JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Structure and *in Vivo* Requirement of the Yeast Spt6 SH2 Domain

Stefan Dengl, Andreas Mayer, Mai Sun and Patrick Cramer*

Gene Center and Center for Integrated Protein Science Munich (CIPSM), Department of Chemistry and Biochemistry, Ludwig-Maximilians-Universität München, Feodor-Lynen-Str. 25, 81377 Munich, Germany

During transcription elongation through chromatin, the Ser2-phosphory-lated C-terminal repeat domain of RNA polymerase II binds the C-terminal Src homology 2 (SH2) domain of the nucleosome re-assembly factor Spt6. This SH2 domain is unusual in its specificity to bind phosphoserine, rather than phosphotyrosine and because it is the only SH2 domain in the yeast genome. Here, we report the high-resolution crystal structure of the SH2 domain from *Candida glabrata* Spt6. The structure combines features from both structural subfamilies of SH2 domains, suggesting it resembles a common ancestor of all SH2 domains. Two conserved surface pockets deviate from those of canonical SH2 domains, and may explain the unusual phosphoserine specificity. Differential gene expression analysis reveals that the SH2 domain is required for normal expression of a subset of yeast genes, and is consistent with multiple functions of Spt6 in chromatin transcription.

© 2009 Elsevier Ltd. All rights reserved.

Edited by R. Huber

Introduction

The gene encoding Spt6 was originally identified in a genetic screen in yeast as a suppressor of transposon insertion in the promoter region of a reporter gene.[1] Spt6 was later described as an essential nuclear protein.[2,3] Like Spt4 and Spt5, Spt6 is required for transcription elongation,[4,5,31,70] and co-localizes with Spt5 and RNA polymerase (Pol) II at transcriptionally active loci on *Drosophila* polytene chromosomes.[6,7] Spt6 interacts with the structured part of histone H3 and can deposit nucleosomes on DNA in a supercoiling assay.[8] Loss of normal Spt6 function leads to transcription from cryptic promoters within the coding region of genes,[9] suggesting that Spt6 is required to re-establish correct chromatin structure after Pol II passage. Spt6 also has a role in the re-assembly of nucleosomes in promoter regions, and is required for repression of transcription reinitiation.[10] Mutations in Spt6 influence mRNA 3′-end formation,[11] splicing, and mRNA export.[12]

*Corresponding author. E-mail address: cramer@LMB.uni-muenchen.de.

Abbreviations used: SH2, Src homology 2; Pol, RNA polymerase; CTD, C-terminal repeat domain of RNA polymerase II subunit Rpb1.

Spt6 is associated with the Pol II machinery,[6,7] and its SH2 domain is apparently important for this. Spt6 binds to Spt5,[4,5] and occupies the 5′-region of the uninduced heat shock gene *hsp70*.[6] *In vivo*, Spt6 and Spt5 co-localize with actively transcribing Pol II that is phosphorylated at Ser2 of the C-terminal repeat domain (CTD) of the Pol II largest subunit.[7] *In vitro*, the C-terminal region of mouse Spt6, which includes the SH2 domain, binds the Ser2-phosphorylated CTD, and this interaction is required for normal mRNA processing and exit from the nucleus.[12]

The Spt6 SH2 domain is of special interest because it apparently has phosphoserine-binding specificity, whereas SH2 domains generally bind to phospho-tyrosine-containing peptides. Further, it is the only SH2 domain encoded in the yeast genome[13] and thus may represent an ancestor of the large variety of SH2 domains encoded in the genome of higher eukaryotes. SH2 domains are highly abundant in animal cells as adaptor modules that connect proteins in tyrosine kinase pathways, and are important mediators in cellular signaling.[14,15]

Structural information is available for SH2 domains and their complexes with phosphopeptides, including that of the Src SH2 domain–phosphopeptide complex.[16] SH2 domains can be divided into two structural subfamilies: the Src-type and the STAT-type subfamilies.[17] Whereas both families of domains share a central fold, the Src-type domains

contain an extra β-sheet (βE-βF motif) that participates in the read-out of residues C-terminal to the phosphotyrosine,[16] and the STAT-type domains show an extended helical structure at their C-terminus (αB′ and αB). Phosphorylated peptides bind two pockets on the SH2 domain surface, one that recognizes the phosphorylated tyrosine side chain, and one that contacts residues C-terminal of the phosphotyrosine, generally at register +3. Binding of the peptide to the second pocket is a major determinant of the interaction specificity.[18] Although in some cases binding of phosphoserine- and phosphothreonine-containing peptides to SH2 domains has been reported,[19,20] there is no structural information on the interaction of a SH2 domain with peptides other than those containing phosphotyrosine.

