doi:10.1016/j.jmb.2008.01.096 J. Mol. Biol. (2008) 377, 1460–1473

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Crystal Structure and RNA Binding of the Tex Protein from *Pseudomonas aeruginosa*

Sean J. Johnson[1]†, Devin Close[2]†, Howard Robinson[3], Isabelle Vallet-Gely[4], Simon L. Dove[4] and Christopher P. Hill[2]*

[1]*Department of Chemistry and Biochemistry, Utah State University, Logan, UT 84322-0300, USA*

[2]*Department of Biochemistry, University of Utah, Salt Lake City, UT 84112-5650, USA*

[3]*Department of Biology, Brookhaven National Laboratory, Upton, NY 11973, USA*

[4]*Division of Infectious Diseases, Children's Hospital, Harvard Medical School, Boston, MA 02115, USA*

Received 18 December 2007; received in revised form 23 January 2008; accepted 25 January 2008
Available online 12 February 2008

Edited by J. Doudna

Tex is a highly conserved bacterial protein that likely functions in a variety of transcriptional processes. Here, we describe two crystal structures of the 86-kDa Tex protein from *Pseudomonas aeruginosa* at 2.3 and 2.5 Å resolution, respectively. These structures reveal a relatively flat and elongated protein, with several potential nucleic acid binding motifs clustered at one end, including an S1 domain near the C-terminus that displays considerable structural flexibility. Tex binds nucleic acids, with a preference for single-stranded RNA, and the Tex S1 domain is required for this binding activity. Point mutants further demonstrate that the primary nucleic acid binding site corresponds to a surface of the S1 domain. Sequence alignment and modeling indicate that the eukaryotic Spt6 transcription factor adopts a similar core structure. Structural analysis further suggests that the RNA polymerase and nucleosome interacting regions of Spt6 flank opposite sides of the Tex-like scaffold. Therefore, the Tex structure may represent a conserved scaffold that binds single-stranded RNA to regulate transcription in both eukaryotic and prokaryotic organisms.

Keywords: transcription; S1 domain; Tex; x-ray crystallography; Spt6

Introduction

The Tex (*t*oxin *ex*pression) protein was originally described in *Bordetella pertussis* as an essential protein involved in expression of critical toxin genes.[1] Tex is a relatively large protein with a domain architecture consisting of several nucleic acid binding domains predicted from primary sequence. The presence of these domains supports the proposal that Tex is a transcription factor that functions in toxin expression and/or pathogen fitness.[1–3] Tex displays a remarkably high degree of identity and similarity across a host of significant pathogens. For example, Tex from *Pseudomonas aeruginosa* shares 65% identity and 78% similarity (at the amino acid level) with Tex from *Vibrio cholerae* (the causative agent of cholera). Similar degrees of identity are seen with Tex proteins from *Shigella flexneri* (the causative agent of dysentery) and *Yersinia pestis* (the causative agent of plague).

Despite being ubiquitous and extremely well conserved, the molecular functions of Tex remain enigmatic. Insight into Tex function is derived from several bacterial studies. Aside from its role in expression of toxin gene products in *B. pertussis*, the *tex* gene from *P. aeruginosa* (PA5201) appears to play an important role in pathogenesis, being required for lung infection in a chronic disease model.[4] In *Streptococcus pneumoniae*, Tex does not alter expres-

*Corresponding author. E-mail address: chris@biochem.utah.edu.

† S.J.J. and D.C. are joint first authors.

Abbreviations used: RNAP, RNA polymerase; HtH, helix–turn–helix; ssRNA, single-stranded RNA; Se-Met, selenomethionine; PDB, Protein Data Bank; HhH, helix–hairpin–helix; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; OB, oligonucleotide/oligosaccharide binding; dsRNA, double-stranded RNA; FP, fluorescence polarization; WT, wild-type; BME, β-mercaptoethanol; NSLS, National Synchrotron Light Source.

0022-2836/$ - see front matter. Published by Elsevier Ltd.

sion of the major pneumococcus toxin pneumolysin but does appear to be a transcription factor involved in pathogen fitness.[3] These studies indicate that Tex may play a role in gene expression or transcript maintenance of either specific toxin or general housekeeping genes.

Tex domain architecture and sequence conservation may extend beyond prokaryotes to the essential eukaryotic transcription elongation factor Spt6.[5–7] Tex is approximately half the size of Spt6 (e.g., 86 kDa for *P. aeruginosa* Tex *versus* 168 kDa for *Saccharomyces cerevisiae* Spt6), with sequence homology spanning the central region of Spt6. The flanking nonhomologous regions of Spt6 include a highly charged N-terminal region and a C-terminal SH2-like domain. Within the region of homology, Tex and Spt6 share ~25% pairwise sequence identity and have a similar predicted domain architecture; primary sequence analysis identified YqgF, HhH, and S1 RNA-binding domains in both proteins.[7] This level of sequence similarity falls in Doolittle's "twilight zone,"[8] indicating that Tex and Spt6 may have similar structures, although direct evidence is lacking.

The sequence similarity may also indicate that Tex and Spt6 have related cellular functions. Although current evidence suggests that Spt6 is a nucleosome chaperone,[9–11] a function unique to eukaryotes, recent studies have shown that Spt6 also interacts directly with both RNA polymerase (RNAP)[12] and mRNA processing factors, including the exosomal RNA degradation machinery.[13] Thus, beyond its role in nucleosome maintenance, Spt6 appears to provide a physical link between transcription and pre-mRNA surveillance, although the relationship between these critical processes is lacking in structural detail. Interestingly, we have recently observed similar interactions with Tex. *P. aeruginosa* Tex copurifies with RNAP, RNase E, and PNPase (I.V.-G. and S.L.D., unpublished data); RNase E and PNPase are components of the prokaryotic RNA degradosome, a 3′–5′ RNA degradation complex analogous to the eukaryotic exosome.[14]

In an effort to better understand the molecular function of Tex, and possibly to gain insight into Spt6, we have determined high-resolution crystal structures of the *P. aeruginosa* Tex protein in two crystal forms. These reveal four putative nucleic acid

Table 1. Data collection and refinement statistics

	Tex Se-Met: crystal form I	Tex native 1: crystal form I	Tex native 2: crystal form II
Data collection			
Beamline	NSLS X29	NSLS X29	Home source
Wavelength (Å)	0.978	1.10000	1.54178
Resolution (Å)	40–2.7	50–2.5	20–2.3
Outer shell (Å)	2.8–2.7	2.59–2.5	2.38–2.3
No. of reflections			
Unique	23,640	35,199	38,347
Total	273,923	448,770	343,965
Mean $I/\sigma(I)$	21.1 (4.8)	30.0 (4.2)	19.0 (3.3)
Completeness (%)	91.2 (61.7)	90.4 (64.8)	99.4 (99.3)
R_{sym}[a]	9.0 (32.0)	8.6 (40.3)	8.4 (50.7)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Unit cell dimensions (Å)			
a	57.0	57.2	56.2
b	135.1	131.8	106.7
c	144.5	144.0	139.7
Phasing (40–3.4 Å)			
FOM, before DM (SOLVE)	0.320		
FOM, after DM (RESOLVE)	0.720		
Refinement			
R_{cryst}[b]/R_{free}[c] (%)		24.0/27.4	22.1/26.6
Nonhydrogen atoms			
Total		5692	5884
Solvent		73	256
RMSD from ideal geometry			
Bond lengths (Å)		0.004	0.004
Bond angles (°)		0.6	0.6
Average isotropic *B* value (Å^2)		66.7	27.4
Protein geometry[d]			
Ramachandran outliers (%)		0.0	0.0
Ramachandran favored (%)		96.2	98.5
Rotamer outliers (%)		0.2	0.2

Values in parentheses correspond to those in the outer resolution shell.
FOM, figure of merit; DM, density modification.
[a] $R_{sym} = (\sum|(I-\langle I\rangle)|)/(\sum I)$, where $\langle I\rangle$ is the average intensity of multiple measurements.
[b] $R_{cryst} = (\sum|F_{obs}-F_{calc}|)/(\sum|F_{obs}|)$.
[c] R_{free} is the *R*-factor based on 5% of the data excluded from refinement.
[d] Geometry statistics were determined by MolProbity.[15]

binding/modifying domains including a helix–turn–helix (HtH) domain that was not predicted from primary sequence. In addition, we have quantitatively examined the ability of Tex to bind various nucleic acid substrates and have found that Tex has a strong preference for single-stranded RNA (ssRNA). Binding appears to be sequence nonspecific, and mutagenesis studies indicate that this interaction is mediated by the flexible S1 domain. In contrast to an earlier proposal,[1,2] we do not observe significant nuclease function associated with the Tex YqgF domain. Our findings provide a structural foundation for understanding Tex function and can guide future studies on the structure and function of Spt6.

Results and Discussion

Structure determination and overall description

The full-length *P. aeruginosa* Tex protein was expressed recombinantly in *Escherichia coli* and purified by Ni-chelate, heparin affinity, and gel filtration chromatography. The C-terminal hexahistidine tag was retained for the structural and biochemical studies. The structure was determined by single-wavelength anomalous dispersion phasing and density modification using data collected to 2.7 Å resolution from a selenomethionine (Se-Met)-substituted crystal. This structure was refined against data collected from an isomorphous native crystal (crystal form I, 2.5 Å resolution) and from a second native crystal that belonged to the same space group but had substantially different cell dimensions (form II, 2.3 Å). The native structures were refined to R/R_{free} values of 24.0/27.4% (form I) and 22.1/26.6% (form II), with good geometry (Table 1). In both crystal forms, residues 1–730 are clearly observed in the electron density, with the exception of a short loop region (residues 246–251) in crystal form II. The 55 C-terminal residues of Tex and the hexahistidine tag are disordered and are not included in the final model.

Tex is ~53% α-helical and 10% β-sheet. The overall structure is notably flat and elongated, with approximate dimensions of 27 × 72 × 107 Å (Fig. 1). The most striking structural feature is a long, central helix (H15) spanning ~72 Å and comprising amino acid

