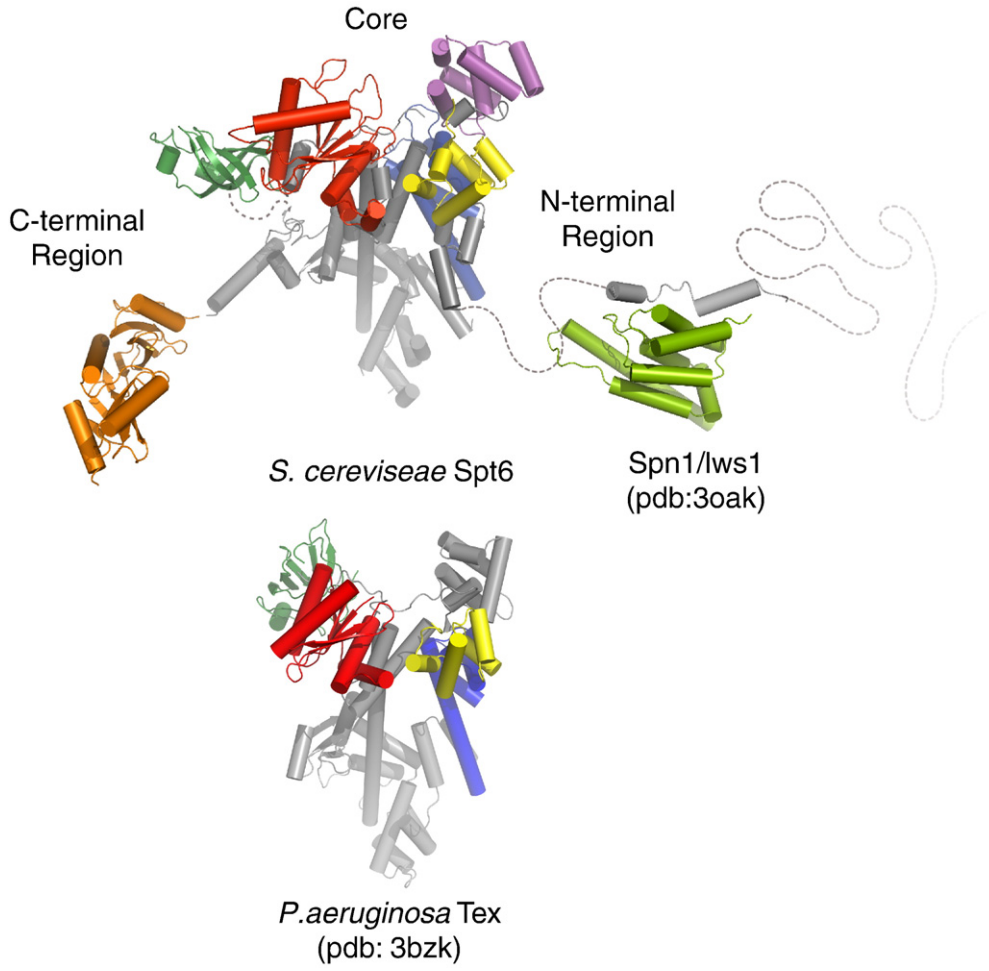


Fig. 10. Comparison of Spt6 and Tex structures. The overall structure of the Spt6 core resembles that of the prokaryotic Tex protein. This similarity implicates the Spt6 core in nucleosome-independent functions, such as transcriptional elongation on naked template DNA. The N-terminal region includes determinants essential for binding Spn1/Iws1 or nucleosomes, which appear to bind competitively with each other.[14] The negatively charged N-terminal region may also be important for modulating binding to nucleic acids. The C-terminal region has been implicated in binding RNAPII.[9]

by a rotation of ~80° and ~25- to 30-Å translation with respect to each other, and the S1 domain appears to be shifted by 14 Å in different crystal forms of Tex.[21] Therefore, it is likely that a highly mobile S1 domain is an important feature of both proteins, although one notable difference is that the Tex S1 binds nucleic acids, whereas our DNA binding data and consideration of surface amino acid residues indicate that the Spt6 S1 domain does not. Although the low conservation of residues in the putative binding cleft of the S1 domain suggests that this region is unlikely to have a highly conserved binding partner, the electrostatic characteristics of this surface (Fig. 5) are consistent with potential binding partners that are positively charged, such as histones.

H44 is visible only in the Spt6(236–1259) structure, where it also appears to be ordered by a lattice contact, and the C-terminal tSH2 domain is expected to be highly dynamic with respect to the rest of the protein. The leading model is that this domain binds RNAPII that is phosphorylated on Ser2 of its CTD.[9] Our binding data are consistent with this view, provided that other determinants contribute to binding/specificity, but also indicate the possibility that other binding partners might be functionally important. For example, Spt5 (discussed below) has been shown to co-localize with Spt6 and contains a phosphorylated C-terminal repeat domain similar to the RNAPII CTD,[46] which could be a physiological ligand for the Spt6 tSH2 domain.

Spt6 has been shown to be functionally associated with a large number of proteins involved in transcription elongation, chromatin maintenance, and RNA processing. The structure of Spt6 presented here will serve as a foundation for a more precise mapping of protein binding partners. Along with protein binding partners, Spt6 is also expected

to interact with nucleic acids in the transcription complex. DNA binding is probably important for nucleosome reassembly and potentially for transcription elongation. Association with nascent RNA transcripts could be important for enhancing the elongation rate or for organizing interactions with RNA modification/export factors such as REF/Aly. Future studies will be needed to see if Spt6 binds RNA and to further map nucleic acid binding to different domains of Spt6. For example, the YqgF or the $(HhH)_2$ domain may mediate binding to specific DNA structures such as four-way (Holliday) junction DNA, structures similar to those found at the DNA entry/exit points of nucleosomes.[47] The tSH2 domain may also contribute to nucleic acid interactions, as it is likely to bind negatively charged substrates containing phosphate groups. A simple electrostatic surface analysis indicates that each of these domains retains properties found in homologous domains with known functions. Furthermore, examination of the histone-binding activity of the various Spt6 domains will be of significant interest in furthering our understanding of nucleosome assembly/reassembly.

The structural similarity between Spt6 and the prokaryotic Tex protein is limited to the core and S1 domains. Consistent with the extent of structural similarity, N- and C-terminal regions that are unique to Spt6 are required for eukaryotic-specific interactions with nucleosomes,[11,13,14] hyperphosphorylated forms of the RNAPII CTD,[9] Spn1/IwsI,[14] and mRNA processing/export factors.[9] An attractive possibility is that the core region provides activities that are conserved among prokaryotes and eukaryotes. In this regard, it is striking that most of the Spt6 core domains belong to structural families whose members function in nucleic acid binding, an activity that is likely to be a key component of transcription factors such as Spt6 and Tex that are capable of stimulating elongation on nucleosome-free DNA templates.[9,16] Consistent with this possibility, our data indicate that the Spt6 core can bind dsDNA. Curiously, some of the putative nucleic-acid-binding surfaces of Spt6 domains are occluded in the structure, although conformational changes might displace residues 298–320 (H1 and H2) to expose a DNA-binding activity on the $(HhH)_2$ domain. Consistent with this model, we find that truncation of ~314 N-terminal residues leads to tighter dsDNA binding. One attractive model is that conformational changes of this nature are induced by binding partners such as histones or Spn1/Iws1. The relationship between Spt6 and Tex proposed here is reminiscent of that between another eukaryotic transcription elongation factor, Spt5, and its bacterial counterpart, NusG. These proteins also display similar core domains, while Spt5 has an acidic N-terminal extension and a C-terminal extension[48] that confers eukaryote-specific functions such as binding to RNAPII,[48] interaction with mRNA capping enzymes,[49] and extensive phosphorylation by RNAPII CTD kinases such as P-TEFb.[50] Therefore, like Spt5, Spt6 is likely to have built on the fundamental transcription activities of its core to accommodate the additional complexities of eukaryotic gene regulation.