Here, we report the high-resolution structure of the SH2 domain of Spt6, suggest a model for how this domain binds to the phosphorylated Pol II CTD, and show that the domain is required for normal expression of a subset of yeast genes. Comparison of known SH2 domain structures reveals that the Spt6

SH2 domain structure is intermediate between SH2 domains of the two subfamilies in higher eukaryotes, reflecting the likely structure of the common ancestor of all SH2 domains.

Results

Structure determination of the Spt6 SH2 domain

To delineate the borders of the SH2 domain in Spt6, we expressed and purified a C-terminal fragment of Spt6 from the yeast *Saccharomyces cerevisiae* (residues 1251–1378) and subjected the pure protein to limited proteolysis (Materials and Methods). This analysis, together with predictions based on known SH2 domain structures,[21] suggested that the domain comprises residues 1251–1349. Crystallization trials with the corresponding recombinant domain from *S. cerevisiae*, *Schizosaccharomyces pombe*, *Candida glabrata*, and *Homo sapiens* resulted in crystals for the *C. glabrata* SH2 domain,
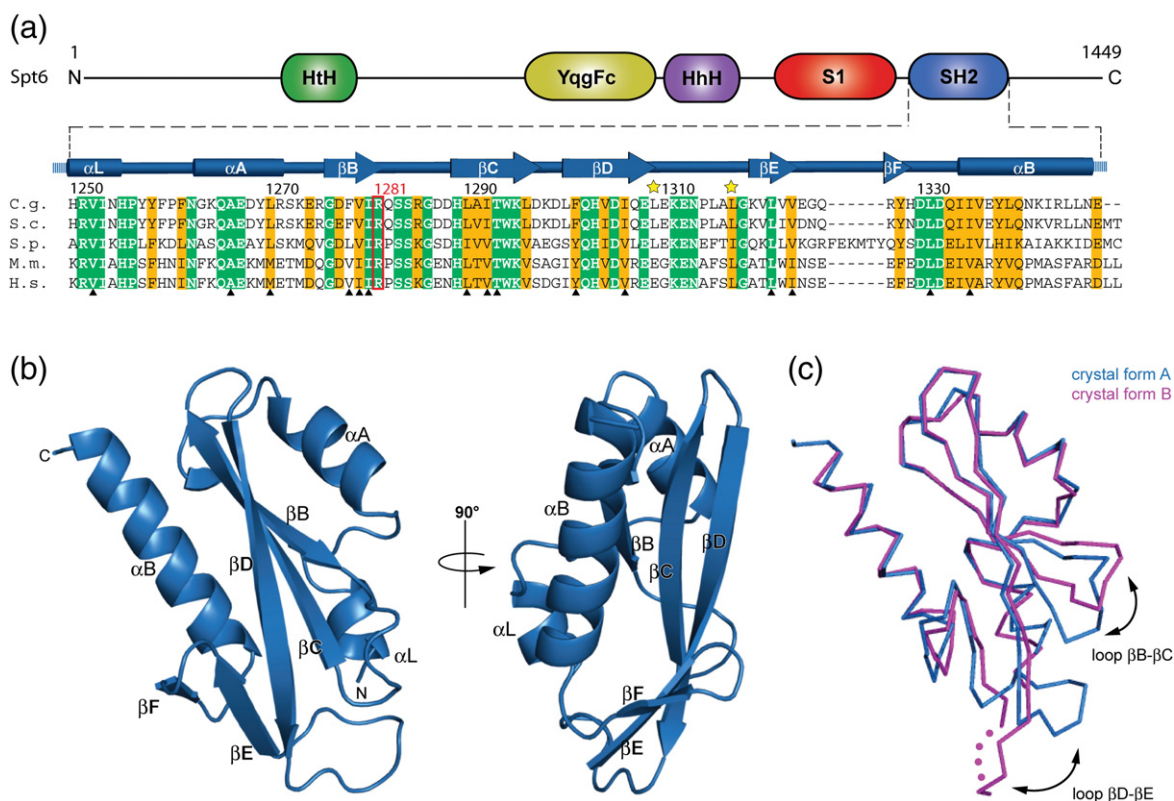
Fig. 1. Structure of the Spt6 SH2 domain. (a) The domain architecture of Spt6[29] and alignment of amino acid sequences of the Spt6 SH2 domains of *C. glabrata* (C.g.), *S. cerevisiae* (S.c.), *S. pombe* (S.p.), *M. musculus* (M.m.), and *H. sapiens* (H.s.). Secondary structure elements are indicated above the alignment (cylinders for α helices, arrows for β strands). Invariant and conserved residues are highlighted in green and orange, respectively. The red box indicates the invariant arginine R1281 that binds the phospho group. Stars indicate the positions where methionine was introduced for selenomethionine phasing (Materials and Methods). Residues forming the hydrophobic core are marked with a triangle. HtH, helix-turn-helix domain, binds to double-stranded DNA; YqgFc, predicted to be resolvase or ribonuclease, but in Spt6 catalytic residues are exchanged, thus probably not active; HhH, triple-helix-domain, binding of double-stranded DNA; S1, RNA-binding domain. (b) Two views of a ribbon model of the Spt6 SH2 domain. Secondary structure elements are labeled according to a. (c) Comparison of the Cα trace of the SH2 domain structure in two different crystal forms (blue, crystal form A; magenta, crystal form B). Structures were aligned by DALILite,[64] resulting in an r.m.s.d. of 1.2 Å for 91 residues. Unstructured parts in crystal form B are indicated by dots.