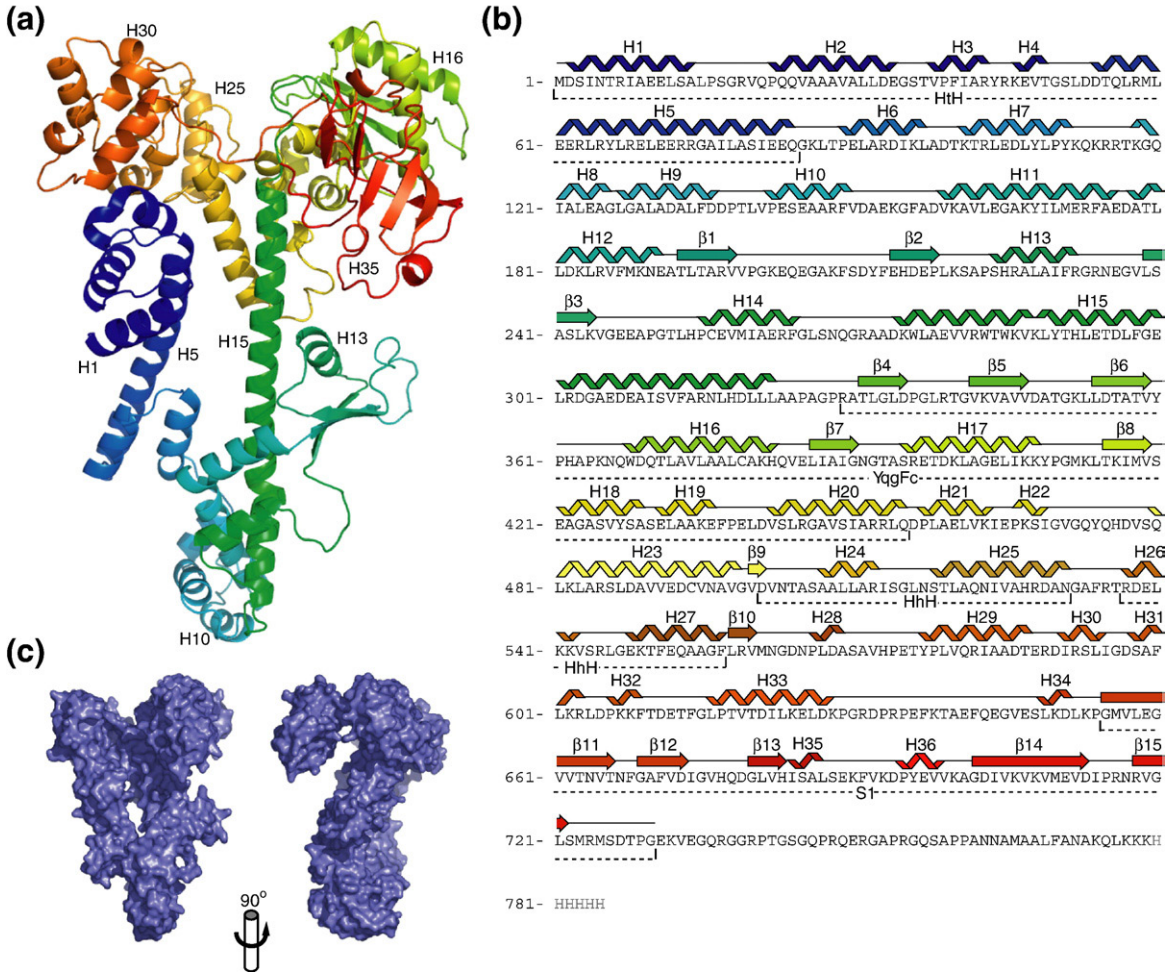


Fig. 1. Crystal structure of Tex. (a) Structure of Tex. The structure is colored from N-terminus (blue) to C-terminus (red). (b) Tex amino acid sequence. The observed secondary structure is indicated above the sequence. Identified domains are indicated below the sequence. (c) Tex surface representation in two orthogonal views. Left-hand view is the same as (a).

residues 274–322. The rest of the protein wraps around the central helix at both the N-terminal and C-terminal ends. The resulting structure has a distinctive question-mark-like appearance (Fig. 1c). The structure closely resembles a lower-resolution Tex structure that was recently submitted to the Protein Data Bank (PDB) by the New York Structural GenomiX Research Consortium (PDB accession code: 2OCE).

Although the Tex structure is fairly compact, it can be largely described as a series of distinct domains. The first 84 residues adopt an HtH structure (Fig. 2), although primary sequence analysis has not previously identified this motif in Tex. The next 189 residues (85–273) wrap around the bottom of the central helix (H15; residues 274–322), which extends nearly the full length of the structure, with a 30° kink occurring at lysine 288. Extending from the C-terminal end of this helix, the rest of the structure forms three domains that had been previously predicted from analysis of the amino acid sequence:[7] YqgF (residues 329–455), helix–hairpin–helix (HhH) (residues 501–557), and S1 (residues 654–730) domains. The S1 domain is tethered to the rest of the structure by a stretch of 25 residues (629–653) that traverses the top of the molecule and contains little defined secondary structure but is clearly ordered in the electron density.

HtH

Despite very low (<15%) sequence identity, the N-terminal HtH domain overlaps with other HtH structures (e.g., DNA helicase hel308, PDB accession code: 2P6U) with a root-mean-square deviation (RMSD) of 2.2 Å over 64 C^{α} atoms. HtH motifs typically bind double-stranded DNA (dsDNA) by inserting a helix into the major groove of the DNA duplex.[16] The observed conformation of the HtH region in Tex does not appear to be competent for canonical dsDNA binding because the DNA would significantly clash with the surrounding Tex structure. However, Tex may be capable of binding a single-stranded nucleic acid substrate through this domain. The strongest "hit" from the DALI search (z score = 5.6, RMSD = 2.2 Å) is to the ratchet domain of an archeal DNA helicase, Hel308.[17] Hel308 unwinds dsDNA and uses the third helix of the HtH motif to bind a single strand of the DNA. Superposition of the DNA-bound Hel308 complex (PDB accession code: 2P6R) onto the Tex structure places the single-stranded DNA (ssDNA) through a narrow, elongated cavity (~6×20 Å) that passes through the center of the Tex structure (Fig. 3). Although we currently have no direct evidence that Tex binds nucleic acid substrates in this region, there is nothing obvious from the Tex structure that would preclude single-stranded nucleic acid binding at this site in some contexts.

YqgF homologous domain

The central YqgF domain of Tex (residues 329–455) belongs to the YqgFc domain family (SMART SM00732[20]), a domain described as RNase H-like and typified by the *E. coli* protein YqgF.[2,7] Little functional or biochemical data are available for YqgF domain-containing proteins, even though these proteins are highly conserved and occur across a wide variety of bacterial genomes.[21] YqgF domain family

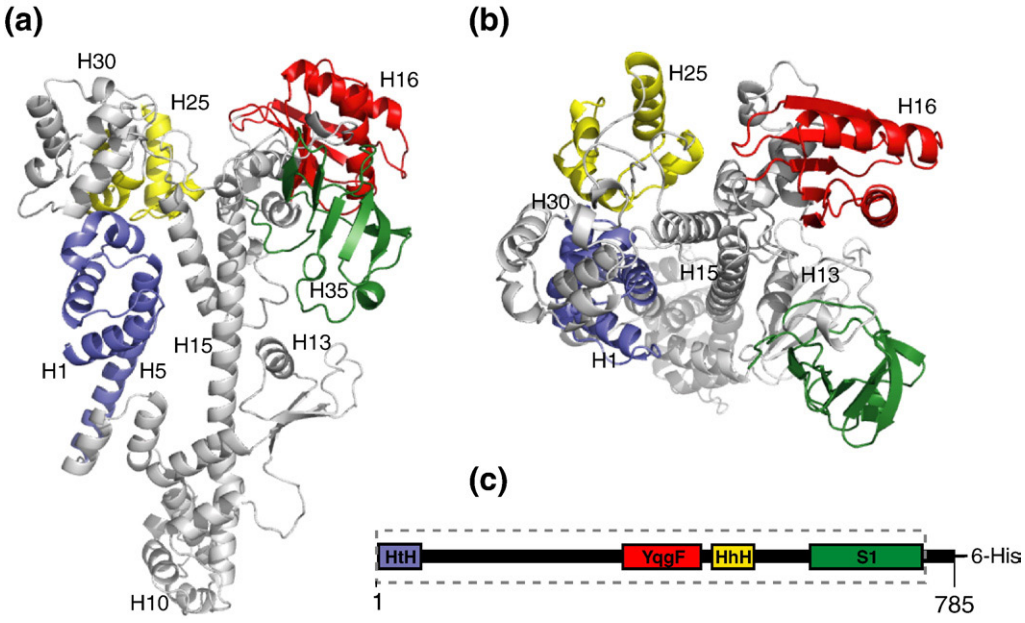


Fig. 2. Tex domain arrangement. (a and b) Orthogonal views of ribbon diagram of the Tex structure. Structural motifs identified from primary sequence and structural analyses are colored. Blue, HtH motif. Red, YqgF homologous domain. Yellow, tandem HhH motif. Green, S1 domain. (c) Tex domain structure, colored as in (a) and (b). The segment boxed with broken lines, comprising residues 1–730, indicates the region of Tex sequence observed in the crystal structure.

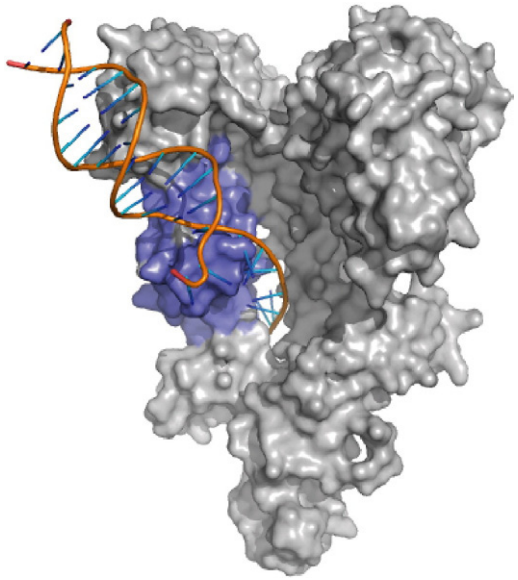


Fig. 3. Model for nucleic acid binding to Tex HtH motif. The archeal Hel308 structure (PDB accession code: 2P6R; only the DNA is shown here) was superimposed on the Tex structure (gray) by aligning HtH regions (blue). The path of the superimposed Hel308-bound ssDNA projects through a hole in the core of the Tex structure. The structure is oriented as in Fig. 1a.

proteins are predicted to be ribonucleases or resolvases based on homology to RuvC Holliday junction resolvases.[2,22,23] Like RuvC, which is structurally and biochemically well characterized, YqgF nuclease domains preserve the overall topology and the majority of structural and sequence elements characteristic of the RNase H fold.[2,24–26] The Tex YqgF domain maintains these core structural elements and aligns especially well with RuvC nucleases (PDB accession code: 1HJR, 120 C^{α}, RMSD = 3.0), although Tex does not appear to possess nuclease activity (see below).

HhH

The Tex structure contains two adjacent HhH motifs, comprising residues 502–531 and 537–557, both of which were predicted from earlier sequence analysis (Fig. 4).[5,6] The two HhH motifs are related by an approximately 90° rotation with respect to each other and pack together through extensive conserved hydrophobic interactions to generate a single, compact unit called an $(HhH)_2$ domain.[27] In contrast to a canonical $(HhH)_2$ structure in which the two HhH motifs are connected by an extra helix, the Tex $(HhH)_2$ domain makes the connection with a short five-residue loop. The typical function of $(HhH)_2$ domains is to bind dsDNA, mediated through nonspecific interactions with nitrogen atoms in the protein backbone and oxygen atoms in the DNA phosphate groups.[27] In Tex, the binding face of this domain lies on the surface of the structure and would be accessible to a potential nucleic acid substrate.

S1 domain

The Tex S1 domain (Fig. 5) adopts the canonical topology characteristic of the S1 RNA-binding domain family.[28] First identified as a motif of the ribosome essential for translational initiation, S1 domains are ubiquitous and found primarily in proteins that bind RNA and/or have nuclease activity.[29,30] The Tex S1 domain adopts the overall five-stranded antiparallel β-barrel topology representative of the ubiquitous oligonucleotide/oligosaccharide binding (OB) fold. OB-fold proteins, including S1 domains, present a common binding cleft for interaction with a variety of different ligands, the most common being nucleic acids.[3,30] This cleft runs perpendicular to the axis of the β-barrel where nucleic acids almost always bind with common polarity. The Tex S1 domain contains a short 3_{10} helix (H35) adjacent to the binding cleft that, along with a strong preference for ssRNA, distinguishes S1 domains from other OB-fold proteins.[31,32]

Evidence for flexibility

Comparison of the two crystal forms, which differ by 25 Å in *b*-axis length and each contain one molecule in the asymmetric unit, indicates that Tex displays flexibility in the disposition of its C-terminal S1 domain. The two Tex structures superimpose closely throughout (RMSD = 1.596 over 510 C^{α} atoms), with the exception of a small rotation in the YqgF domain and a 14-Å displacement of the S1 domain (Fig. 6). The S1 domain rearrangement is accomplished by rotation in the loop that tethers the domain to the rest of the Tex structure. C-terminal to the S1 domain, there is no discernable electron density for residues 731–785 in either crystal form,

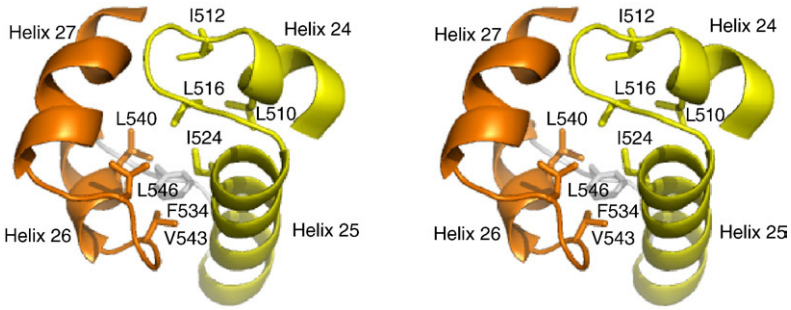


Fig. 4. The Tex $(HhH)_2$ domain stereoview. Tandem HhH motifs (yellow and orange) are linked by a short loop (gray) and pack together to form a single $(HhH)_2$ domain. Conserved hydrophobic residues that comprise the core of the HhH packing surface are indicated. The view is looking down on the surface that binds dsDNA in other structures.