Materials and Methods

Protein expression and purification

The protein constructs were expressed from pET151-D/TOPO vectors (Invitrogen) in BL21 codon plus (RIL) *E. coli* cells (Stratagene). Cultures were grown in autoinduction media[51] in baffled 1.8-l flasks at 37 °C with continuous shaking. After 4–8 h, the cultures were shifted to 23 °C and grown for an additional 16–24 h. Harvested cells were stored at −80 °C. Cells were thawed and lysed in buffer containing lysozyme and protease inhibitors, followed by sonication and centrifugation (25,000–30,000*g*). The soluble fraction was applied to nickel agarose resin (Qiagen) and eluted in buffer containing 300 mM imidazole and 100 mM NaCl, immediately followed by application to a heparin column (5-ml HiTrap Heparin; GE Healthcare Life Sciences) and elution over a NaCl gradient. Fractions containing Spt6 were pooled and processed overnight at room temperature in buffer containing tobacco etch virus protease. A nickel agarose column was used to remove the tagged tobacco etch virus protease and unprocessed Spt6 protein, and the flow-through was concentrated and loaded onto a size-exclusion column [Superdex 200 or S75 (for 1247–1451 constructs); GE Healthcare Life Sciences]. SeMet protein was expressed using an autoinduction protocol[51] for selenomethionine incorporation and purified by the same protocol as native protein. All crystals were grown in sitting drops, transferred briefly to a cryoprotection solution, suspended in a nylon loop, and plunged into liquid nitrogen (Table 1).

Crystal structure determinations and refinements

Data were processed with HKL2000.[52] Spt6(236–1259) was determined by the multiple-wavelength anomalous diffraction method. SOLVE[53] was used to locate selenium atoms, and RESOLVE[54] was used for density modification and preliminary model building. AutoSol in PHENIX[26] was used to determine the Spt6(1247–1451) structure by the single-wavelength anomalous diffraction method. The Spt6 (239–1451) structure was determined by molecular replacement using AutoMR in PHENIX[26] to a resolution of 3.3 Å. A homology model built by TASSER[55] of the Spt6 S1 domain was used as a guide for model building. The native Spt6 (1247–1451) structure was determined by molecular replacement using Phaser.[56] PHENIX[26] and TLSMD[57] were used for refinement, Coot[58] was used for model building, and MolProbity[27] was used for structure validation. The following residues were ordered and included in the refined models: Spt6(236–1259) 297–455, 464–484, 501–561, 567–1002, 1009–1128, and 1219–1248; Spt6(239–1451) 312–455, 464–489, 509–552, 567–649, 653–1001, and 1014–1210; and

Spt6(1247–1451) 1247–1440. Structural alignments were performed using Dali[59] and SSM.[60] PyMOL[61] was used to create the figures. Electrostatic surface representations were calculated using PDB2PQR and APBS tools[62,63] using the AMBER force field and colored from red ($-5\ kT/e$) to blue ($+5\ kT/e$).

DNA binding experiments

Widom 601 DNA (177 bp) with a 5′ Cy3 fluorophore was generated as described[45] followed by precipitation, gel purification, and electroelution. Electrophoretic mobility shift binding experiments were performed, and K_d values were determined as described previously.[21] In short, 2-fold serial dilutions of the respective purified protein construct were mixed with nucleic acid substrate at room temperature in binding buffer [15 mM Tris–HCl (pH 7.5), 100 mM NaCl, 10% glycerol, and 0.5 mM ethylenediaminetetraacetic acid] where the final concentration of dsDNA was 10- to 20-fold below the estimated K_d for the interaction. After incubation for 30 min, samples were run on 4–20% TBE native gels (Bio-Rad Laboratories) and imaged and quantified using a TYPHOON imaging system with ImageQuant software (GE Healthcare Life Sciences). The fraction bound was calculated by quantifying the DNA_{total} (total fluorescence in entire lane) and DNA_{free}. DNA of slower mobility than the DNA_{free} was considered bound. The fraction bound $= 1 - ([DNA]_{free}/[DNA]_{total})$. Dissociation constants (K_d values) were calculated by plotting data points and curve fitting in the program R[64] using the Hill formalism where fraction bound $= 1/(1 + (K_d n/[P]n))$. In all cases, standard deviations are calculated from at least three measurements, except for the Spt6(315–1117) construct, which was repeated twice.

FA binding experiments

Peptides were synthesized by the University of Utah Core Facility or purchased commercially through AnaSpec Inc. (San Jose, CA), purified to >98% purity by HPLC, and confirmed by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Purified Spt6(1247–1451) was titrated in 1.5- to 2.0-fold serial dilutions against a constant concentration of fluorescein-labeled peptide (10- to 20-fold below estimated K_d) in 20 mM Tris–Cl (pH 7.5), 100 mM NaCl, and 5% glycerol. Samples were incubated at room temperature for at least 15 min prior to reading. Parallel and perpendicular fluorescence intensity was measured in a multi-well format using a Tecan Infinite 200 microplate reader using excitation/emission wavelengths of 485 nm/535 nm. Anisotropy values were calculated, normalized, and plotted as a function of protein concentration. K_d values were determined by fitting the data using the equation[65] $A = ((A_T \times ([pro]/K_d))/(1 + [pro]/K_d))$, where A is the measured anisotropy, A_T is the total change in anisotropy, and [pro] is the protein concentration.

Genetic analysis of tSH2 mutations

The following alleles of *SPT6* were screened for phenotypes in strains isogenic with the A364a genetic background: wild type (WT), *spt6-R1282H*, *spt6-S1284D*, *spt6-R1286A*, *spt6-Q1303E*, *spt6-E1313A*, *N1314A*, *spt6-K1343E*, *spt6-P1390A*, and *spt6-K1411E*. Mutations were introduced into the genomic copy of *SPT6* such that expression was from the native promoter at the normal locus except for the introduction of a *URA3* or *TRP1* marker downstream of the open reading frame. Strains were tested for growth on rich medium at 30 °C and 38 °C; on medium lacking lysine at 30 °C and 37 °C (all strains had the *lys2-128∂* allele; thus, growth would reveal an Spt^- phenotype); and on media containing 150 mM hydroxyurea, 75 μg/ml 6-azauracil, 0.6 μg/ml 4-nitroquinolone, 10 mM caffeine, 3% formamide, 1.2 M NaCl, 45 μg/ml mycophenolic acid, or 6% ethanol (all at 30 °C). None of the mutants were sensitive to any of the stress conditions relative to the WT strain. *spt6-S1284D* and *spt6-Q1303E* strains were somewhat more resistant to 3% formamide than the WT, and *spt6-Q1303E* strains displayed a very weak Spt^- phenotype (faint growth after 7 days). Mutants were not tested for a defect in cryptic initiation.

PDB accession numbers

Coordinates and structure factors for Spt6(236–1259), Spt6(239–1451), and SeMet Spt6(1247–1451) and native Spt6(1247–1451) have been deposited in the PDB with accession numbers 3psf, 3psi, 3psj, and 3psk, respectively.

Acknowledgements

We thank Hua Xin and Charisse Kettelkamp for technical assistance and Heidi Schubert for advice with the crystallographic analysis. Portions of this work were performed in Core Facilities at the University of Utah, which were supported by P30CA042014 from the National Cancer Institute. Some of the X-ray diffraction data for this study were measured at the National Synchrotron Light Source (NSLS). Financial support for NSLS comes principally from the Offices of Biological and Environmental Research and of Basic Energy Sciences of the U.S. Department of Energy and from the National Center for Research Resources of the National Institutes of Health (NIH). Portions of this research were performed at the Stanford Synchrotron Radiation Laboratory (SSRL), a national user facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The SSRL Structural Molecular Biology Program is supported by the Department of Energy, Office of Biological and Environmental Research, and the NIH, National Center for Research Resources, Biomedical Technology Program, and the National Institute of General Medical Sciences. S.J.J. was supported by a postdoctoral fellowship (GM074368). This work was supported by NIH grant RO1 GM076242.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2011.03.002