which is highly homologous to that of other eukaryotic species (Fig. 1). The X-ray structure was solved at 1.9 Å resolution by introduction of selenomethionine residues and multiwavelength anomalous diffraction (crystal A, Table 1, Methods). A second, independent structure was obtained in a different crystal form at 2.4 Å resolution by molecular replacement (crystal B, Table 1, Materials and Methods).

Unique fold of the Spt6 SH2 domain

The structure reveals the classical core fold of SH2 domains,[18] with a central three-stranded antiparallel β-sheet (βB–βD) sandwiched between two α-helices (αA and αB, Fig. 1b). In addition to this core fold, the structure contains an α-helix N-terminal of αA (called αL here), a small anti-parallel β-sheet inserted between βB and βD (βE-βF), loops βB-βC and βD-βE that both adopt alternative conformations in the two different crystal forms (Fig. 1c), and an extended C-terminal α-helix (Figs. 1b and 2). In addition, Spt6 lacks a phenylalanine in the βB core motif that contains the phosphate-binding arginine (green in Fig. 2). This phenylalanine is highly conserved only in STAT-type domains.[17]

To investigate how these structural features are related to known SH2 domain structures, we compared the new structure to known SH2 domain structures, including 32 of the Src-type family and four of the STAT-type (Supplementary Data Table S1). Whereas the βE-βF motif is found only in the Src-type subfamily[17], helix αL is a feature that

occurs adjacent to several SH2 domains of both subfamilies,[22] and might be related to a preceding linker domain, as first observed in STAT proteins (Fig. 2).[23,24] The extended helix B is almost twice as long as the respective helices in known SH2 structures, except for the APS SH2 domain,[25] a feature that might resemble the extended helix αB' seen in STAT domains. Taken together, the yeast Spt6 SH2 domain combines structural features from both SH2 subfamilies of higher eukaryotes.