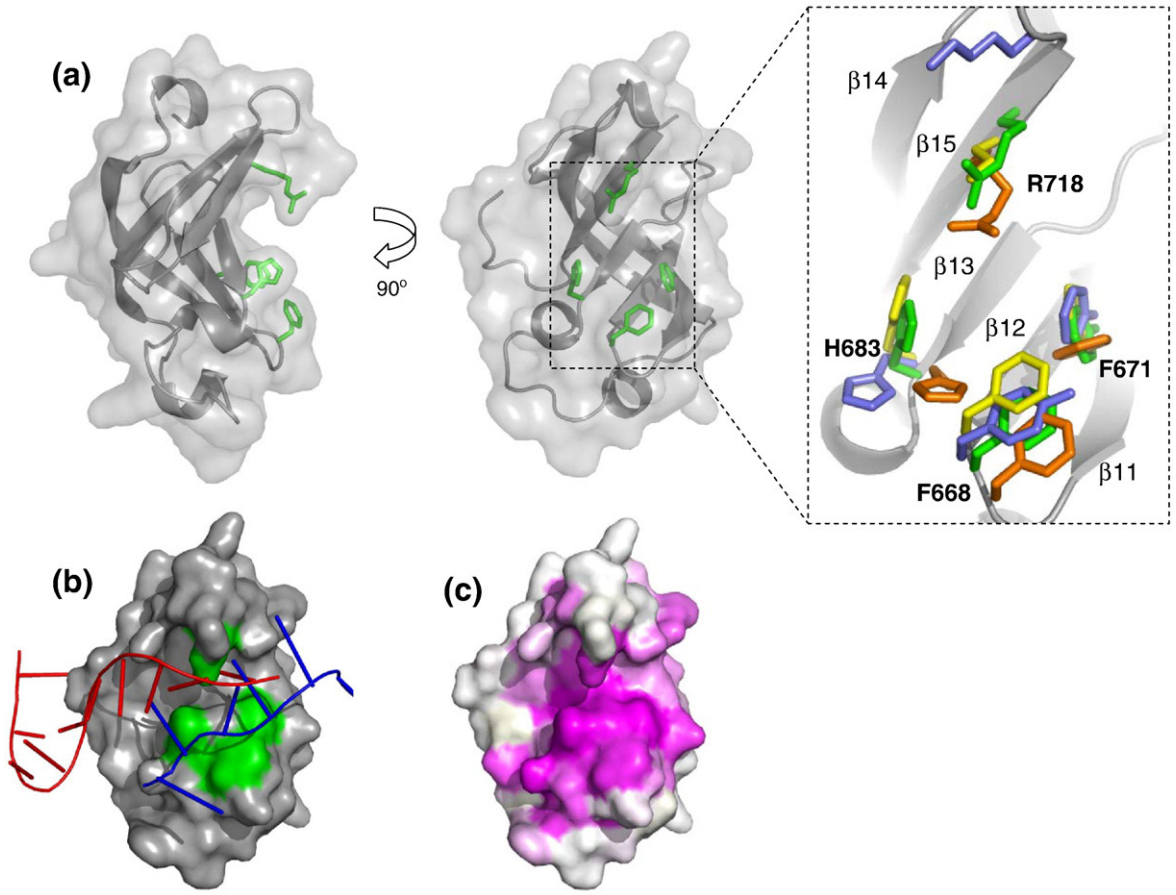


Fig. 5. The Tex S1 domain. (a) Two orthogonal cartoon representations with conserved residues R718, H683, F668, and F671 shown in green. Right, a close-up view of an alignment of conserved residues from structurally related S1 domains: green, Tex; orange, PNPaseS1 (PDB accession code: 1SRO); yellow, archealRPB4/7 (PDB accession code: 1GO3); blue, archealIFα (PDB accession code: 1ZY6). (b) A hypothetical model for RNA binding to the S1 domain binding cleft. The crystal structures of S1 domains with bound RNA from RNase E (PDB accession code: 2COB, red) and RNase II (PDB accession code: 2IX1, blue) were aligned with Tex S1 using DALI.[18] As illustrated, ssRNA binds the same face of the different S1 domains but considerable differences in detail are apparent. (c) Surface representation showing primary sequence conservation as assigned by the ConSurf server.[19] Conservation is indicated as a gradient from magenta (high) to white (low). In contrast to the view shown here, minimal conservation is observed on the opposite face of the S1 domain.

indicating a high degree of flexibility. In addition, very few contacts are observed in either structure between the S1 domain and the rest of the Tex protein. These observations indicate that the S1 domain is relatively unrestrained and able to adopt a range of orientations in solution.

Tex binds oligonucleic acids

Although the precise function of Tex is not known, the structural motifs observed in the Tex structure suggest binding to nucleic acids. It has recently been demonstrated by Southwestern and Northwestern analysis that recombinant *S. pneumoniae* Tex can interact with RNA and DNA,[3] although Tex–nucleic acid interactions have not been explored in detail. We therefore quantified the ability of Tex to bind various nucleic acid substrates using electrophoretic mobility shift assays. Random 25-mer ssDNA, dsDNA, ssRNA, and double-stranded RNA (dsRNA) sequences were tested for binding (Fig. 7). Tex bound all four types of nucleic acids, with a strong preference for ssRNA ($K_d = 210 \pm 50$ nM, $n = 9$), and binding affinities were not altered by the presence of Mg^{2+}. Binding to dsDNA ($K_d = 3720 \pm 150$ nM, $n = 3$), dsRNA ($K_d = 4200 \pm 147$ nM, $n = 4$), and ssDNA ($K_d = 5100 \pm 150$ nM, $n = 4$) was more than 10-fold weaker (Table 2). Binding to an RNA/DNA hybrid was also confirmed, though not quantified. Because Tex binds RNA and DNA chosen at random, including a poly-U ssRNA sequence (data not shown), binding appears to be sequence independent, although the possibility of some strong sequence preferences cannot be excluded at this time.

To supplement our electrophoretic mobility shift data and provide a solution-based estimate of binding, we used fluorescence polarization (FP). The K_d for 25-nt ssRNA binding to Tex was 56.6 ± 6.2 nM ($n = 5$). The approximately fourfold decrease in K_d compared to the value determined by electrophore-

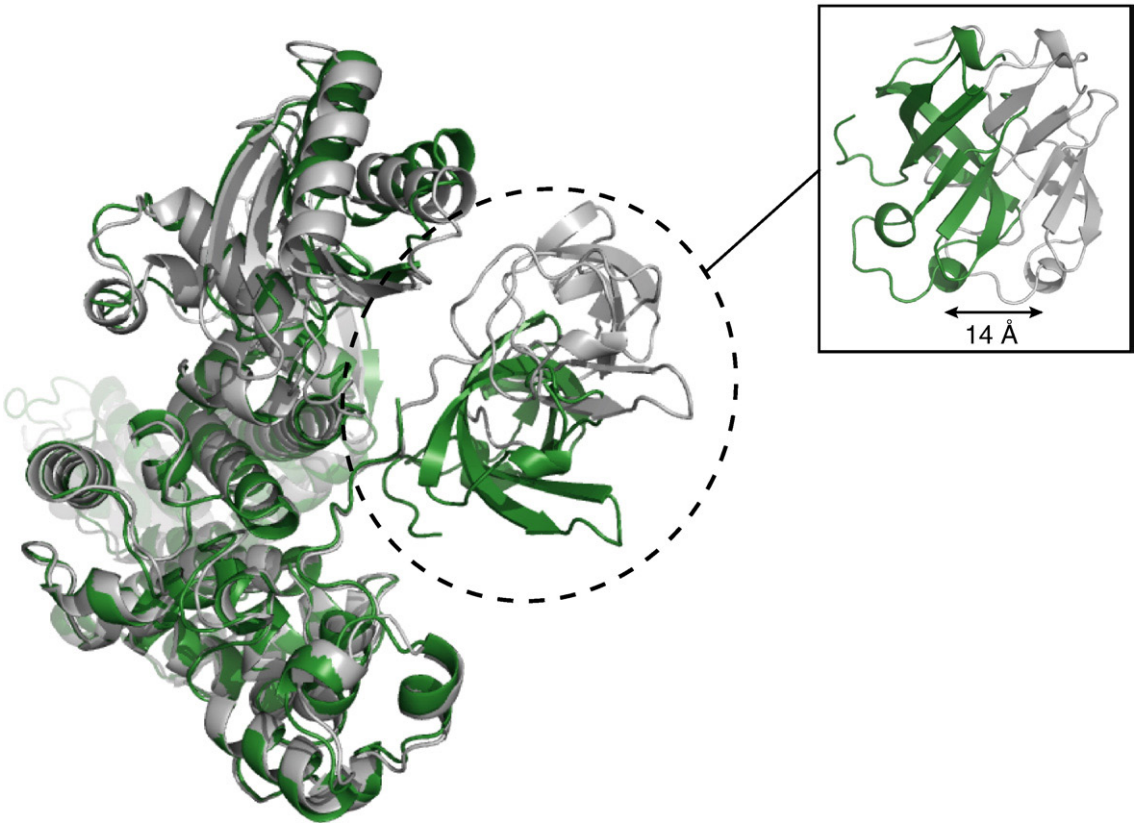


Fig. 6. S1 domain mobility. Significant rearrangement of the S1 domain is observed in different crystal forms of the Tex structure. The 2.5-Å (crystal form I, gray) and 2.3-Å (crystal form II, green) crystal structures are superimposed. A side view of the superimposed crystal structures (inset) highlights a 14-Å shift between the two S1 domains.

tic mobility shift likely results from differences in equilibrium considerations in solution studies *versus* gel-based electrophoretic methods.[33]

The length dependence of ssRNA binding was examined using 10-mer, 13-mer, 16-mer, 20-mer, and 25-mer sequences. Negligible differences in binding were observed between the 20-mer and 25-mer sequences, and a modest reduction in binding was observed for the 16-mer sequence. In contrast, the 13-nt ssRNA bound with markedly reduced affinity, and binding was not detected with the 10-mer sequence.

Our electrophoretic mobility shift assays reveal two distinct, shifted bands for ssRNA and ssDNA (Fig. 7). Although this may be an artifact of the native gel electrophoresis, it may also suggest non-stoichiometric binding or protein multimerization during the binding event. Studies of other proteins, such as RNase E and PNPase, have shown that S1-domain-mediated multimerization may be critical for substrate binding and enzymatic activity.[31,34] We therefore investigated Tex RNA-binding stoichiometry using FP and gel filtration. Apo Tex elutes from a gel filtration column as a single peak with a retention time expected for a monomer, even at a very high concentration. Similarly, a Tex:ssRNA complex, prepared by premixing Tex and an excess of 20 nt ssRNA, elutes from the sizing column as a single peak with the retention time expected for a 1:1 stoichiometry. FP was also used to estimate stoichiometry by titrating Tex protein into a solution containing a saturating concentration (>20-fold above K_d) of 25 nt ssRNA. Polarization values (P) were read at each Tex concentration, and the values were plotted as P *versus* the molar ratio of Tex to ssRNA. The inflection in this plot represents the point at which Tex has saturated all the binding sites on the RNA substrate. The inflection point for this experiment occurred at a molar ratio of 1:1, indicating that a single molecule of Tex binds one 25-nt ssRNA molecule (data not shown).