References

1. Li, B., Carey, M. & Workman, J. L. (2007). The role of chromatin during transcription. *Cell*, **128**, 707–719.
2. Luna, R., Gaillard, H., Gonzalez-Aguilera, C. & Aguilera, A. (2008). Biogenesis of mRNPs: integrating different processes in the eukaryotic nucleus. *Chromosoma*, **117**, 319–331.
3. Clark-Adams, C. D. & Winston, F. (1987). The *SPT6* gene is essential for growth and is required for delta-mediated transcription in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **7**, 679–686.
4. Kok, F. O., Oster, E., Mentzer, L., Hsieh, J. C., Henry, C. A. & Sirotkin, H. I. (2007). The role of the SPT6 chromatin remodeling factor in zebrafish embryogenesis. *Dev. Biol.* **307**, 214–226.
5. Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W. & Lis, J. T. (2009). Spt6 enhances the elongation rate of RNA polymerase II *in vivo*. *EMBO J.* **28**, 1067–1077.
6. Nishiwaki, K., Sano, T. & Miwa, J. (1993). *emb-5*, a gene required for the correct timing of gut precursor cell division during gastrulation in *Caenorhabditis elegans*, encodes a protein similar to the yeast nuclear protein SPT6. *Mol. Gen. Genet.* **239**, 313–322.
7. Shen, X., Xi, G., Radhakrishnan, Y. & Clemmons, D. R. (2009). Identification of novel SHPS-1-associated proteins and their roles in regulation of insulin-like growth factor-dependent responses in vascular smooth muscle cells. *Mol. Cell. Proteomics*, **8**, 1539–1551.
8. Baniahmad, C., Nawaz, Z., Baniahmad, A., Gleeson, M. A., Tsai, M. J. & O'Malley, B. W. (1995). Enhancement of human estrogen receptor activity by SPT6: a potential coactivator. *Mol. Endocrinol.* **9**, 34–43.
9. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. (2007). The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* **21**, 160–174.
10. Vanti, M., Gallastegui, E., Respaldiza, I., Rodriguez-Gil, A., Gomez-Herreros, F., Jimeno-Gonzalez, S. *et al.* (2009). Yeast genetic analysis reveals the involvement of chromatin reassembly factors in repressing HIV-1 basal transcription. *PLoS Genet.* **5**, e1000339.
11. Adkins, M. W. & Tyler, J. K. (2006). Transcriptional activators are dispensable for transcription in the absence of Spt6-mediated chromatin reassembly of promoter regions. *Mol. Cell*, **21**, 405–416.
12. Kaplan, C. D., Laprade, L. & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, **301**, 1096–1099.
13. Bortvin, A. & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, **272**, 1473–1476.
14. McDonald, S. M., Close, D., Xin, H., Formosa, T. & Hill, C. P. (2010). Structure and biological importance of the Spn1–Spt6 interaction, and its regulatory role in nucleosome binding. *Mol. Cell*, **40**, 725–735.
15. Yoh, S. M., Lucas, J. S. & Jones, K. A. (2008). The Iws1: Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* **22**, 3422–3434.
16. Endoh, M., Zhu, W., Hasegawa, J., Watanabe, H., Kim, D. K., Aida, M. *et al.* (2004). Human Spt6 stimulates transcription elongation by RNA polymerase II *in vitro*. *Mol. Cell. Biol.* **24**, 3324–3336.
17. Andrulis, E. D., Werner, J., Nazarian, A., Erdjument-Bromage, H., Tempst, P. & Lis, J. T. (2002). The RNA processing exosome is linked to elongating RNA polymerase II in *Drosophila*. *Nature*, **420**, 837–841.
18. Kaplan, C. D., Holland, M. J. & Winston, F. (2005). Interaction between transcription elongation factors and mRNA 3′-end formation at the *Saccharomyces cerevisiae* GAL10–GAL7 locus. *J. Biol. Chem.* **280**, 913–922.
19. Pawson, T. (2004). Specificity in signal transduction: from phosphotyrosine–SH2 domain interactions to complex cellular systems. *Cell*, **116**, 191–203.
20. Maclennan, A. J. & Shaw, G. (1993). A yeast SH2 domain. *Trends Biochem. Sci.* **18**, 464–465.
21. Johnson, S. J., Close, D., Robinson, H., Vallet-Gely, I., Dove, S. L. & Hill, C. P. (2008). Crystal structure and RNA binding of the Tex protein from *Pseudomonas aeruginosa*. *J. Mol. Biol.* **377**, 1460–1473.
22. Sun, M., Lariviere, L., Dengl, S., Mayer, A. & Cramer, P. (2010). A tandem SH2 domain in transcription elongation factor Spt6 binds the phosphorylated RNA polymerase II CTD. *J. Biol. Chem.* **285**, 41597–41603.
23. Diebold, M. L., Loeliger, E., Koch, M., Winston, F., Cavarelli, J. & Romier, C. (2010). A non-canonical tandem SH2 enables interaction of elongation factor SPT6 with RNA polymerase II. *J. Biol. Chem.* **285**, 38389–38398.
24. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
25. Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652.
26. Zwart, P. H., Afonine, P. V., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J. *et al.* (2008). Automated structure solution with the PHENIX suite. *Methods Mol. Biol.* **426**, 419–435.
27. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X. *et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **35**, W375–W383.
28. Dengl, S., Mayer, A., Sun, M. & Cramer, P. (2009). Structure and *in vivo* requirement of the yeast Spt6 SH2 domain. *J. Mol. Biol.* **389**, 211–225.
29. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J. Mol. Biol.* **337**, 635–645.
30. Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M. & Iyer, L. M. (2005). The many faces of the helix–turn–helix domain: transcription regulation and beyond. *FEMS Microbiol. Rev.* **29**, 231–262.
31. Grant, R. P., Marshall, N. J., Yang, J. C., Fasken, M. B., Kelly, S. M., Harreman, M. T. *et al.* (2008). Structure of the N-terminal Mlp1-binding domain of the