Conserved phospho-binding site

The domain surface shows a conserved patch that includes a small pocket (pocket 1, Fig. 3) that contains an invariant arginine residue (R1281). This arginine is present in all known SH2 domains and interacts directly with the phosphate group of the target phosphopeptide in the SH2 domain-phosphopeptide complex (Fig. 3).[26] In our crystals, R1281 binds a succinate ion that was present in the crystallization buffer. R1281 side chains from two neighboring domains in the crystal each bind a carboxylate of the succinate ion (Fig. 3b). This observation is consistent with a high affinity of pocket 1 for negatively charged chemical groups, and the conserved phospho-binding function of R1281. Indeed, mutation of R1281 decreased the interaction of the SH2 domain with the phosphory-lated CTD *in vitro*.[12]

We modeled the possible CTD interaction with the use of the known Src SH2 domain–phosphopeptide complex structure (Fig. 3d).[16] We chose the Src-SH2 domain, because it shows the highest degree of similarity in a structural comparison using the DALI-server,[65] and because it is well described in the literature. The two structures were superim-posed with their conserved residues in the central β-sheet of the core domain fold. The phospho-binding pockets are highly similar in structure (Fig. 3d). The positions of Cα atoms of the residues that form pocket 1 are essentially identical, except for the residues of the phosphate-binding loop βB-βC that deviates slightly from those of Src (broken blue line in Fig. 3d). However, the Spt6 residue R1281 aligns perfectly with the phosphotyrosine-binding residue R175 of Src and is thus in a position to make contacts with the phosphate by specific hydrogen bonding interactions between the two terminal nitrogen and two phosphate oxygen atoms as partially mimicked by the succinate (Figs. 3b and e). Additional contacts of the phosphate group are likely conserved, including hydrogen bonds to a backbone amide (E178 in Src, S1284 in Spt6) and to the side-chain hydroxyl group of at least one of two residues (S177 and T179 of Src, apparently corresponding to S1283 and S1284, respectively, of Spt6).
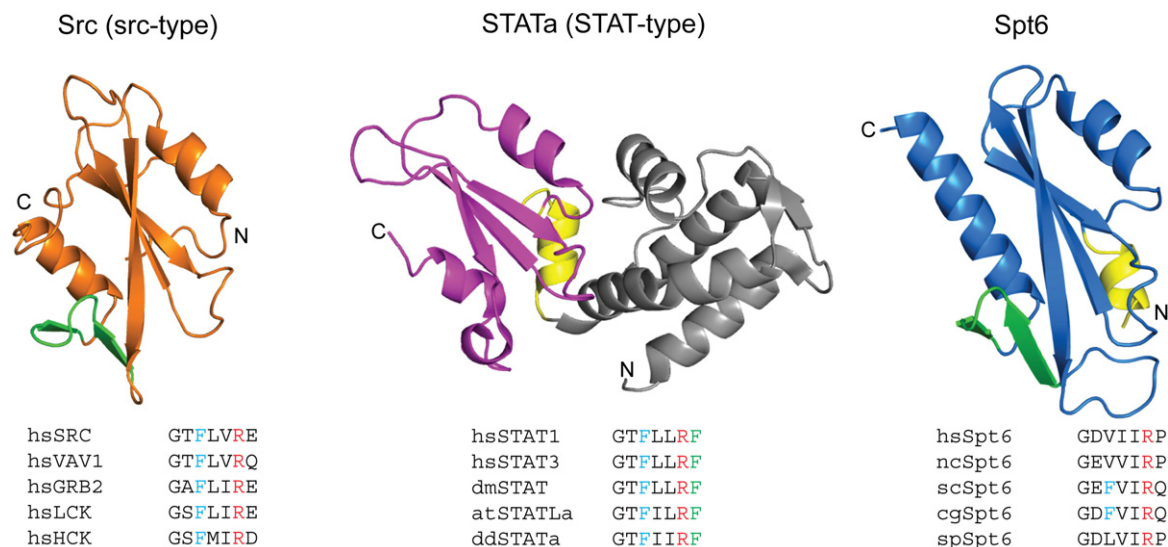
Unusual phosphoserine specificity

Whereas the contacts to the phosphate group are likely conserved, modeling suggests that contacts to an aromatic ring of a phosphotyrosine side chain are