Oligonucleic acids bind the Tex S1 domain

Tex's preference for binding ssRNA made the S1 domain an obvious candidate for mediating binding of nucleic acid. In support of this possibility, we found that protein lacking the S1 domain (TexΔS1) was unable to bind ssRNA, dsRNA, ssDNA, or dsDNA in our electrophoretic mobility shift assays (Table 2). It is unlikely that the loss of nucleic acid binding is due to misfolding because the TexΔS1 protein behaves similarly to the full-length protein throughout the purification process, including gel filtration. Furthermore, significant repositioning of the S1 domain (with respect to the rest of the protein) in different Tex crystal forms has little effect on the rest of the Tex structure. These data indicate that although Tex binds a variety of nucleic acids

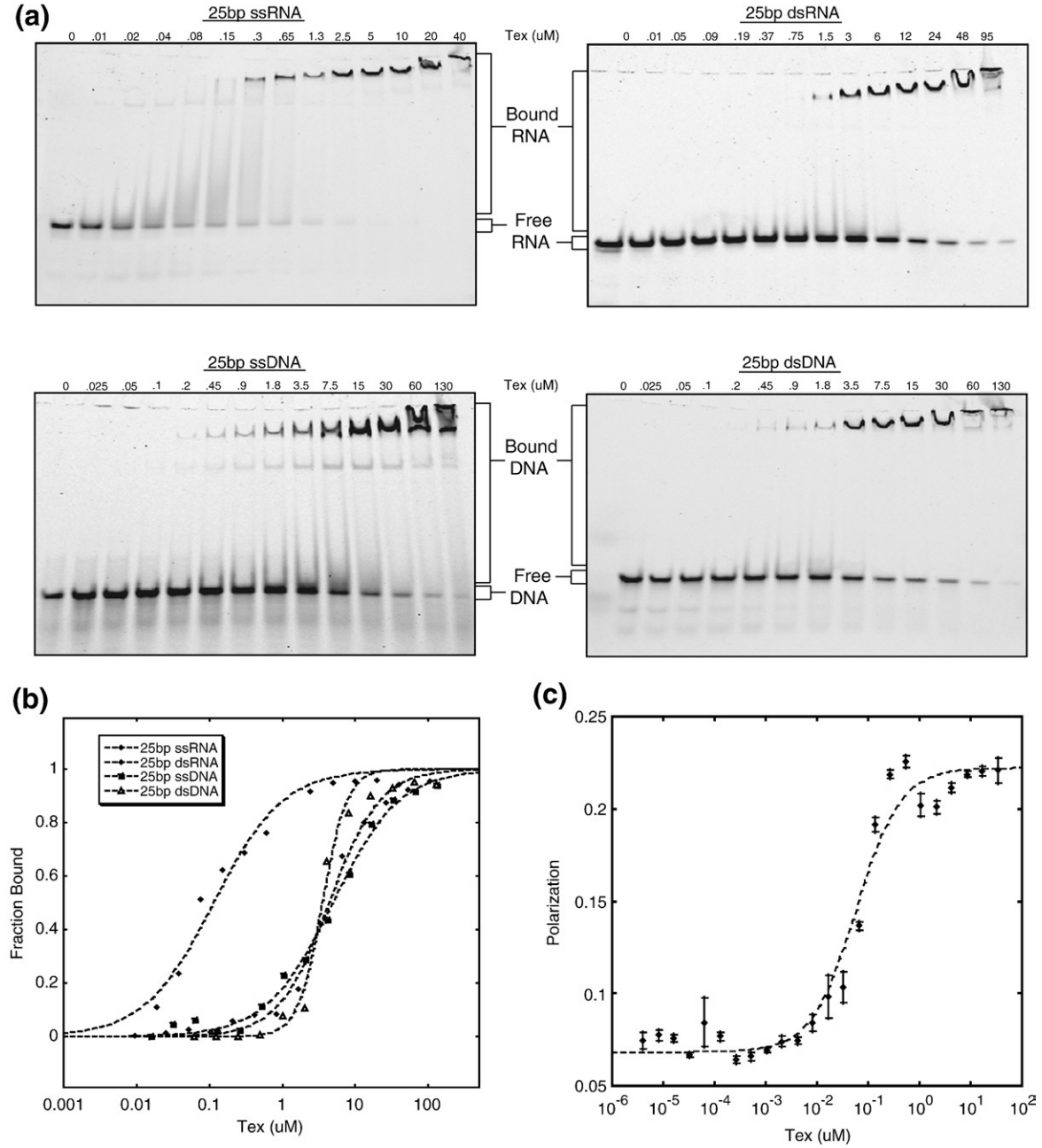


Fig. 7. Electrophoretic mobility shift data for Tex binding different substrates. (a) Full-length Tex protein was added in increasing concentrations to 5′-fluorescein-labeled ssRNA (top left), dsRNA (top right), ssDNA (bottom left), or dsDNA (bottom right). Nucleic-acid-bound Tex complexes were resolved from free substrate by native gel electrophoresis. (b) Representative binding isotherms for the gel shifts represented in (a). Shifts and respective isotherms were repeated at least three times for each substrate. Resulting K_d values with standard error (in micromolar) were 0.21±0.05 for 25 bp ssRNA, 4.2±1.47 for 25 bp dsRNA, 5.1±1.5 for 25 bp ssDNA, and 3.7±0.15 for 25 bp ssDNA. (c) FP binding isotherm for Tex binding to fluorescein-labeled 25 bp ssRNA. Data points with error bars (standard error) representing average with polarization values (P) were measured for increasing concentrations of Tex. K_d (±standard error) for Tex binding ssRNA based on FP experiments is 0.057±0.006 μM.

and is composed of a number of putative nucleic acid binding motifs, binding does not occur in the absence of the S1 domain.

In the numerous S1/OB-fold complex structures determined to date, nucleic acid substrate is coordinated in the binding cleft via surface-exposed aromatic side chains such as phenylalanine and more polar groups such as lysine or arginine.[30,35] The structure of Tex reveals the characteristic S1/OB binding cleft and candidate contact residues: F668, F671, H683, and R718 (Fig. 5). When using DALI[18] to align the Tex S1 domain with other S1 structures, Tex residues F668, F671, H683, and R718 superimpose closely with comparable residues in the S1 domain RNA-binding clefts of PNPase (PDB accession code: 1SRO, 73 C^{α}, RMSD=2.0 Å), aIFα (PDB

Table 2. Summary of binding affinities and relative binding affinity of Tex and Tex mutants based on gel mobility shift analysis

Protein	K_d[a] (μM)	Percentage of binding relative to WT
25 nt ssRNA		
WT Tex	0.21±0.05	100
ΔS1 Tex	nd	<0.1
R718E	28.4±11.4	0.74±0.30
F668D/F671D	13.7±0.7	1.53±0.08
F668D	15.5±5.5	1.35±0.48
F671D	9.8±1.7	2.14±0.37
H683E	10.7±3.5	1.96±0.64
25 bp dsRNA		
WT Tex	4.2±1.47	100
ΔS1 Tex	nd	<0.1
R718E	46.9±6.3	8.96±1.20
F668D/F671D	21.2±11.5	19.8±10.6
25 bp dsDNA		
WT Tex	3.7±0.15	100
ΔS1 Tex	nd	<0.1
R718E	48.3±5.2	7.66±0.82
F668D/F671D	20.4±3.8	18.1±3.38
25 nt ssDNA		
WT Tex	5.1±1.5	100
ΔS1 Tex	nd	<0.1
R718E	118.7±10.7	4.30±0.39
F668D/F671D	55.3±6.0	9.22±1.00

nd, no binding detected.
[a] Values represent the average K_d from multiple experiments ($2 \le n \le 9$)±standard error.

accession code: 1YZ6, 79 C^{α}, RMSD = 1.8 Å), and the archeal homolog of the RNAPII subunit RPB7 (PDB accession code: 1GO3, 72 C^{α}, RMSD = 1.8 Å) (Fig. 5). These proteins are all reported to bind RNA transcripts in a sequence-nonspecific manner with the interaction being important for RNA decay,[34] general translation initiation,[36] and transcription initiation.[37,38] Given the close alignment of these conserved residues, it is likely that cellular substrates for Tex are also sequence-nonspecific RNA transcripts.

Although putative RNA-binding residues align well among some S1 proteins, precise alignment of critical residues does not appear to be required for binding similar RNA substrates. The crystal structures of the ribonucleases RNase E (PDB accession code: 2C0B) and RNase II (PDB accession code: 2IX1) represent costructures of RNA-bound S1 domains that interact with RNA independent of sequence. S1 residues making significant contacts in these structures do not align precisely with one another even though the general composition of side chains is maintained and the interaction occurs across the same S1/OB binding cleft. When aligning Tex S1 with these structures, a similar theme is observed, where F668, F671, H683, and R718 are clustered in the same binding cleft as equivalent RNase E and RNase II residues. Based on these observations, we propose a model for Tex binding to ssRNA via the S1/OB binding cleft (Fig. 5b).

In order to map the nucleic acid binding surface further and to test the binding model, we assayed a variety of Tex S1 domain point mutants for RNA binding. Mutation of the conserved S1 binding cleft residues F668, F671, H683, and R718 to aspartate or glutamate disrupts ssRNA binding by at least 46-fold [<2.2% binding relative to wild-type (WT) Tex] (Table 2), with a 135±50-fold reduction (0.74±0.30% of WT Tex) in ssRNA binding when the mutant R718E was assayed by gel shift. The double mutation F668D/F671D results in a 65±3-fold reduction (1.53±0.08% of WT Tex) in ssRNA binding and implies that hydrophobic base-stacking and/or packing interactions on the S1 surface are additionally important. Consistent with the model that all nucleic acid binding in our assay conditions is to the S1 domain and that ssRNA is the preferred ligand, affinities for all substrates were reduced in the S1 point mutants, with the greatest effect seen on binding of ssRNA.

Overall, the structural and binding data indicate that the S1 domain is a highly dynamic module that is required for nucleic acid binding and displays a strong preference for ssRNA. Binding is likely to be sequence independent, as has been found for the majority of other described S1-domain-containing proteins, including those that align well with Tex. Although S1 is the primary nucleic acid binding domain, we cannot rule out minor contributions to binding from other regions, and it is possible that specific *in vivo* contexts, such as binding to another partner, might open additional surfaces for binding to nucleic acid substrates.

Consideration of minimal oligonucleotide length required for Tex binding suggests that S1 is not the only region that binds the various nucleic acid substrates. Ten-nucleotide oligonucleotides do not bind Tex in our assay, and a significant decrease in binding affinity was observed for ssRNA lengths less than 20 nt. An ssRNA molecule of 10 nt or longer would extend beyond the available binding surface of the S1 domain, implying that additional contacts outside the S1 domain occur in our assay. This is a familiar theme for S1-containing proteins. For example, the RNase E S1 domain is a dynamic module that serves as a molecular clamp for correctly orienting ssRNA substrate.[35] Given the modularity of the Tex S1 domain, the presence of multiple other nucleic acid binding domains, and a minimal substrate length that spans a surface larger than that offered by the S1 domain, one attractive possibility is that Tex may utilize an S1 molecular clamp binding model similar to RNase E.[35]

Putative nuclease activity

Based on the presence of the RNase H fold YqgF domain and the observation that Tex negatively regulates transcription when overexpressed, Tex is predicted to have ribonuclease activity.[1,2] In conflict with this prediction, however, we have not detected nuclease activity. There is no indication of nucleic acid degradation in our gel shift experiments. We performed a qualitative nuclease assay in which the positive controls (RNase T1 and micrococcal nuclease) were active, but activity was at background levels for Tex (data not shown). In addition, we

observed no differences in nucleic acid binding or the very low background nuclease activity between WT Tex and a double mutant in which two putative catalytic residues in the YqgF domain (Asp335 and Glu421) were changed to alanine (data not shown). The lack of nuclease activity could be explained by the fact that the Tex YqgF domain lacks a critical and highly conserved carboxylate residue that is required for metal coordination in known RNase H fold ribonucleases (e.g., D141 of *E. coli* RuvC) (Fig. 8). Therefore, in contrast to the earlier prediction,[1,2] these data suggest that Tex is not a ribonuclease.