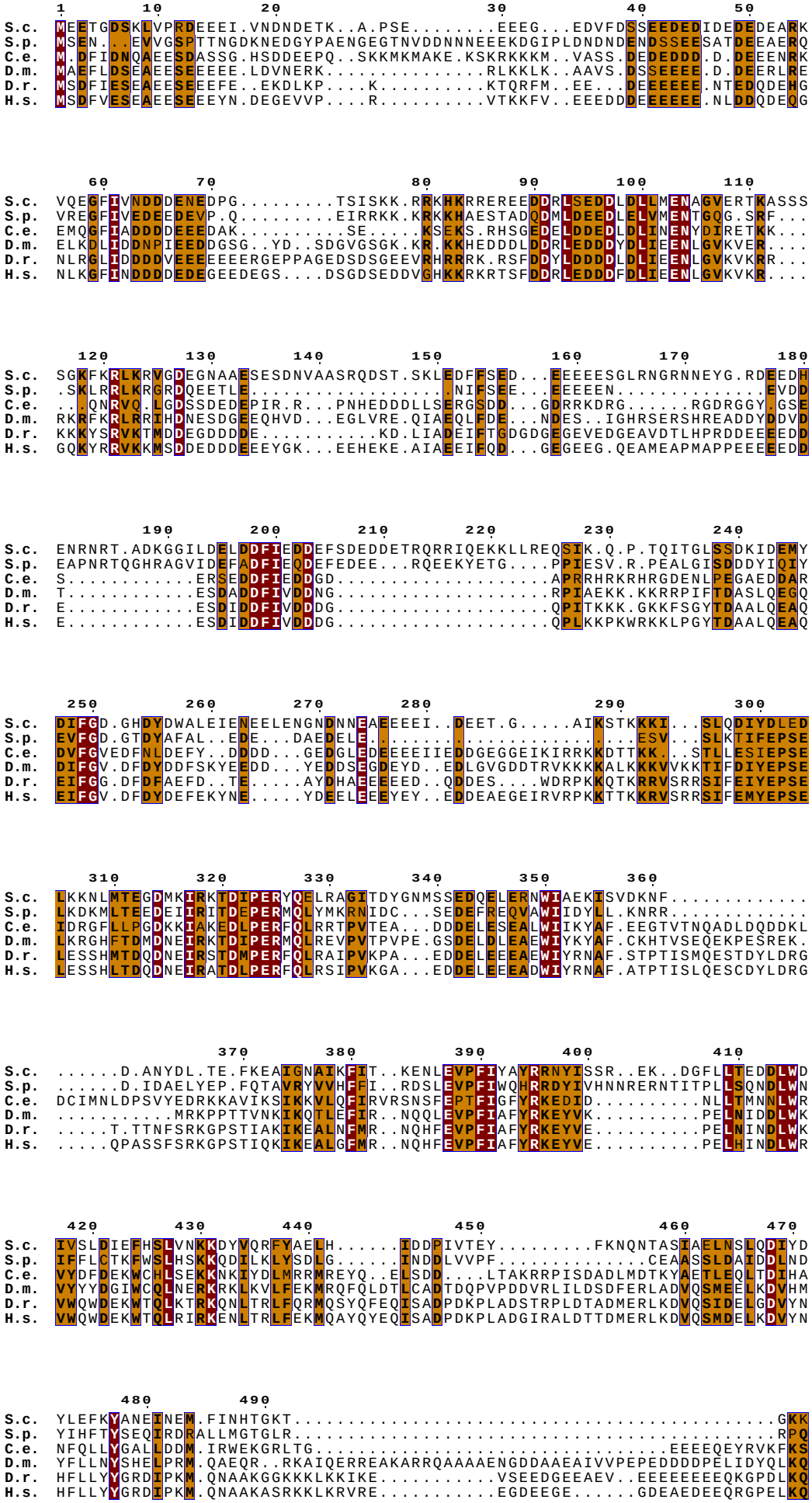
Saccharomyces cerevisiae mRNA-binding protein, Nab2. *J. Mol. Biol.* **376**, 1048–1059.
32. Saito, A., Iwasaki, H., Ariyoshi, M., Morikawa, K. & Shinagawa, H. (1995). Identification of four acidic amino acids that constitute the catalytic center of the RuvC Holliday junction resolvase. *Proc. Natl Acad. Sci. USA*, **92**, 7470–7474.
33. Shao, X. & Grishin, N. V. (2000). Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res.* **28**, 2643–2650.
34. Qiao, F. & Bowie, J. U. (2005). The many faces of SAM. *Sci. STKE*, **2005**, re7.
35. Park, H. H., Lo, Y. C., Lin, S. C., Wang, L., Yang, J. K. & Wu, H. (2007). The death domain superfamily in intracellular signaling of apoptosis and inflammation. *Annu. Rev. Immunol.* **25**, 561–586.
36. Theobald, D. L., Mitton-Fry, R. M. & Wuttke, D. S. (2003). Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.* **32**, 115–133.
37. Yu, E. Y., Wang, F., Lei, M. & Lue, N. F. (2008). A proposed OB-fold with a protein-interaction surface in *Candida albicans* telomerase protein Est3. *Nat. Struct. Mol. Biol.* **15**, 985–989.
38. Waksman, G., Shoelson, S. E., Pant, N., Cowburn, D. & Kuriyan, J. (1993). Binding of a high affinity phosphotyrosyl peptide to the Src SH2 domain: crystal structures of the complexed and peptide-free forms. *Cell*, **72**, 779–790.
39. Lee, C. H., Kominos, D., Jacques, S., Margolis, B., Schlessinger, J., Shoelson, S. E. & Kuriyan, J. (1994). Crystal structures of peptide complexes of the amino-terminal SH2 domain of the Syp tyrosine phosphatase. *Structure*, **2**, 423–438.
40. Duyster, J., Baskaran, R. & Wang, J. Y. (1995). Src homology 2 domain as a specificity determinant in the c-Abl-mediated tyrosine phosphorylation of the RNA polymerase II carboxyl-terminal repeated domain. *Proc. Natl Acad. Sci. USA*, **92**, 1555–1559.
41. Lunde, B. M., Reichow, S. L., Kim, M., Suh, H., Leeper, T. C., Yang, F. *et al.* (2010). Cooperative interaction of transcription termination factors with the RNA polymerase II C-terminal domain. *Nat. Struct. Mol. Biol.* **17**, 1195–1201.
42. Ladbury, J. E., Lemmon, M. A., Zhou, M., Green, J., Botfield, M. C. & Schlessinger, J. (1995). Measurement of the binding of tyrosyl phosphopeptides to SH2 domains: a reappraisal. *Proc. Natl Acad. Sci. USA*, **92**, 3199–3203.
43. Mayer, A., Lidschreiber, M., Siebert, M., Leike, K., Soding, J. & Cramer, P. (2010). Uniform transitions of the general RNA polymerase II transcription complex. *Nat. Struct. Mol. Biol.* **17**, 1272–1278.
44. Phatnani, H. P. & Greenleaf, A. L. (2006). Phosphorylation and functions of the RNA polymerase II CTD. *Genes Dev.* **20**, 2922–2936.
45. Lowary, P. T. & Widom, J. (1998). New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42.
46. Liu, Y., Warfield, L., Zhang, C., Luo, J., Allen, J., Lang, W. H. *et al.* (2009). Phosphorylation of the transcription elongation factor Spt5 by yeast Bur1 kinase stimulates recruitment of the PAF complex. *Mol. Cell. Biol.* **29**, 4852–4863.
47. Zlatanova, J. & van Holde, K. (1998). Binding to four-way junction DNA: a common property of architectural proteins? *FASEB J.* **12**, 421–431.
48. Guo, M., Xu, F., Yamada, J., Egelhofer, T., Gao, Y., Hartzog, G. A. *et al.* (2008). Core structure of the yeast spt4–spt5 complex: a conserved module for regulation of transcription elongation. *Structure*, **16**, 1649–1658.
49. Pei, Y. & Shuman, S. (2002). Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem.* **277**, 19639–19648.
50. Ivanov, D., Kwak, Y. T., Guo, J. & Gaynor, R. B. (2000). Domains in the SPT5 protein that modulate its transcriptional regulatory properties. *Mol. Cell. Biol.* **20**, 2970–2983.
51. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.* **41**, 207–234.
52. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In (Simon, M. I., Abelson, J. N., Carter, C. W. & Sweet, R. M., eds), Academic Press, New York, NY.
53. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **55**, 849–861.
54. Terwilliger, T. C. (2003). Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **59**, 38–44.
55. Roy, A., Kucukural, A. & Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc.* **5**, 725–738.
56. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **63**, 32–41.
57. Painter, J. & Merritt, E. A. (2006). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **62**, 439–450.
58. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **60**, 2126–2132.
59. Holm, L., Kaariainen, S., Rosenstrom, P. & Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. *Bioinformatics*, **24**, 2780–2781.
60. Krissinel, E. & Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr., Sect. D. Biol. Crystallogr.* **60**, 2256–2268.
61. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System.* Delano Scientific, Palo Alto, CA.
62. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
63. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
64. Team, R. D. C.. (2010). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
65. LiCata, V. J. & Wowor, A. J. (2008). Applications of fluorescence anisotropy to the study of protein–DNA interactions. *Methods Cell Biol.* **84**, 243–262.