Table 1. X-ray diffraction data and refinement statistics

	Crystal form A			Crystal form B
A. *Data collection*				
Space group	P65			P32
Cell dimensions				
a, b (Å)	54.5, 54.5			71.6, 71.6
c (Å)	253.4			87.6
	Peak	Remote	Inflection	
Wavelength (nm)	0.97973	0.90810	0.97987	0.97971
Resolution (Å)	20–1.9	20–1.9	20–1.9	20–2.4
R_{sym} (%)	5.2 (12.9)	5.3 (22.3)	3.9 (12.6)	6.2 (19.0)
$I/\sigma s\ I$	42.9 (7.3)	34.4 (7.5)	31.2 (6.0)	45.1 (7.7)
Completeness (%)	99.4 (96.6)	99.9 (100)	99.3 (95.7)	99.5 (95.4)
Redundancy	4.2 (2.8)	7.7 (6.7)	4.1 (2.8)	3.9 (3.7)
B. *Refinement*				
Resolution (Å)	1.9			2.4
No. reflections	33,162			18,513
R_{work} (%)	19.6			25.3
R_{free} (%)	24.1			28.5
No. atoms				
Protein	3292			3160
Ligand/ion	20			—
Water	428			76
B-factors				
Protein (Å²)	26.6			29.8
Ligand/ion (Å²)	23.5			-
Water (Å²)	35.4			27.8
r.m.s.d. from ideal				
Bond lengths (Å)	0.005			0.007
Bond angles (°)	1.2			1.4

Src (src-type)   STATa (STAT-type)   Spt6

hsSRC	GTFLVRE
hsVAV1	GTFLVRQ
hsGRB2	GAFLIRE
hsLCK	GSFLIRE
hsHCK	GSFMIRD

hsSTAT1	GTFLLRF
hsSTAT3	GTFLLRF
dmSTAT	GTFLLRF
atSTATLa	GTFILRF
ddSTATa	GTFIIRF

hsSpt6	GDVIIRP
ncSpt6	GEVVIRP
scSpt6	GEFVIRQ
cgSpt6	GDFVIRQ
spSpt6	GDLVIRP

Fig. 2. Unique fold of the Spt6 SH2 domain. Comparison of the overall structures of the SH2 domains of *H. sapiens* Src (orange[16]), of *D. discoideum* STATa (magenta[43]), and of *C. glabrata* Spt6 (blue, this work). The βE-βF-motif is colored green, the Spt6 helix αL, as well as STATa linker domain helix α11 are colored yellow. The remainder of the STAT linker domain is colored gray. Alignments of the SH2 βB core motif of five representative sequences are shown below each structure (hs, *H. sapiens*; dm, *D. melanogaster*; at, *A. thaliana*; dd, *D. discoideum*; nc, *N. crassa*; sc, *S. cerevisiae*; cg, *C. glabrata*; sp, *S. pombe*). The highly conserved arginine residue that binds the phospho group is highlighted in red, a conserved phenylalanine residue that is a feature of STAT-type domains is highlighted in green. A highly conserved phenylalanine that is variable in Spt6 is in cyan.

apparently not possible in the Spt6 domain, consistent with binding to a phosphoserine peptide. The specific recognition of the phosphotyrosine aromatic ring by Src is realized by amino–aromatic interactions, where the π-electrons of the aromatic ring interact with the amino groups of residues R155 and K203 in Src (Fig. 3e). The aromatic ring of the phosphotyrosine is sandwiched between these two residues (Fig. 3d). While the two residues corresponding to Src residues R155 and K203 are conserved in the majority of SH2 domains, some SH2 domains lack one of them (Supplementary Data Table S1). However, the SH2 domains of Spt6 and Cbl are the only SH2 domains within our structural comparisons, in which both residues are not conserved. Src residues R155 and K203 are replaced by residues G1263 and D1305 in Spt6. The nega-

tively charged residue D1305 was observed only in Spt6 and could repel aromatic π-electrons, consistent with the selection of serine over tyrosine, but may also form a salt bridge with residue R1285 of loop βB-βC (Fig. 3e). The alignment shows also that pocket 1 is shallower in Spt6 than in Src (Fig. 3f). Thus, the Spt6 pocket 1 is suited for phosphate binding, and R1281 is more readily accessible from the solvent than in the Src SH2 domain, enabling its interaction with a short phosphorylated serine side chain.