Implications for Spt6

Many S1-domain-containing proteins are factors involved in general processes such as transcription, translation initiation, and mRNA decay.[28,31,36,37] Consistent with this idea, comparative genomic and evolutionary studies suggested that Tex represents a bacterial ortholog of the eukaryotic transcription elongation factor Spt6.[5,7] Although Spt6 is twice the size of Tex, it is predicted to possess YqgF, HhH, and S1 domains in the same order and to possess 15% sequence identity (27% similarity) over these regions (Fig. 9a). The predicted secondary structural elements of the Spt6 sequence also show good agreement with the Tex structure (data not shown).

The nature of the structural similarity between Tex and Spt6 is further clarified by aligning the Spt6 sequence with the Tex structure. In particular, all of the differences in sequence length between Tex and Spt6 (e.g., Spt6 insertion sequence) occur on the surface of the Tex structure, which appears able to accommodate additional sequences without disrupting the core structural scaffold (data not shown). Additionally, evolutionary conservation scores based on Tex- and Spt6-related sequences were assigned for each Tex amino acid residue and mapped onto the Tex structure using the ConSurf server.[19] Most of the conserved residues identified by this method appear to be involved in packing interactions in the Tex structure. These observations suggest that Spt6 retains the core Tex structure, with variations on the periphery, along with additional domain features at the N- and C-terminal ends (described below).

Unlike Tex, Spt6 possesses an SH2-like domain C-terminal to the S1 domain.[39] Given the proximity of the SH2-like sequence to the S1 domain, it is likely that this domain lies along one face of the Spt6 structure, as indicated in Fig. 9. The Spt6 SH2-like domain is reported to mediate interactions with the C-terminal domain of RNAPII.[12] Based on the mobility of the S1 domain observed in the Tex structures, it is likely that Spt6 binds RNAPII via a flexible tether.

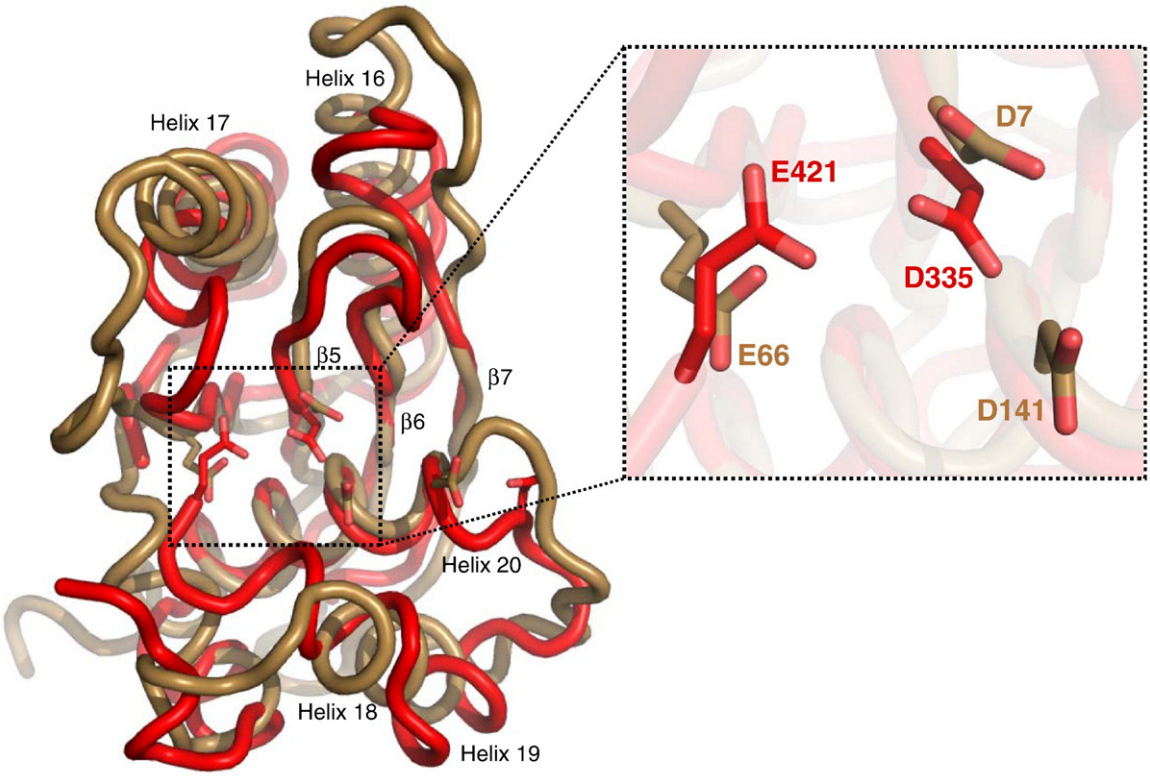


Fig. 8. The Tex Yqgf (red) superimposed on the *E. coli* RuvC (sand) (PDB accession code: 1HJR, 120 C^{α}, RMSD = 3.0) and an exploded view of the catalytic center. The RuvC catalytic residues D7, E66, D141, and D138 are shown in stick representation aligned with Tex residues D335, D421, and D441 that share the same basic geometric orientation in the catalytic center. Although three of the four conserved catalytic residues are present in Tex, there is no equivalent acidic residue present at the location corresponding to RuvC D141.

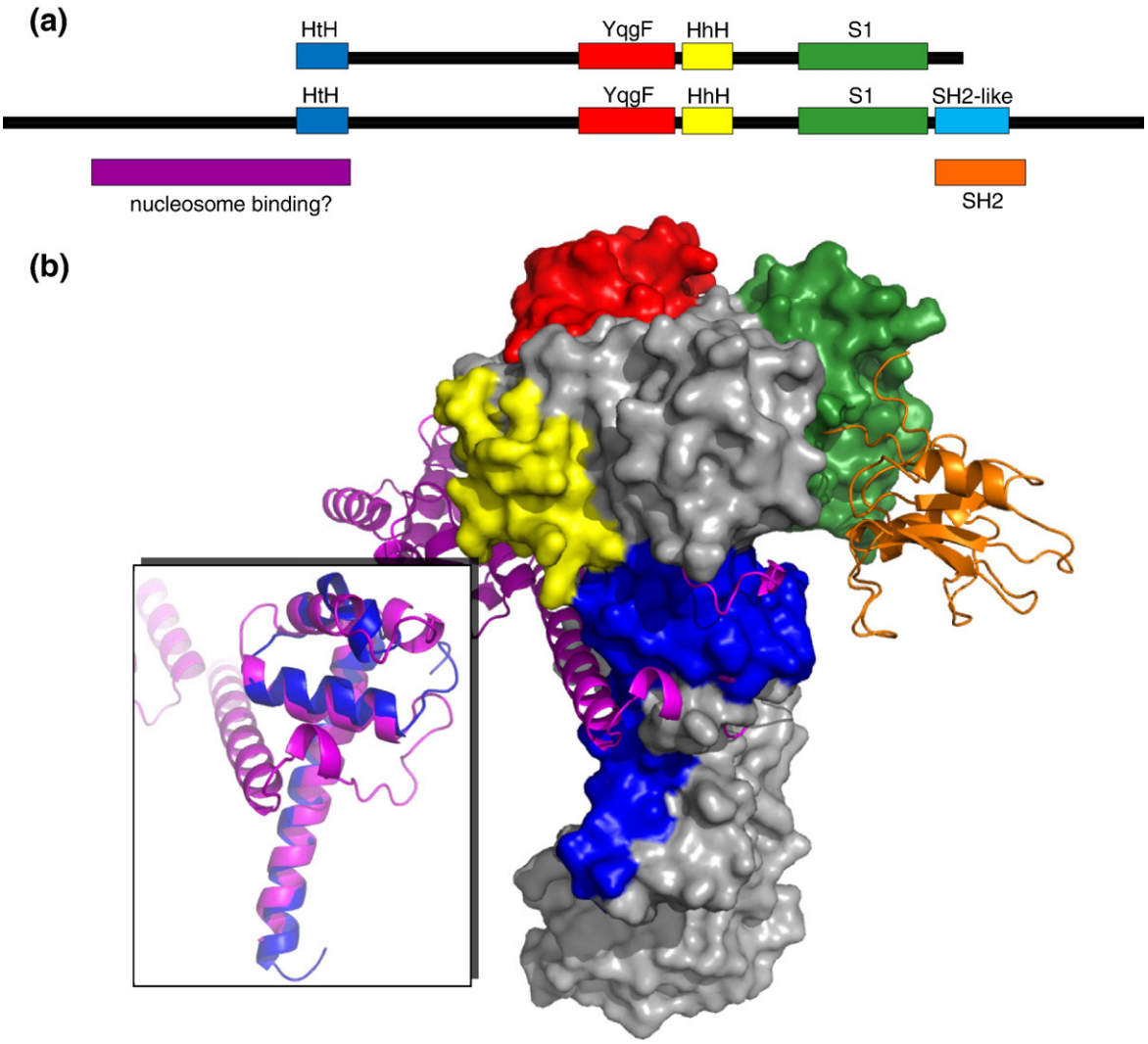


Fig. 9. Model for Spt6 structure. (a) Comparison of Tex (top) and Spt6 (bottom) domain structures. (b) The Tex structure is used to model the central portion of the Spt6 structure (surface representation). A proposed nucleosome-binding domain (magenta, inset) is modeled based on structural alignment with the C-terminal portion of the ISWI nucleosome interacting domain (PDB accession code: 1OFC). An SH2-like domain (orange, PDB accession code: 1PIC) is modeled at the C-terminal end of the S1 domain.

Structural comparisons suggest that a region toward the Spt6 N-terminus may possess histone chaperone activity. A DALI search of the Tex HtH domain identified significant structural homology (z score = 5.2, RMSD = 2.4 Å) with the nucleosome-binding SLIDE domain of ISWI (PDB accession code: 1OFC).[40] While there is insufficient sequence for Tex to form a complete SLIDE-like nucleosome-binding domain, Spt6 contains an additional sequence N-terminal to the region of Tex homology that could potentially fulfill this role. In this model, histone chaperone activity would be located on the face of Spt6 opposite from the S1 and SH2-like domains (Fig. 9).

Spt6 has demonstrated eukaryotic exosome-recruiting faculties,[12] and we have observed that Tex copurifies with RNase E and PNPase, which, in *E. coli*, are components of the RNA degradosome (I.V.-G. and S.L.D., unpublished data). This may at least partially explain why Tex appears to negatively effect transcription when overexpressed[1,2] but does not itself appear to possess ribonuclease activity in our assays; it may be coordinating the recruitment, or influencing the activities, of degradosome-associated ribonucleases. Further, Spt6 interacts with an elongating RNAP at the C-terminal domain based on elongation-specific phosphorylation.[12] The observation that Tex from *P. aeruginosa* copurifies with components of RNAP (I.V.-G. and S.L.D., unpublished data) suggests that it may be associated, either directly or indirectly, with the transcription machinery; functional parallels may therefore exist between Spt6 and Tex.

In summary, the Tex crystal structures reveal an elongated helical protein comprising several putative nucleic acid binding domains. Biochemical characterization revealed that Tex binds ssDNA, dsDNA, ssRNA, and dsRNA substrates with a preference for ssRNA, with a primary interface being mediated by interactions along the canonical OB-fold binding

cleft of the Tex S1 domain. The Tex structure provides a model for the core of the eukaryotic transcription factor Spt6 and raises the possibility that the N-terminal portion of Spt6 constitutes a nucleosome-binding domain that evolved from an HtH domain.

Materials and Methods

Tex protein expression and purification

Full-length Tex from *P. aeruginosa* strain PAO1 was cloned into a pET24 *kan* expression vector containing a C-terminal hexahistidine tag. The plasmid was transformed into cells of *E. coli* BL21-codonplus-(DE3)-RP (Stratagene). Cells were grown in LB media and induced with 1 mg/ml IPTG at an OD_{600} of 0.6 or alternatively grown using an autoinduction method as described in Ref. 41. In both cases, cells were grown at 37 °C for 5 h and then transferred to 20 °C and grown to saturation. Harvested cells were stored at −80 °C.