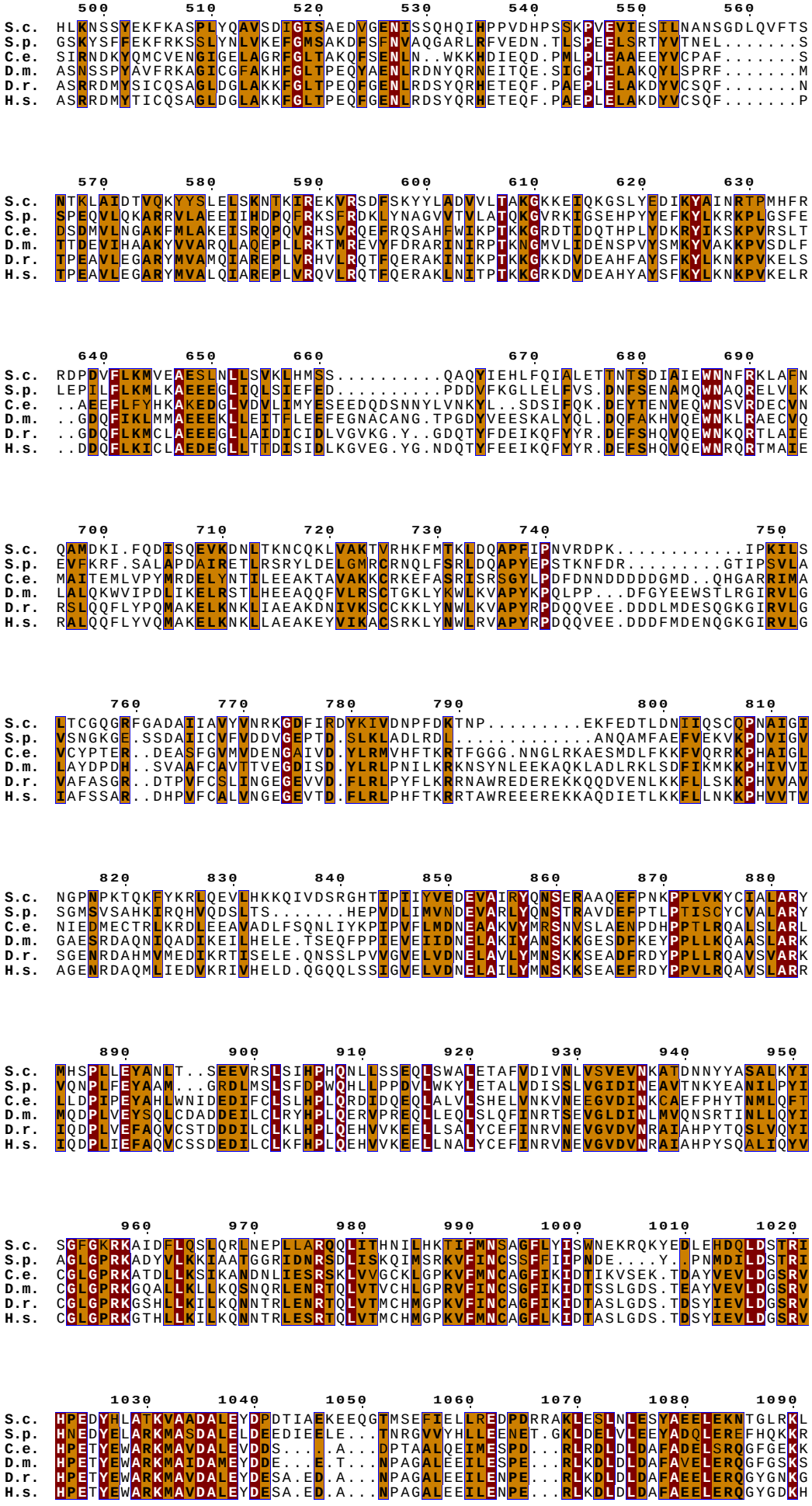
Supplemental Figure Legend

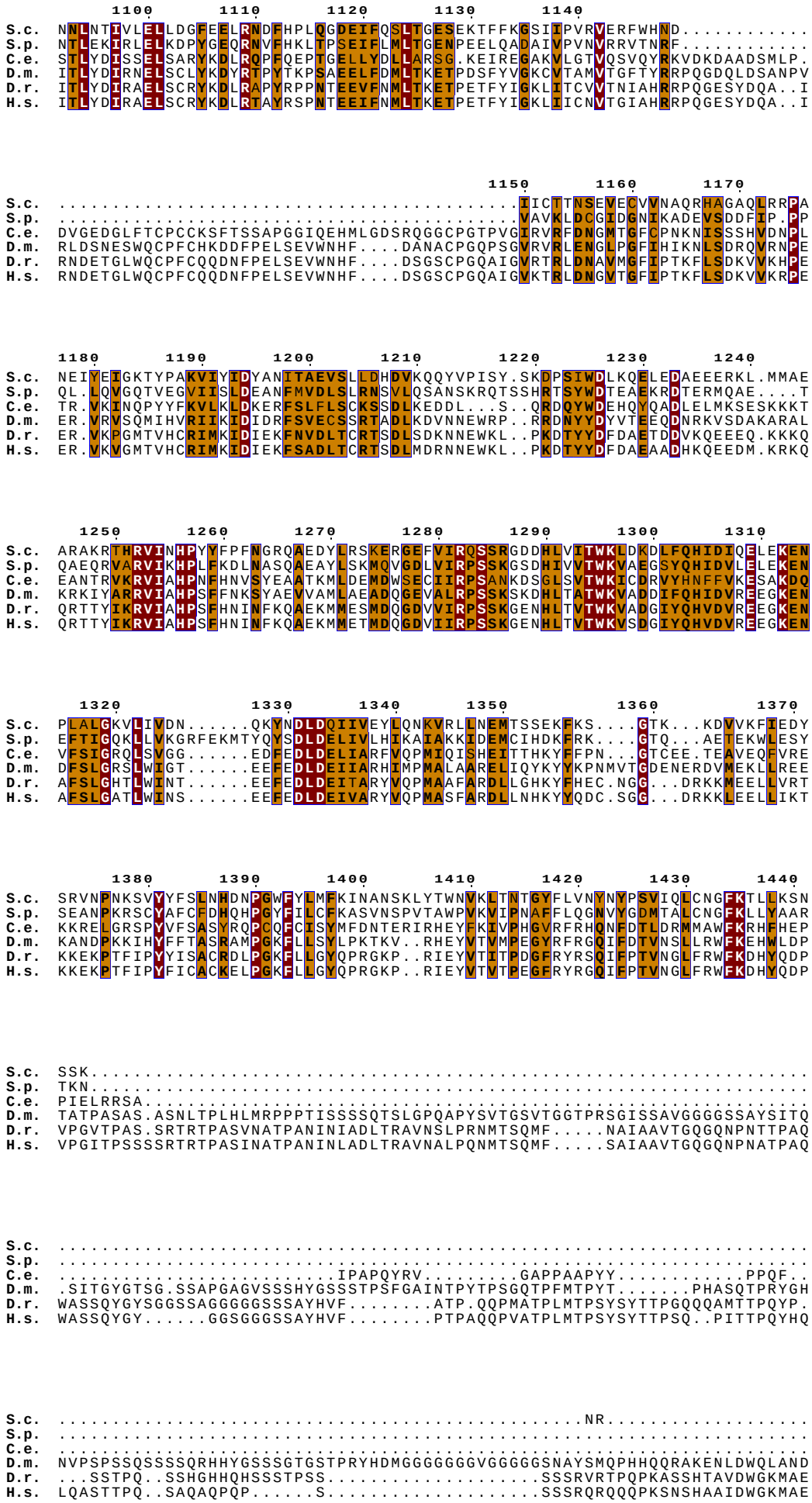
Figure S1. Spt6 Conservation.

Spt6 protein from various eukaryotic species were aligned using T-Coffee[1] and ESPript[2] was used for visualization. Coloring of sequence represents degree of conservation, dark red background (invariant), orange font (conserved), in an alignment of proteins from *Saccharomyces cerevisiae* (S.c.), *Schizosaccharomyces pombe* (S.p.), *Caenorhabditis elegans* (C.e.), *Drosophila melanogaster* (D.m.), *Danio rerio* (D.r.), and *Homo sapiens* (H.s.).

1. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-17.
2. Gouet, P., Robert, X. & Courcelle, E. (2003). ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Research* **31**, 3320-3323.