Model for CTD binding

In the structure of the human Src protein with its high-affinity target peptide,[16] the peptide runs across the binding surface from pocket 1 to pocket

Fig. 3. Surface properties of the Spt6 SH2 domain. (a) Conservation of surface residues, which are colored according to the alignment in Fig. 1a, with green for invariant and orange for conserved. The putative peptide-binding pockets 1 and 2, and the surface residues implicated in peptide binding are indicated. (b) $2F_o - F_c$ electron density map, contoured at 1 σ, for the arginine residue R1281 of neighboring domains in the crystal and a bridging succinate ion. (c) Surface charge distribution of the Spt6 SH2 domain (left) and the Src SH2 domain (right;[16] PDB code *1SPS*). The phosphopeptide bound to the SH2 domain is shown as a stick model in red, only four residues of the original structure are shown. (d) Superposition of the Src SH2 domain-phosphopeptide complex structure (orange/red[16]) and the Spt6 SH2 domain (blue). Structures were aligned using the most conserved residues (G1276–R1281, L1290–K1295, Q1302–E1308) of the central β-sheet strands. The β-sheets aligned very well, resulting in an RMSD of 0.4 Å for 19 Cα atoms. A close-up view of the phospho-binding site is shown at the right. Src residues that interact with the phosphotyrosine are in orange, and corresponding residues in Spt6 are in blue. The broken blue line indicates the position of loop βB-βC of Spt6. (e) Conservation of the phospho-binding pocket. The modeling in d was used to delineate the Spt6 residues predicted to be involved in phospho group binding based on the Src SH2 domain–peptide complex structure. Interactions of Src residues with the phosphotyrosine[26] are in orange, and potential corresponding interactions of the Spt6 residues with the phospho group are in blue. Phospho-mimetic interactions of residues with succinate are in green. Stars indicate the residues depicted in d. (f) Different shape of the phospho-binding pocket. The molecular surfaces of the phospho-binding pockets in Spt6 and Src are in blue and orange, respectively, and the phosphotyrosine residue of the peptide bound to Src is in red.

2 (Fig. 3c, right). The polarity is dictated by the interaction of the phosphotyrosine with pocket 1 and an isoleucine C-terminal of the phosphotyrosine with pocket 2. This binding mode is used by most SH2-domains ("two-pronged plug") and likely also by the SH2-domain of Spt6, since a stretch of conserved surface residues extends from pocket 1 to a second pocket (pocket 2, Fig. 3a).