Cells were thawed and resuspended in lysis buffer [50 mM Tris, pH 7.5, 500 mM NaCl, 10% glycerol, 10 mM imidazole, and 2 mM β-mercaptoethanol (BME)] in the presence of lysozyme and protease inhibitors. Following sonication and centrifugation (25,000*g*), the soluble fraction was applied to Ni-NTA agarose resin (Qiagen) and eluted with 300 mM imidazole. Protein was dialyzed (50 mM Tris, pH 7.0, 10% glycerol, 2 mM BME, and 35 mM NaCl), applied to a heparin column (5 ml HiTrap Heparin HP, GE Healthcare), and eluted over a NaCl gradient. ΔS1 Tex was purified in the same way, but a Q ion-exchange column (5 ml HiTrap Q HP, GE Healthcare) was used instead of the heparin column as the ΔS1 construct did not bind heparin. Peak fractions from either heparin or Q columns were pooled, dialyzed (50 mM Tris, pH 7.5, 5% glycerol, 100 mM NaCl, and 2 mM BME), and run over a size-exclusion column (Superdex 200 26/70; GE Healthcare). Se-Met-substituted Tex protein was expressed[42] and purified using the same protocol as native protein.

Tex crystallization and structure determination

Crystals (Se-Met and crystal form I) were grown by sitting drop vapor diffusion in 19% (w/v) polyethylene glycol 3350, 0.1 M Bis–Tris, pH 5.5, 0.17 M ammonium sulfate, and 10% (v/v) glycerol. A second crystal form (crystal form II) was grown in 18% w/v polyethylene glycol 4000 and 100 mM sodium acetate, pH 4.6. Data were collected at the National Synchrotron Light Source (NSLS) at Brookhaven National Lab for crystal form I and on a home source (Rigaku Raxis IV) for crystal form II (Table 1). X-ray diffraction data were processed using HKL2000.[43]

Phases were determined for the form I crystals by the single-wavelength anomalous dispersion method using Se-Met-substituted Tex. The programs SOLVE[44] and RESOLVE[45] were used to identify selenium positions (12 out of 13 potential sites were identified) to calculate initial maps to 3.4 Å resolution followed by phase extension to 2.7 Å. Data from native crystals grown in the same conditions (crystal form I) were subsequently collected to extend the resolution of the Tex structure to 2.5 Å. A 2.3-Å structure (crystal form II) was determined using crystal form I as a model for molecular replacement in PHASER.[46] Refinement was performed using CNS,[47] Phenix,[48] and Refmac.[49] O,[50] Coot,[51] and MolProbity[15] were used for model building and validation. PyMOL[52] was used to prepare figures, superimpose the Tex structures, and perform electrostatic calculations (APBS tools). The ConSurf server[19] was used to calculate evolutionary-based conservation scores.

Nucleic acid binding assays

Oligonucleotides were designed using random sequence and purchased from the University of Utah DNA/Peptide Core facility. Oligonucleotide sequences used in this study are the following (5′ to 3′): UCUUUUCCUGUGUUUUUCCGCAAUC (25 nt ssRNA and 25 bp dsRNA, sense), GATTGCGGAAAAACACAGGAAAAGA (25 nt ssDNA), CGCAGGCCCGGCGCGAGGCCGAGGG (25 bp dsDNA, sense), UCCUGUGUUUUUCCGCAAUC (20 nt ssRNA), UUGUUUUUCCGCAAUC (16 nt ssRNA), UUUUUCCGCAAUC (13 nt ssRNA), and UUCCGCAAUC (10 nt ssRNA). Prior to binding studies, oligonucleotides were gel purified on a 20% acrylamide/7 M urea denaturing gel. Double-stranded substrates were mixed in equimolar amounts in 10 mM Tris–HCl, pH 7.5, and 40 mM KCl with one strand being end labeled with fluorescein. Samples were annealed by boiling (5 min) and slow (2 h) cooling to room temperature and gel purified using a nondenaturing 20% acrylamide gel.

Gel mobility shift assays were performed by mixing varying concentrations of protein with nucleic acid substrate in binding buffer [15 mM Tris–HCl, pH 7.5, 100 mM NaCl, 10% glycerol, and 1 U/μl RNasin (Promega) or 1 U/μl RNaseOUT (Invitrogen)]. The final concentration of nucleic acid used in each experiment was at least 20-fold below the K_d for the respective substrate/protein. Reactions were incubated at room temperature for 30 min and then assayed by electrophoresis using 4–20% TBE native gels (Bio-Rad Laboratories) at room temperature. Gels were imaged and quantified using a TYPHOON imaging system with ImageQuant software (GE Health Sciences). The fraction bound was calculated by quantifying the RNA/DNA_{total} (total fluorescence in entire lane) and RNA/DNA_{free}. All RNAs/DNAs of slower mobility than the RNA/DNA_{free} were considered bound. The fraction bound = $1-([RNA/DNA]_{free}/[RNA/DNA]_{total})$. Dissociation constants (K_d values) were calculated by plotting data points and curve fitting using the Hill formalism where fraction bound = $1/(1+(K_d^n/[P]^n))$. Average K_d values were determined by fitting data points from individual experiments and then averaging the calculated dissociation constants. All plots and curve fits were performed using the program KaleidaGraph (Synergy Software).

FP

FP was performed in 96-well format using a Tecan fluorimeter with the same fluorescein-labeled substrates and buffer conditions as the shift analysis. Tex concentrations were varied in individual wells and mixed with RNA at a final concentration that was at least 10-fold below the K_d. Samples were incubated at room temperature for at least 30 min prior to measuring polarization. Polarization values (P) were measured and plotted as a function of Tex concentration. Data points were fit using $P=((P_{bound}-P_{free})\ [Tex]/K_d+[Tex])+P_{free}$. The free and total protein concentrations are assumed to be equal because the RNA concentration is at least 10-fold lower than the K_d.

When using FP to evaluate binding stoichiometry, polarization measurements were performed following the procedure described in Ref. 53. RNA was mixed in

binding buffer solution at a concentration 20-fold above the K_d. Protein was titrated into the RNA solution until polarization values leveled off, and polarization values were plotted as a function of the concentration ratio of Tex protein *versus* RNA substrate. The Tex/RNA ratio where an inflection in the data occurs represents the binding stoichiometry, as this is the point where Tex has been saturated by RNA and polarization values change modestly as protein concentration is increased.

PDB accession numbers

Coordinates and structure factors for Tex crystal form I and crystal form II have been deposited in the PDB with accession numbers 3BZC and 3BZK, respectively.

Acknowledgements

We thank Brenda Bass and members of the Bass Lab for use of materials and reagents as well as help with binding studies. This work was supported by National Institutes of Health Grants GM074368 (S.J.J.), T32-GM008537 (D.C.), AI057754 (S.L.D.), and GM076242 (C.P.H.). Financial support for use of the 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.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2008.01.096

References

1. Fuchs, T. M., Deppisch, H., Scarlato, V. & Gross, R. (1996). A new gene locus of *Bordetella pertussis* defines a novel family of prokaryotic transcriptional accessory proteins. *J. Bacteriol.* **178**, 4445–4452.
2. Aravind, L., Makarova, K. S. & Koonin, E. V. (2000). Survey and summary: holliday junction resolvases and related nucleases: identification of new families, phyletic distribution and evolutionary trajectories. *Nucleic Acids Res.* **28**, 3417–3432.
3. He, X., Thornton, J., Carmicle-Davis, S. & McDaniel, L. S. (2006). Tex, a putative transcriptional accessory factor, is involved in pathogen fitness in *Streptococcus pneumoniae*. *Microb. Pathog.* **41**, 199–206.
4. Potvin, E., Lehoux, D. E., Kukavica-Ibrulj, I., Richard, K. L., Sanschagrin, F., Lau, G. W. & Levesque, R. C. (2003). *In vivo* functional genomics of *Pseudomonas aeruginosa* for high-throughput screening of new virulence factors and antibacterial targets. *Environ. Microbiol.* **5**, 1294–1308.
5. Anantharaman, V., Koonin, E. V. & Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* **30**, 1427–1464.
6. Kaplan, C. D., Morris, J. R., Wu, C. & Winston, F. (2000). Spt5 and spt6 are associated with active transcription and have characteristics of general elongation factors in *D. melanogaster*. *Genes Dev.* **14**, 2623–2634.
7. Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucleic Acids Res.* **30**, 3643–3652.
8. Doolittle, R. F. (1986). *Urfs and Orfs: A Primer on How to Analyze Derived Amino Acid Sequences,* University Science Books, Mill Valley, CA.
9. 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.
10. Bortvin, A. & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, **272**, 1473–1476.
11. Kaplan, C. D., Laprade, L. & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, **301**, 1096–1099.
12. 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.
13. 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.
14. Carpousis, A. J. (2002). The *Escherichia coli* RNA degradosome: structure, function and relationship in other ribonucleolytic multienzyme complexes. *Biochem. Soc. Trans.* **30**, 150–155.
15. 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.
16. 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.
17. Buttner, K., Nehring, S. & Hopfner, K. P. (2007). Structural basis for DNA duplex separation by a superfamily-2 helicase. *Nat. Struct. Mol. Biol.* **14**, 647–652.
18. Holm, L. & Sander, C. (1998). Touring protein fold space with Dali/FSSP. *Nucleic Acids Res.* **26**, 316–319.
19. Landau, M., Mayrose, I., Rosenberg, Y., Glaser, F., Martz, E., Pupko, T. & Ben-Tal, N. (2005). ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302.
20. Letunic, I., Copley, R. R., Pils, B., Pinkert, S., Schultz, J. & Bork, P. (2006). SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **34**, D257–D260.
21. Rocha, E. P., Cornet, E. & Michel, B. (2005). Comparative and evolutionary analysis of the bacterial homologous recombination systems. *PLoS Genet.* **1**, e15.
22. Ichiyanagi, K., Iwasaki, H., Hishida, T. & Shinagawa, H. (1998). Mutational analysis on structure–function relationship of a holliday junction specific endonuclease RuvC. *Genes Cells*, **3**, 575–586.
23. Yoshikawa, M., Iwasaki, H. & Shinagawa, H. (2001). Evidence that phenylalanine 69 in *Escherichia coli* RuvC resolvase forms a stacking interaction during binding and destabilization of a Holliday junction DNA substrate. *J. Biol. Chem.* **276**, 10432–10436.
24. Ariyoshi, M., Vassylyev, D. G., Iwasaki, H., Nakamura, H., Shinagawa, H. & Morikawa, K. (1994). Atomic