1 10 20 30 40 50
S.c. MEETGDSKLVPRDEEEI.VNDNDETK..A.PSE.....EEEG...EDVFDSSEEDEDIDEDEDEARK
S.p. MSEN...EVVGSPTTNGDKNEDGYPAENGEGTNVDDNNNEEEKDGIPLDNDNDENDSSEESATDEEAERQ
C.e. M.DFIDNQAEESDASSG.HSDDEEPQ..SKKMKMAKE.KSKRKKKM..VASS.DEDEDDD.D.DEEENRK
D.m. MAEFLDSEAEESEEEEE.LDVNERK.....RLKKLK..AAVS.DSSEEEE.D.DEERLRE
D.r. MSDFIESEAEESEEEFE..EKDLKP....K.....KTQRFM..EE...DEEEEEE.NTEDQDEHG
H.s. MSDFVESEAEESEEEYN.DEGEVVP....R.....VTKKFV..EEEDDDEEEEEE.NLDDQDEQG
60 70 80 90 100 110
S.c. VQEGFIVNDDDENEDPG.....TSISKK.RRKHKRREREEDDRLSEDDLDLLMENAGVERTKASSS
S.p. VREGFIVEDEEDEVP.Q.....EIRRKK.KRKKHAESTADQDMLDEEDLELVMENTGQG.SRF...
C.e. EMQGFIADDDDEEEDAK.....SE.....KSEKS.RHSGEDELDDEDLDLINENYDIRETKK...
D.m. ELKDLIDDNPIEEDDGSG..YD..SDGVGSGK.KR.KKHEDDDLDDRLEDDDYDLIEENLGVKVER....
D.r. NLRGLIDDDDVEEEEEEERGEPPAGEDSDSGEEVRHRRRK.RSFDDYLDDDDLDLIEENLGVKVKRR...
H.s. NLKGFINDDDDEDEGEEDEGS....DSGDSEDDVGHKKRKRTSFDDRLEDDDFDLIEENLGVKVKR....
120 130 140 150 160 170 180
S.c. SGKFKFKRLKRVGDEGNAAESESDNVAASRQDST.SKLEDFFSEED...EEEEESGLRNGRNNEYG.RDEEDH
S.p. .SKLRRLKRGRDQEETLE.....NIFSEE...EEEEEN.....EVDD
C.e. ..QNRVQ.LGDSSDEDEPIR.R...PNHEDDDLLSERGSDD...GDRRKDRG.....RGDRGGY.GSE
D.m. RKRFKFKLRRIHDNESDGEEQHVD...EGLVRE.QIAEQLFDE...NDES..IGHRSERSHREADDYDDVD
D.r. KKKYSRVKTMDDEGDDDDE.....KD.LIADEIFTGDGDGEGEVEDGEAVDTLHPRDDEEEEEDD
H.s. GQKYRRVKKMSDDEDDDEEEYGK...EEHEKE.AIAEEIFQDD...GEGEEG.QEAMEAPMAPPEEEEEEDD
190 200 210 220 230 240
S.c. ENRNRT.ADKGGILDELDDFIEDDEFSDEDDETRQRRIQEKKLLREQSIK.Q.P.TQITGLSSDKIDEMY
S.p. EAPNRTQGHRAGVIDEFADFIEQDEFEDEE...RQEEKYETG....PPIESV.R.PEALGISDDDYIQIY
C.e. S.....ERSEDDFIEDDGD.....APRRHRKRHRGDENLPEGAEDDAR
D.m. T.....ESDADDFIVDDNG.....RPIAEKK.KKRRPIFTDASLQEGQ
D.r. E.....ESDIDDFIVDDDG.....QPITKKK.GKKFSGYTDAALQEAQ
H.s. E.....ESDIDDFIVDDDG.....QPLKKPKWRKKLPGYTDAALQEAQ
250 260 270 280 290 300
S.c. DIFGD.GHDYDWALEIENEELENGNDNNEAEEEEI..DEET.G.....AIKSTKKKI...SLQDIYDLED
S.p. EVFGD.GTDYAFAL..EDE...DAEDELE.....ESV...SLKTIFEPSE
C.e. DVFGVEDFNLDEFY..DDDD...GEDGLEDEEEEIIEDDGEGGEIKIRRKKDTTKK...STLLESIEPSE
D.m. DIFGV.DFDYDDFSKYEEDD...YEDDSEGDEYD..EDLGVGDDTRVKKKKALKKKVVKKTIFDIYEPSE
D.r. EIFGG.DFDFAEFD..TE.....AYDHAEEEEED..QDDES....WDRPKKQTKRRVSRRSIFEIYEPSE
H.s. EIFGV.DFDYDEFEKYNE.....YDEELEEEYEY..EDDEAEGEIRVRPKKTTKKRVSRRSIFEMYEPSE
310 320 330 340 350 360
S.c. LKKNLMTEGDMKIRKTDIPERYQELRAGITDYGNMSSEDQELERNWIAEKISVDKNF.....
S.p. LKDKMLTEEDEIIRITDEPERMQLYMKRNIDC...SEDEFREQVAWIIDYLL.KNRR.....
C.e. IDRGFLLPGDKKIAKEDLPERFQLRRTPVTEA...DDDELESEALWIIKYAF.EEGTVTNQADLDQDDKL
D.m. LKRGHFTDMDNEIRKTDIPERMQLREVPVTPVPE.GSDELDLEAEWIYKYAF.CKHTVSEQEKPESREK.
D.r. LESSHMTDQDNEIRSTDMPERFQLRAIPVKPA...EDDELEEEAEWIYRNAF.STPTISMQESTDYLDRG
H.s. LESSHLTDQDNEIRATDLPERFQLRSIPVKGA...EDDELEEEADWIYRNAF.ATPTISLQESCDYLDRG
370 380 390 400 410
S.c. ......D.ANYDL.TE.FKEAIGNAIKFIT..KENLEVPFIYAYRRNYISSR..EK..DGFLLTEDDLWD
S.p. ......D.IDAELYEP.FQTAVRYVVHFFI..RDSLEVPFIWQHRRDYIVHNNRERNTITPLLSQNDLWN
C.e. DCIMNLDPSVYEDRKKAVIKSIKKVLQFIRVRSNSFEPTFIGFYRKEDID.....NLLTMNNLWR
D.m. .....MRKPPTTVNKIKQTLEFIR..NQQLEVPFIAFYRKEYVK.....PELNIDDLWK
D.r. .....T.TTNFSRKGPSTIAKIKEALNFMR..NQHFEVPFIAFYRKEYVE.....PELNINDLWK
H.s. .....QPASSFSRKGPSTIQKIKEALGFMR..NQHFEVPFIAFYRKEYVE.....PELHINDLWR
420 430 440 450 460 470
S.c. IVSLDIEFHSLVNKKDYVQRFYAELH.....IDDPIVTEY.....FKNQNTASIAELNSLQDIYD
S.p. IFFLCTKFWSLHSKKQDILKLYSDLG.....INDDLVVPF.....CEAASSLDAIDDLND
C.e. VYDFDEKWCHLSEKKNKIYDLMRRMREYQ..ELSDD....LTAKRRPISDADLMDTKYAETLEQLTDIHA
D.m. VYYYDGIWCQLNERKRKLKVLFEKMRQFQLDTLCADTDQPVPDDVRLILDSDFERLADVQSMEELKDVHM
D.r. VWQWDEKWTQLKTRKQNLTRLFQRMQSYQFEQISADPDKPLADSTRPLDTADMERLKDVQSIDELGDVYN
H.s. VWQWDEKWTQLRIRKENLTRLFEKMQAYQYEQISADPDKPLADGIRALDTTDMERLKDVQSMDELKDVYN
480 490
S.c. YLEFKYANEINEM.FINHTGKT.....GKK
S.p. YIHFTYSEQIRDRALLMGTGLR.....RPQ
C.e. NFQLLYGALLDDM.IRWEKGRLTG.....EEEEQEYRVKFKS
D.m. YFLLNYSHELPRM.QAEQR..RKAIQERREAKARRQAAAAENGDDAAEAIVVPEPEDDDDPELIDYQLKQ
D.r. HFLLYYGRDIPKM.QNAAKGGKKKLKKIKE.....VSEEDGEEAEV..EEEEEEEEQKGPDLKQ
H.s. HFLLYYGRDIPKM.QNAAKASRKKLKRVRE.....EGDEEGE.....GDEAEDEEQRGPELKQ