In Src, pocket 2 is important mainly for substrate specificity,[16,27] and accommodates the residue that is three residues C-terminal of the phosphotyrosine (Fig. 3c). Due to the shorter side chain of the phosphoserine, this register would shift to the peptide N-terminus, as suggested by modeling (not shown). This would bring residues at position +3 to +5 in the proximity of pocket 2 (Fig. 4, left).
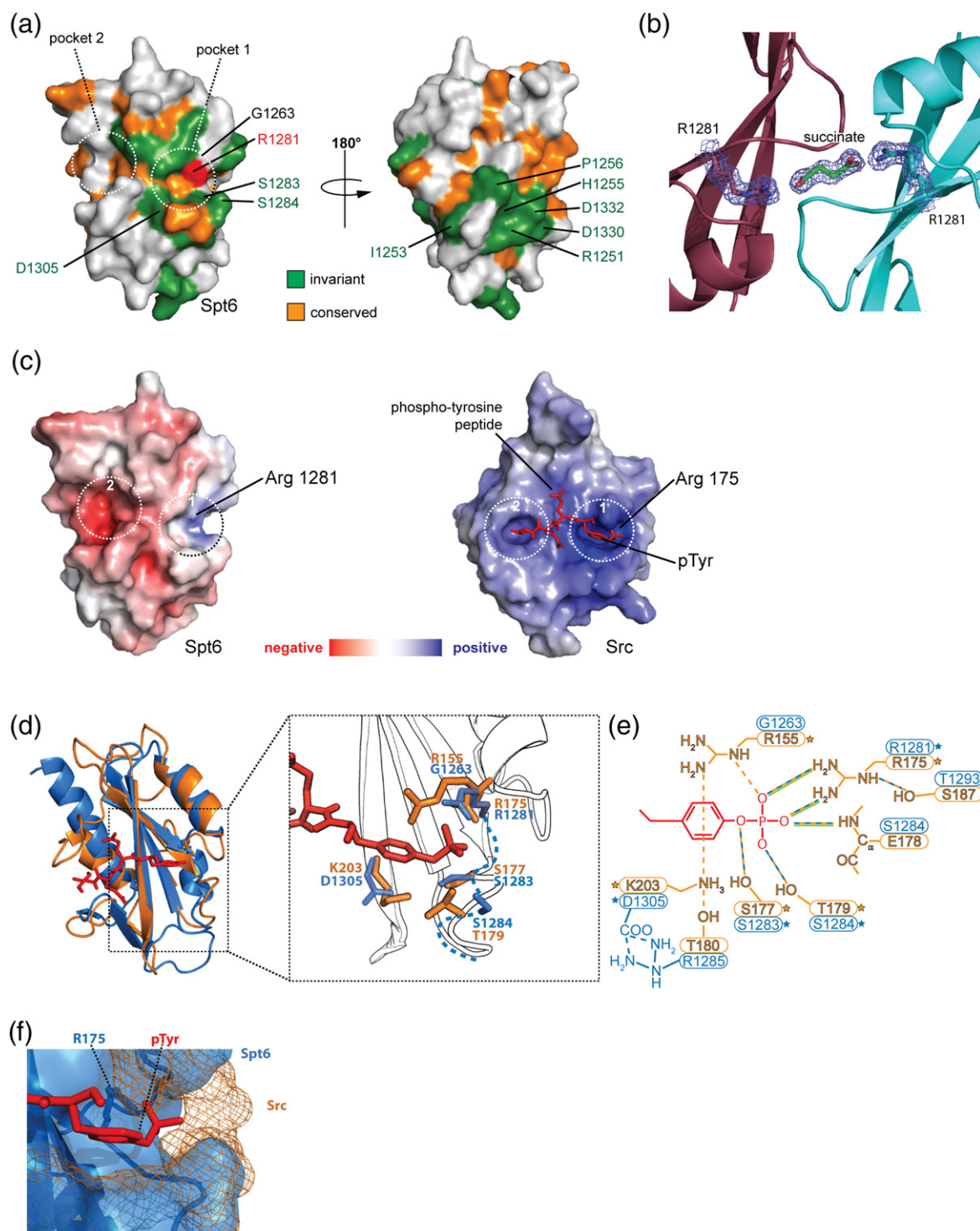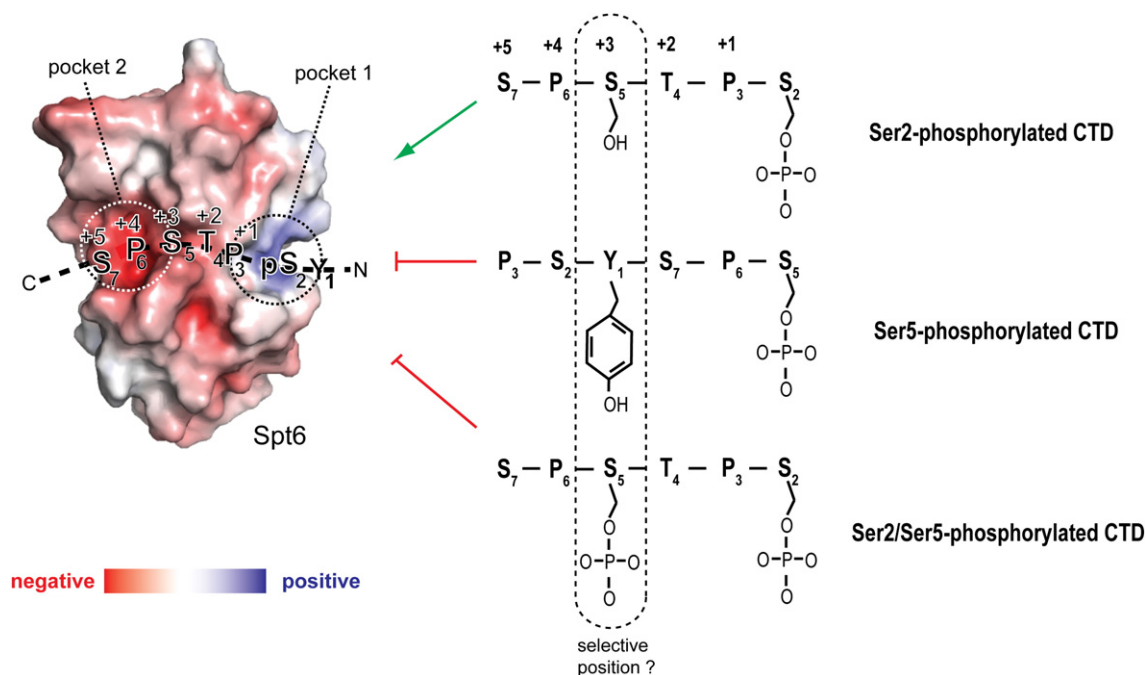