structure of the RuvC resolvase: a holliday junction-specific endonuclease from *E. coli*. *Cell*, **78**, 1063–1072.
25. 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.
26. Yoshikawa, M., Iwasaki, H., Kinoshita, K. & Shinagawa, H. (2000). Two basic residues, Lys-107 and Lys-118, of RuvC resolvase are involved in critical contacts with the Holliday junction for its resolution. *Genes Cells*, **5**, 803–813.
27. Shao, X. & Grishin, N. V. (2000). Common fold in helix–hairpin–helix proteins. *Nucleic Acids Res.* **28**, 2643–2650.
28. Bycroft, M., Hubbard, T. J., Proctor, M., Freund, S. M. & Murzin, A. G. (1997). The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235–242.
29. Subramanian, A. R. (1983). Structure and functions of ribosomal protein S1. *Prog. Nucleic Acid Res. Mol. Biol.* **28**, 101–142.
30. 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.
31. Amblar, M., Barbas, A., Gomez-Puertas, P. & Arraiano, C. M. (2007). The role of the S1 domain in exoribonucleolytic activity: substrate specificity and multimerization. *RNA*, **13**, 317–327.
32. Schubert, M., Edge, R. E., Lario, P., Cook, M. A., Strynadka, N. C., Mackie, G. A. & McIntosh, L. P. (2004). Structural characterization of the RNase E S1 domain and identification of its oligonucleotide-binding and dimerization interfaces. *J. Mol. Biol.* **341**, 37–54.
33. Hill, J. J. & Royer, C. A. (1997). Fluorescence approaches to study of protein–nucleic acid complexation. *Methods Enzymol.* **278**, 390–416.
34. Stickney, L. M., Hankins, J. S., Miao, X. & Mackie, G. A. (2005). Function of the conserved S1 and KH domains in polynucleotide phosphorylase. *J. Bacteriol.* **187**, 7214–7221.
35. Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G. & Luisi, B. F. (2005). Structure of *Escherichia coli* RNase E catalytic domain and implications for RNA turnover. *Nature*, **437**, 1187–1191.
36. Yatime, L., Schmitt, E., Blanquet, S. & Mechulam, Y. (2005). Structure–function relationships of the intact aIF2alpha subunit from the archaeon *Pyrococcus abyssi*. *Biochemistry*, **44**, 8749–8756.
37. Todone, F., Brick, P., Werner, F., Weinzierl, R. O. & Onesti, S. (2001). Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell*, **8**, 1137–1143.
38. Meka, H., Werner, F., Cordell, S. C., Onesti, S. & Brick, P. (2005). Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II. *Nucleic Acids Res.* **33**, 6435–6444.
39. Maclennan, A. J. & Shaw, G. (1993). A yeast SH2 domain. *Trends Biochem. Sci.* **18**, 464–465.
40. Grune, T., Brzeski, J., Eberharter, A., Clapier, C. R., Corona, D. F., Becker, P. B. & Muller, C. W. (2003). Crystal structure and functional analysis of a nucleosome recognition module of the remodeling factor ISWI. *Mol. Cell*, **12**, 449–460.
41. Studier, F. W. (2005). Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234.
42. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
43. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. In *Methods in Enzymology* (Carter, J. C. W. & Sweet, R. M., eds), Macromolecular Crystallography, Part A, vol. 276, pp. 307–326. Academic Press, New York, NY.
44. Terwilliger, T. C. & Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **55**, 849–861.
45. Terwilliger, T. C. (2000). Maximum-likelihood density modification. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **56**, 965–972.
46. McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). Likelihood-enhanced fast translation functions. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **61**, 458–464.
47. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54**, 905–921.
48. Adams, P. D., Grosse-Kunstleve, R. W., Hung, L. W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W. *et al.* (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **58**, 1948–1954.
49. Collaborative Computational Project (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **50**, 760–763.
50. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.* **A47**, 110–119.
51. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 2126–2132.
52. DeLano, W. L. (2002). *The PyMOL Molecular Graphics System*, DeLano Scientific, Palo Alto, CA.
53. Hoffmann, K. M., Williams, D., Shafer, W. M. & Brennan, R. G. (2005). Characterization of the multiple transferable resistance repressor, MtrR, from *Neisseria gonorrhoeae*. *J. Bacteriol.* **187**, 5008–5012.

Figure S1. Determination of Tex:25nt ssRNA binding stoichiometry using fluorescence polarization. Fluorescein labeled RNA was held at a constant concentration 20-fold above the K_d. Tex protein was titrated in at increasing concentration. An inflection point representing the molar ratio where the binding sites have become saturated occurs at equimolar concentrations of Tex and 25nt ssRNA indicating a 1:1 binding stoichiometry. The assay was performed in 15mM Tris 7.5, 100mM NaCl, 5% glycerol and measured using a Tecan fluorimeter.

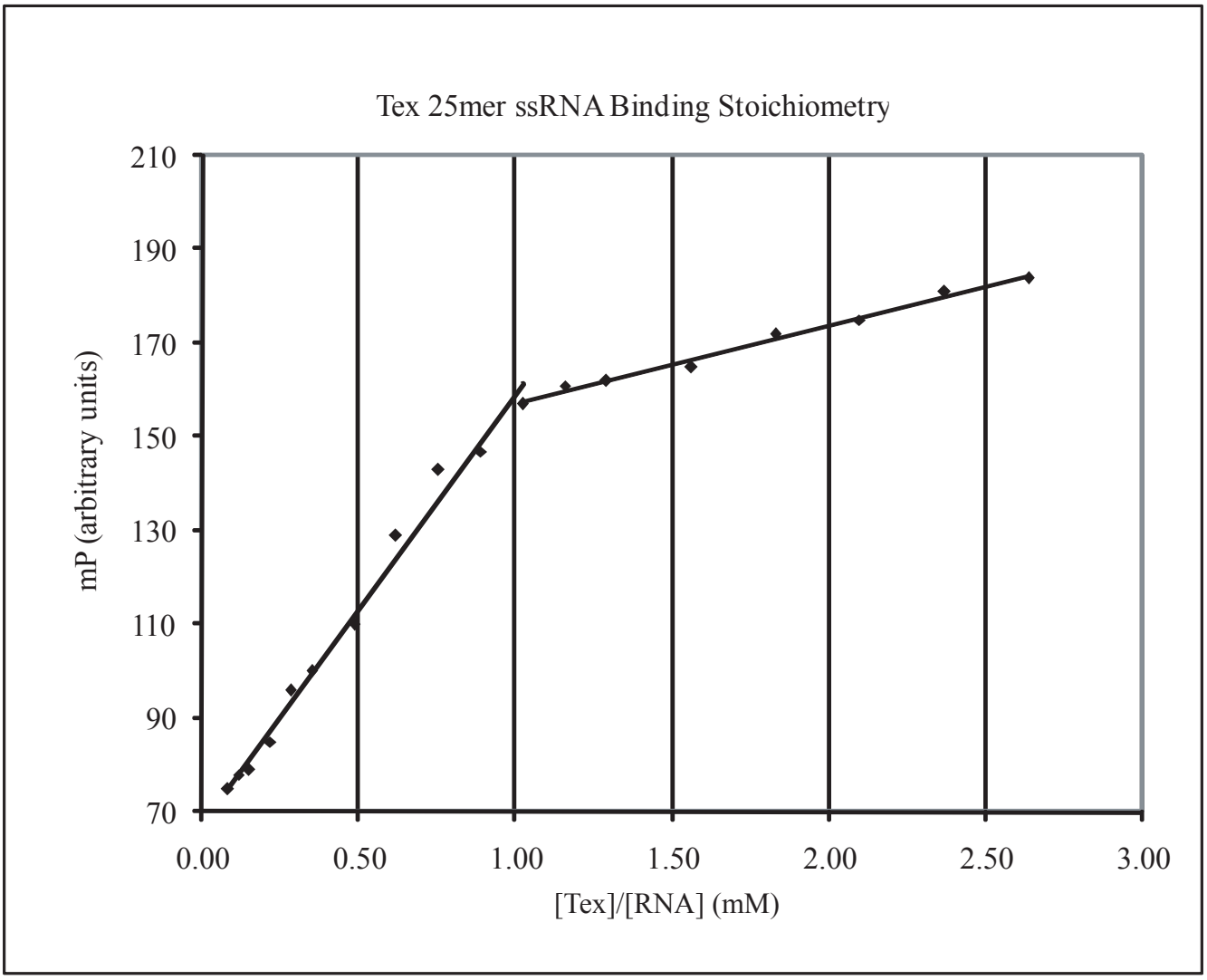


Figure S2. Tex does not appear to have nuclease activity. WT Tex was assayed in a time course experiment in parallel with RNase T1, Micrococcal Nuclease, and the Tex YqgF domain active site mutant E335A/D421A. In the assay, the Tex proteins were at a molar concentration 100x that of Micrococcal Nuclease. Substrate used in this assay is a 5' radio-labeled 300 nt single-stranded RNA and the assay was performed in 15mM Tris 7.5, 100mM NaCl, 5% glycerol, 3mM $MgCl_2$, 1mM $CaCl_2$ at 37 C. Reactions were quenched by the addition of an equal volume of phenol/chloroform pH 6.6 and 25mM EDTA. Samples were spun down and the aqueous phase extracted and mixed with 2x formamide loading dye. Samples were run on a 12% acrylamide (19:1)/7M Urea denaturing gel in TBE buffer and the gel was scanned using a TYPHOON imaging system (GE Healthsciences).

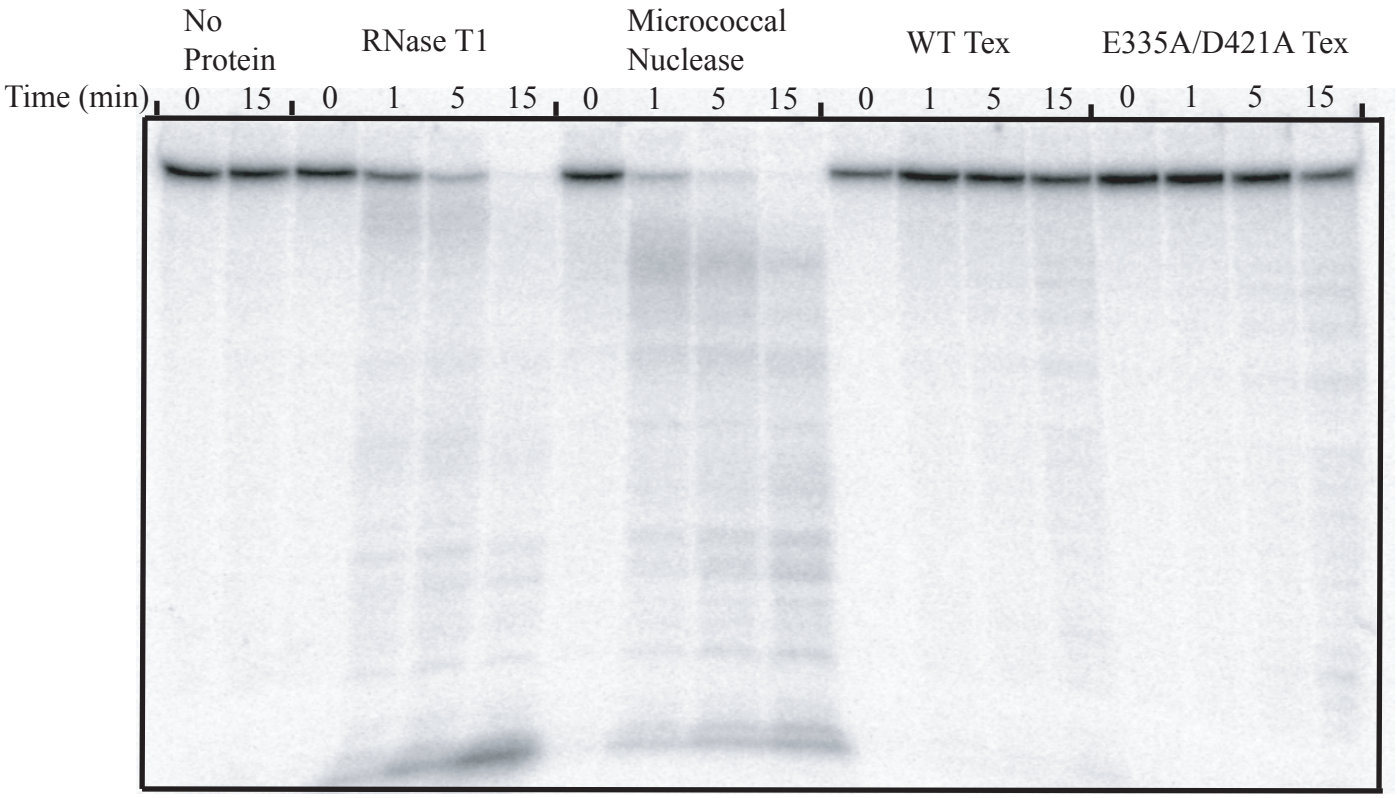


Figure S3. Predicted sites of sequence insertion for Spt6, with respect to the Tex structure. Front (A) and back (B) stereo views of the Tex structure. Highlighted regions (purple) indicate sites where additional sequence exists in Spt6 when aligned with Tex sequences. All highlighted regions lie on the surface of the Tex structure and appear to be able to accommodate additional sequence without disrupting the core structural scaffold.