500 510 520 530 540 550 560
S.c. HLKNSSYEKFKASPLYQAVSDIGISAEDVGENISSQHQIHPPVDHPSSKPVEVIESILNANSGDLQVFTS
S.p. GSKYSFFEKFRKSSLYNLVKEFGMSAKDFSFNVAQGARLRFVEDN.TLSPEELSRTYVTNEL.......S
C.e. SIRNDKYQMCVENGIGELAGRFGLTAKQFSENLN..WKKHDIEQD.PMLPLEAAEEYVCPAF.......S
D.m. ASNSSPYAVFRKAGICGFAKHFGLTPEQYAENLRDNYQRNEITQE.SIGPTELAKQYLSPRF.......M
D.r. ASRRDMYSICQSAGLDGLAKKFGLTPEQFGENLRDSYQRHETEQF.PAEPLELAKDYVCSQF.......N
H.s. ASRRDMYTICQSAGLDGLAKKFGLTPEQFGENLRDSYQRHETEQF.PAEPLELAKDYVCSQF.......P
570 580 590 600 610 620 630
S.c. NTKLAIDTVQKYYSLELSKNTKIREKVRSDFSKYYLADVVLTAKGKKEIQKGSLYCDIKYAINRTPMHFR
S.p. SPEQVLQKARRVLAEEIIHDPQFRKSFRDKLYNAGVVTVLATQKGVRKIGSEHPYYEFKYLKRKPLGSFE
C.e. DSDMVLNGAKFMLAKEISRQPQVRHSVRQEFRQSAHFWIKPTKKGRDTIDQTHPLYDKRYIKSKPVRSLT
D.m. TTDEVIHAAKYVVARQLAQEPLLRKTMREVYFDRARINIRPTKNGMVLIDENSPVYSMKYVAKKPVSDLF
D.r. TPEAVLEGARYMVAMQIAREPLVRHVLRQTFQERAKINIKPTKKGKKDVDEAHFAYSFKYLKNKPVKELS
H.s. TPEAVLEGARYMVALQIAREPLVRQVLRQTFQERAKLNITPTKKGRKDVDEAHYAYSFKYLKNKPVKELR
640 650 660 670 680 690
S.c. RDPDVFLKMVEAESLNLLSVKLHMSS..........QAQYIEHLFQIALETTNTSDIAIEWNFRKLAFN
S.p. LEPILFLKMLKAEEEGLIQLSIEFED..........PDDVFKGLLELFVS.DNFSENAMQWNAQRELVLK
C.e. ..AEEFLFYHKAKEDGLVDVLIMYESEEDQDSNNYLVNKYL..SDSIFQK.DEYTENVEQWNSVRDECVN
D.m. ..GDQFIKLMMAEEEKLLEITFLEEFEGNACANG.TPGDYVEESKALYQL.DQFAKHVQEWNKLRAECVQ
D.r. ..GDQFLKMCLAEEEGLLAIDICIDLVGVKG.Y..GDQTYFDEIKQFYYR.DEFSHQVQEWNKQRTLAIE
H.s. ..DDQFLKICLAEDEGLLTTDISIDLKGVEG.YG.NDQTYFEEIKQFYYR.DEFSHQVQEWNRQRTMAIE
700 710 720 730 740 750
S.c. QAMDKI.FQDISQEVKDNLTKNCQKLVAKTVRHKFMTKLDQAPFIPNVRDPK.....IPKILS
S.p. EVFKRF.SALAPDAIRETLRSRYLDELGMRCRNQLFSRLDQAPYEPSTKNFDR.....GTIPSVLA
C.e. MAITEMLVPYMRDELYNTILEEAKTAVAKKCRKEFASRISRSGYLPDFDNNDDDDDGMD..QHGARRIMA
D.m. LALQKWVIPDLIKELRSTLHEEAQQFVLRSCTGKLYKWLKVAPYKPQLPP...DFGYEEWSTLRGIRVLG
D.r. RSLQQFLYPQMAKELKNKLIAEAKDNIVKSCCKKLYNWLKVAPYRPDQQVEE.DDDLMDESQGKGIRVLG
H.s. RALQQFLYVQMAKELKNKLLAEAKEYVIKACSRKLYNWLRVAPYRPDQQVEE.DDDFMDENQGKGIRVLG
760 770 780 790 800 810
S.c. LTCGQGRFGADAIIAVYVNRKGDFIRDYKIVDNPFDKTNP.....EKFEDTLDNIIQSCQPNAIGI
S.p. VSNGKGE.SSDAIICVFVDDVGEPTD.SLKLADLRDL.....ANQAMFAEFVEKVKPDVIGV
C.e. VCYPTER..DEASFGVMVDENGAIVD.YLRMVHFTKRTFGGG.NNGLRKAESMDLFKKFVQRRKPHAIGL
D.m. LAYDPDH..SVAAFCAVTTVEGDISD.YLRLPNILKRKNSYNLEEKAQKLADLRKLSDFIKMKKPHIVVI
D.r. VAFASGR..DTPVFCSLINGEGEVVD.FLRLPYFLKRRNAWREDEREKKQQDVENLKKFLLSKKPHVVAV
H.s. IAFSSAR..DHPVFCALVNGEGEVTD.FLRLPHFTKRRTAWREEEREKKAQDIETLKKFLLNKKPHVVTV
820 830 840 850 860 870 880
S.c. NGPNPKTQKFYKRLQEVLHKKQIVDSRGHTIPIIYVEDEVAIRYQNSERAAQEFPNKPPLVKYCIALARY
S.p. SGMSVSAHKIRQHVQDSLTS.....HEPVDLIMVNDEVARLYQNSTRAVDEFPTLPTISCYCVALARY
C.e. NIEDMECTRLKRDLEEAVADLFSQNLIYKPIPVFLMDNEAAKVYMRSNVSLAENPDHPPTLRQALSLARL
D.m. GAESRDAQNIQADIKEILHELE.TSEQFPPIEVEIIDNELAKIYANSKKGESDFKEYPPLLKQAASLARK
D.r. SGENRDAHMVMEDIKRTISELE.QNSSLPVVGVELVDNELAVLYMNSKKSEADFRDYPPLLRQAVSVARK
H.s. AGENRDAQMLIEDVKRIVHELD.QGQQLSSIGVELVDNELAILYMNSKKSEAEFRDYPPVLRQAVSLARR
890 900 910 920 930 940 950
S.c. MHSPLLEYANLT..SEEVRSLSIHPHQNLLSSEQLSWALETAFVDIVNLVSVEVNKATDNNYYASALKYI
S.p. VQNPLFEYAAM...GRDLMSLSFDPWQHLLPPDVLWKYLETALVDISSLVGIDINEAVTNKYEANILPYI
C.e. LLDPIPEYAHLWNIDEDIFCLSLHPLQRDIDQEQLALVLSHELVNKVNEEGVDINKCAEFPHYTNMLQFT
D.m. MQDPLVEYSQLCDADDEICLRYHPLQERVPREQLLEQLSLQFINRTSEVGLDINLMVQNSRTINLLQYI
D.r. IQDPLVEFAQVCSTDDDICLKLHPLQEHVVKEELLSALYCEFINRVNEVGVDVNRAIAHPYTQSLVQYI
H.s. IQDPLIEFAQVCSSDEDICLKFHPLQEHVVKEELLNALYCEFINRVNEVGVDVNRAIAHPYSQALIQYV
960 970 980 990 1000 1010 1020
S.c. SGFGKRKAIDFLQSLQRLNEPLLARQQLITHNILHKTIFMNSAGFLYISWNEKRQKYEDLEHDQLDSTRI
S.p. AGLGPRKADYVLKKIAATGGRIDNRSDLISKQIMSRKVFINCSSFFIIPNDE....Y..PNMDILDSTRI
C.e. CGLGPRKATDLLKSIKANDNLIESRSKLVVGCKLGPKVFMNCAGFIKIDTIKVSEK.TDAYVEVLDGSRV
D.m. CGLGPRKGQALLKLLKQSNQRLENRTQLVTVCHLGPRVFINCSGFIKIDTSSLGDS.TEAYVEVLDGSRV
D.r. CGLGPRKGSHLLKILKQNNTRLENRTQLVTMCHMGPKVFINCAGFIKIDTASLGDS.TDSYIEVLDGSRV
H.s. CGLGPRKGTHLLKILKQNNTRLESRTQLVTMCHMGPKVFMNCAGFLKIDTASLGDS.TDSYIEVLDGSRV
1030 1040 1050 1060 1070 1080 1090
S.c. HPEDYHLATKVAADALEYDPDTIAEKEEQGTMSEFIELLREDPDRRAKLESLNLESYAEELEKNTGLRKL
S.p. HNEDYELARKMASDALELDEEDIEELE...TNRGVVYHLLEENET.GKLDELVLEEYADQLEREFHQKKR
C.e. HPETYEWARKMAVDALEVDDS....A...DPTAALQEIMESPD...RLRDLDLDAFADELSRQGFGEKK
D.m. HPETYEWARKMAIDAMEYDDE...E.T...NPAGALEEILESPE...RLKDLDLDAFAVELERQGFGSKS
D.r. HPETYEWARKMAVDALEYDESA.ED.A...NPAGALEEILENPE...RLKDLDLDAFAEELERQGYGNKG
H.s. HPETYEWARKMAVDALEYDESA.ED.A...NPAGALEEILENPE...RLKDLDLDAFAEELERQGYGDKH


1100 1110 1120 1130 1140
S.c. NNLNTIVLELLDGFEELRNDFHPLQGDEIFQSLTGESEKTFFKGSITPVRVERFWHND
S.p. NTLEKIRLELKDPYGEQRNVFHKLTPSEIFLMLTGENPEELQADAIVPVNVRRVTNRF
C.e. STLYDISSELSARYKDLRQPFQEPTGELLYDLLARSG.KEIREGAKVLGTVQSVQYRKVDKDAADSMLP.
D.m. ITLYDIRNELSCLYKDYRTPYTKPSAEELFDMLTKETPDSFYVGKCVTAMVTGFTYRRPQGDQLDSANPV
D.r. ITLYDIRAELSCRYKDLRAPYRPPNTEEVFNMLTKETPETFYIGKLITCVVTNIAHRRPQGESYDQA..I
H.s. ITLYDIRAELSCRYKDLRTAYRSPNTEEIFNMLTKETPETFYIGKLIICNVTGIAHRRPQGESYDQA..I
1150 1160 1170
S.c. .....IICTTNSEVECVVNAQRHAGAQLRRPA
S.p. .....VAVKLDCGIDGNIKADEVSDDFIP.PP
C.e. DVGEDGLFTCPCCKSFTSSAPGGIQEHMLGDSRQGGCPGTPVGIRVRFDNGMTGFCPNKNISSSHVDNPL
D.m. RLDSNESWQCPFCHKDDFPELSEVWNHF....DANACPGQPSGVRVRLENGLPGFIHIKNLSDRQVRNPE
D.r. RNDETGLWQCPFCQQDNFPELSEVWNHF....DSGSCPGQAIGVRTRLDNAVMGFIPTKFLSDKVVKHPE
H.s. RNDETGLWQCPFCQQDNFPELSEVWNHF....DSGSCPGQAIGVKTRLDNGVTGFIPTKFLSDKVVKRPE
1180 1190 1200 1210 1220 1230 1240
S.c. NEIYEIGKTYPAKVIYIDYANITAEVSLLDHDVKQQYVPISY.SKDPSIWDLKQELEDAEEERKL.MMAE
S.p. QL.LQVGQTVEGVIISLDEANFMVDLSLRNSVLQSANSKRQTSSHRTSYWDTEAEKRDTERMQAE....T
C.e. TR.VKINQPYYFKVLKLDKERFSLFLSCKSSDLKEDDL...S..QRDQYWDEHQYQADLELMKSESKKKT
D.m. ER.VRVSQMIHVRIIKIDIDRFSVECSSRTADLKDVNNEWRP..RRDNYYDYVTEEQDNRKVSDAKARAL
D.r. ER.VKPGMTVHCRIMKIDIEKFNVDLTCRTSDLSDKNNEWKL..PKDTYYDFDAETDDVKQEEEQ.KKKQ
H.s. ER.VKVGMTVHCRIMKIDIEKFSADLTCRTSDLMDRNNEWKL..PKDTYYDFDAEAADHKQEEDM.KRKQ
1250 1260 1270 1280 1290 1300 1310
S.c. ARAKRTHRVINHPYYFPFNGRQAEDYLRSKERGEFVIRQSSRGDDHLVITWKLDKDLFQHIDIQELEKEN
S.p. QAEQRVARVIKHPLFKDLNASQAEAYLSKMQVGDLVIRPSSKGSDHIVVTWKVAEGSYQHIDVLELEKEN
C.e. EANTRVKRVIAHPNFHNVSYEAATKMLDEMDWSECIIRPSANKDSGLSVTWKICDRVYHNFFVKESAKDQ
D.m. KRKIYARRVIAHPSFFNKSYAEVVAMLAEADQGEVALRPSSKSKDHLTATWKVADDIFQHIDVREEGKEN
D.r. QRTTYIKRVIAHPSFHNINFKQAEKMMESMDQGDVVIRPSSKGENHLTVTWKVADGIYQHVDVREEGKEN
H.s. QRTTYIKRVIAHPSFHNINFKQAEKMMETMDQGDVIIRPSSKGENHLTVTWKVSDGIYQHVDVREEGKEN
1320 1330 1340 1350 1360 1370
S.c. PLALGKVLIVDN......QKYNDLDQIIVEYLQNKVRLLNEMTSSEKFKS....GTK...KDVVKFIEDY
S.p. EFTIGQKLLVKGRFEKMTYQYSDLDELIVLHIKAIAKKIDEMCIHDKFRK....GTQ...AETEKWLESY
C.e. VFSIGRQLSVGG......EDFEDLDELIARFVQPMIQISHEITTHKYFFPN...GTCEE.TEAVEQFVRE
D.m. DFSLGRSLWIGT......EEFEDLDEIIARHIMPMALAARELIQYKYYKPNMVTGDENERDVMEKLLREE
D.r. AFSLGHTLWINT......EEFEDLDEITARYVQPMAAFARDLLGHKYFHEC.NGG...DRKKMEELLVRT
H.s. AFSLGATLWINS......EEFEDLDEIVARYVQPMASFARDLLNHKYYQDC.SGG...DRKKLEELLIKT
1380 1390 1400 1410 1420 1430 1440
S.c. SRVNPNKSVYYFSLNHDNPGWFYLMFKINANSKLYTWNVKLTNTGYFLVNYNYPSVIQLCNGFKTLLKSN
S.p. SEANPKRSCYAFCFDHQHPGYFILCFKASVNSPVTAWPVKVIPNAFFLQGNVYGDMTALCNGFKLLYAAR
C.e. KKRELGRSPYVFSASYRQPCQFCISYMFDNTERIRHEYFKIVPHGVRFRHQNFDTLDRMMAWFKRHFHEP
D.m. KANDPKKIHYFFTASRAMPGKFLLSYLPKTKV..RHEYVTVMPEGYRFRGQIFDTVNSLLRWFKEHWLDP
D.r. KKEKPTFIPYYISACRDLPGKFLLGYQPRGKP..RIEYVTITPDGFRYRSQIFPTVNGLFRWFKDHYQDP
H.s. KKEKPTFIPYFICACKELPGKFLLGYQPRGKP..RIEYVTVTPEGFRYRGQIFPTVNGLFRWFKDHYQDP
S.c. SSK.....
S.p. TKN.....
C.e. PIELRRSA.....
D.m. TATPASAS.ASNLTPLHLMRPPPTISSSSQTSLGPQAPYSVTGSVTGGTPRSGISSAVGGGGSSAYSITQ
D.r. VPGVTPAS.SRTRTPASVNATPANINIADLTRAVNSLPRNMTSQMF.....NAIAAVTGQGQNPNTTPAQ
H.s. VPGITPSSSSRTRTPASINATPANINLADLTRAVNALPQNMTSQMF.....SAIAAVTGQGQNPNATPAQ
S.c. .....
S.p. .....
C.e. .....IPAPQYRV.....GAPPAAPYY.....PPQF..
D.m. .SITGYGTSG.SSAPGAGVSSSHYGSSSTPSFGAINTPYTPSGQTPFMTPYT.....PHASQTPRYGH
D.r. WASSQYGYSGGSSAGGGGGSSSAYHVF.....ATP.QQPMATPLMTPSYSYTTPGQQQAMTTPQYP.
H.s. WASSQYGY.....GGSGGGSSAYHVF.....PTPAQQPVATPLMTPSYSYTTPSQ..PITTPQYHQ
S.c. .....NR.....
S.p. .....
C.e. .....
D.m. NVPSPSSQSSSSQRHHYGSSSGTGSTPRYHDMGGGGGGGVGGGGGSNAYSMQPHHQQRAKENLDWQLAND
D.r. ...SSTPQ..SSHGHHQHSSSTPSS.....SSSRVRTPQPKASSHTAVDWGKMAE
H.s. LQASTTPQ..SAQAQPQP.....S.....SSSRQRQQQPKSNSHAAIDWGKMAE

```
S.c.  ..........................................................................
S.p.  ..........................................................................
C.e.  ..........................................................................
D.m.  AWARRRPQQHQSHQSYHAQQQHHHSQQQPHMGMSMNMGITMSLGRGTGGGGGGGYGSTPVNDYSTGGGHN
D.r.  QWLQEKEAERRKQK..............................................................
H.s.  QWLQEKEAERRKQK..............................................................
```

```
                                          1450
                                           .
S.c.  ....................................MNNYR
S.p.  ....................................FRRM
C.e.  ....................................VGYH
D.m.  RGMSSKASVRSTPRTNASPHS....MN.LGDATPLYD..EN
D.r.  .......TPRMTPRPSPSPMIESTPMSIAGDATPLLDEMDR
H.s.  ........QRLTPRPSPSPMIESTPMSIAGDATPLLDEMDR
```