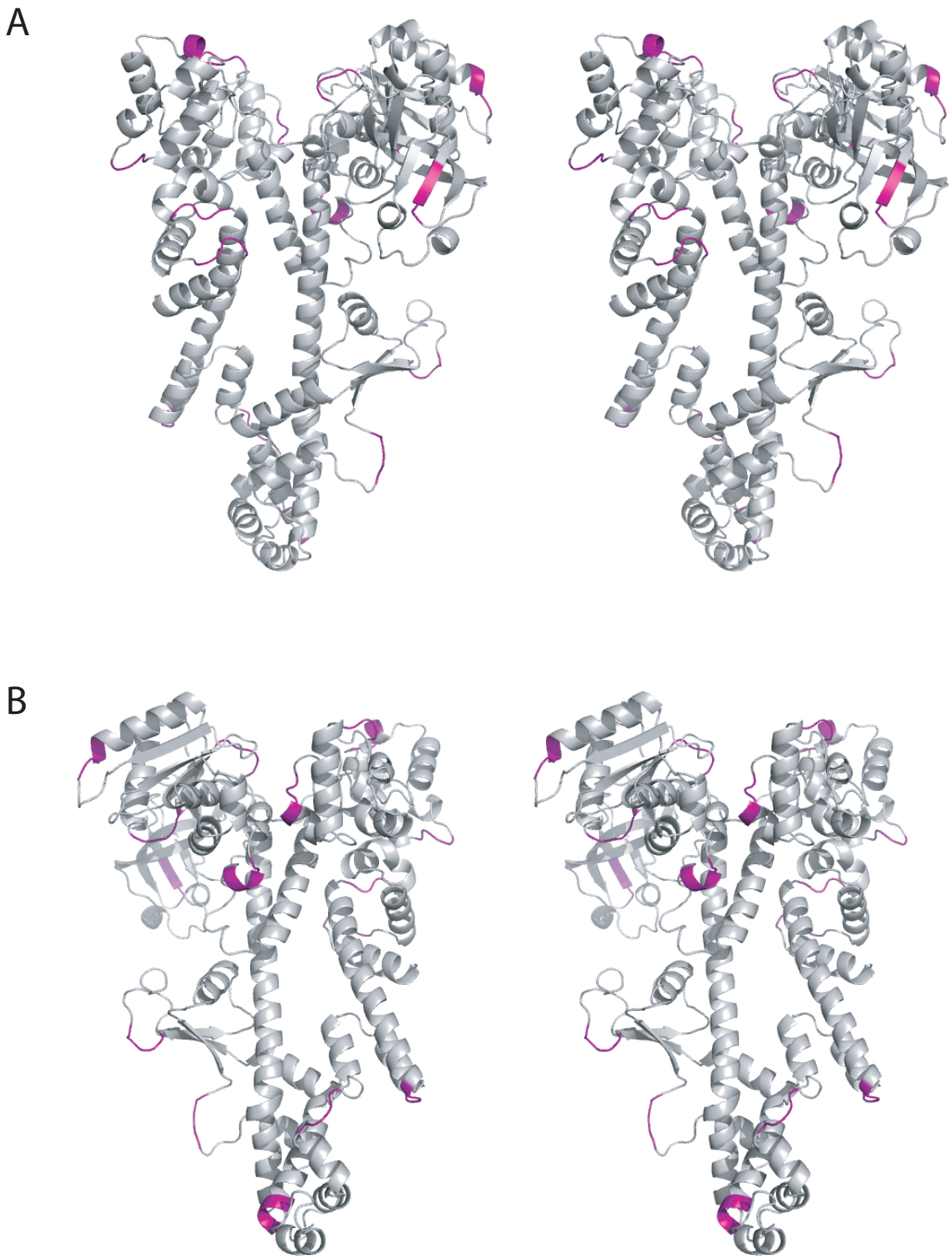
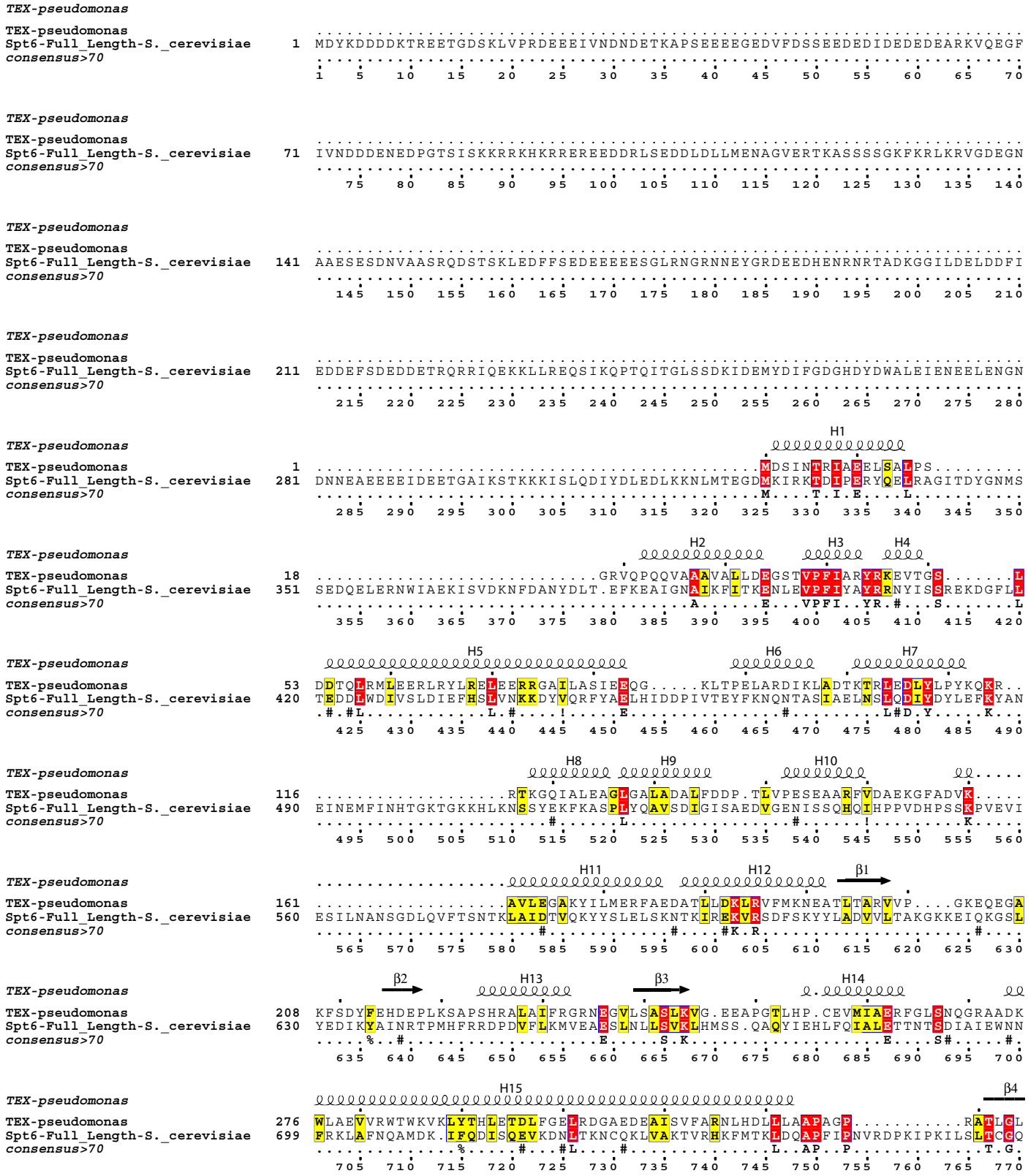
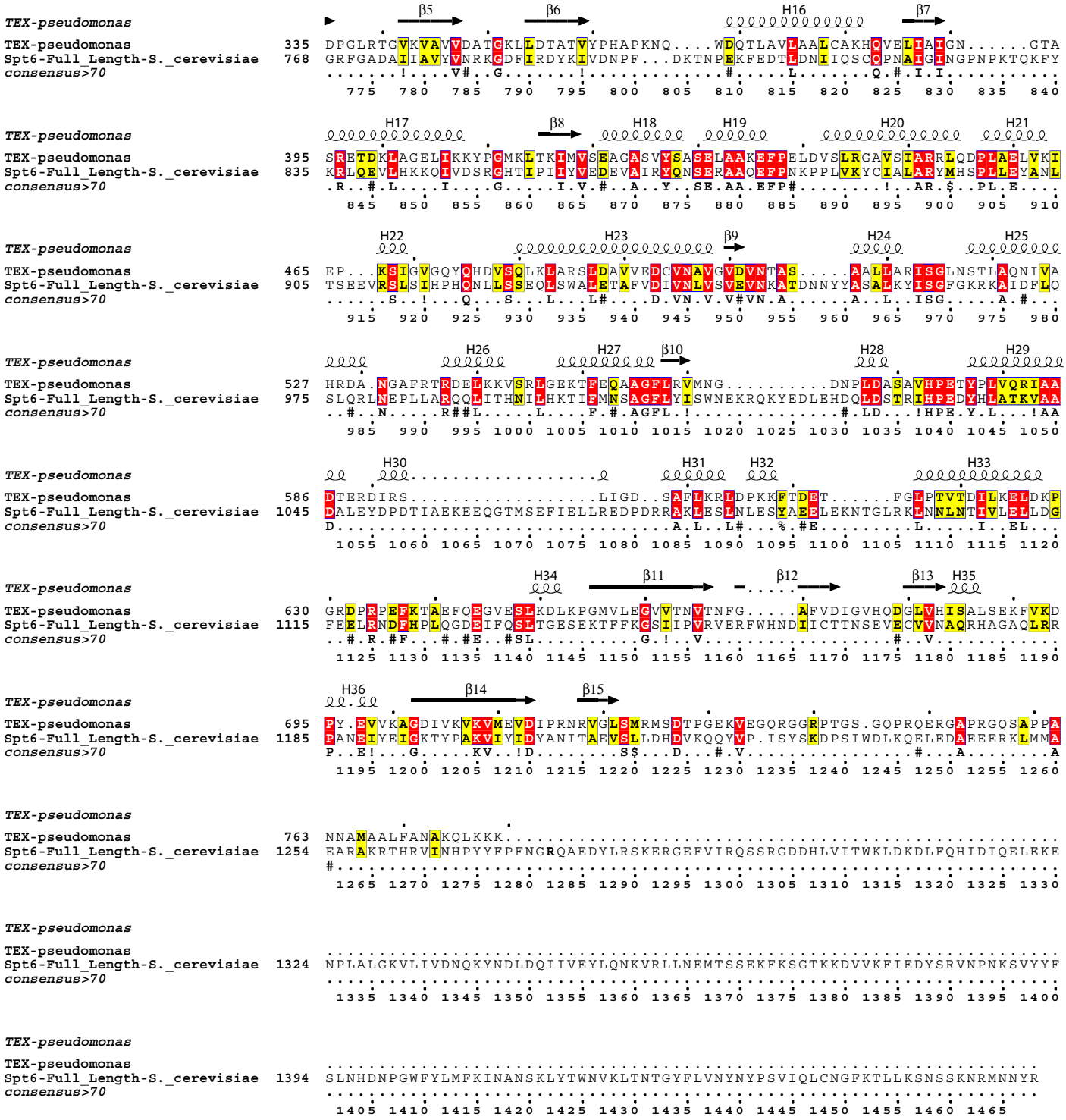


Figure S4. Amino acid sequence alignment of Tex (*P. aeruginosa*) and Spt6 (*S. cerevisiae*). Observed Tex secondary structure is indicated above the Tex sequence. Highlighted regions indicate identical (red) and similar (yellow) residues. Sequence alignment was performed using ClustalW.[1] Figure was created using ESPript.[2]





References

1. Higgins D., Thompson J., Gibson T., Thompson J.D., Higgins D.G., Gibson T.J. (1994). CLUSTAL W: improving the sensitivity of progressivemultiple sequence alignment through sequence weighting,position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.

2. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-308.