Fig. 3 *(legend on previous page)*

Fig. 4. Model for binding of the Spt6 SH2 domain to the Pol II CTD. Suggested path for CTD phosphopeptide binding to the SH2 domain (broken black line) and implications for specific binding to the Ser2-phosphorylated CTD. A peptide position that might be important for the selectivity of the interaction is marked by a broken line. The Spt6 SH2 domain is shown with its surface charge distribution as in Fig. 3c. See the text for details.

Pocket 2 is negatively charged in Spt6, whereas it is slightly positively charged in Src and is hydrophobic or positively charged in most SH2 domains. Amongst ten Src-type and four STAT-type SH2 domains that we picked randomly (Supplementary Data Table S1), pocket 2 was negatively charged in only two domains, but generally hydrophobic or positively charged (Fig. S2). The unusual negatively charged pocket 2 may be important for specific phosphopeptide binding.

Assuming that a Ser2-phosphorylated CTD peptide runs along the conserved surface patch towards pocket 2 (Fig. 3a), the CTD residues Ser5, Pro6, and Ser7 at register +3, +4, and +5, respectively, from the phosphoserine would be in proximity of pocket 2 and thus could be crucial for the discrimination between different CTD phosphorylation states. If Ser5 is phosphorylated, the binding determinants change. The phosphorylated Ser5 could also bind to pocket 1, which would position Tyr1 of the CTD in the vicinity of pocket 2 (Fig. 4). Selection could be accomplished in this position by rejection of the bulky tyrosine side chain in comparison to a serine, which resides in this position in the Ser2-phosphorylated peptide. A tyrosine in position +3 would probably result in a lower hydrogen-bonding potential between surface residues and the peptide. In the case of a CTD that is doubly phosphorylated at Ser2 and Ser5, binding of the phosphorylated Ser2 to pocket 1 would position the phosphorylated Ser5 near pocket 2. The negatively charged pocket 2 most likely cannot bind the phosphorylated Ser5 residue (Fig. 4). Taken together, our modeling efforts offer an explanation of why the Spt6 SH2 domain binds to

the Ser2-phosphorylated CTD and not to CTD phosphorylated at Ser5,[12] but the resulting rationale awaits experimental confirmation.

The SH2 domain is insufficient for binding short CTD phosphopeptides

To test our model for CTD binding by the Spt6 SH2 domain, we tried to co-crystallize the complex, unfortunately without success. We also tried to soak phospho-CTD peptides into SH2 domain crystals, in particular with the use of crystal form B, in which the putative peptide-binding pocket was not involved in crystal contacts, also without success, likely because of the high ionic strength of the crystallization conditions (4.3 M NaCl, Materials and Methods). Thus, despite extensive trials we were unable to obtain structural information on the SH2 domain-CTD peptide complex.

We next tested whether the Spt6 SH2 domain binds to short CTD-derived peptides that were phosphorylated at Ser2 residues. We tested the SH2 domains of Spt6 from *S. cerevisiae* (residues 1251–1351) and human (residues 1327–1427), and used two different peptides, one comprising a single CTD heptad repeat and a single phosphorylated Ser2 (pS), and one comprising two repeats and two phospho-Ser2 residues (SPSYpSPTS and YpSPTSP-SYpSPTSPS, respectively). We assayed binding by fluorescence anisotropy, by surface plasmon resonance (BIAcore), and by isothermal titration calorimetry, but could not detect significant binding. The published Spt6-CTD interaction[12] used a C-terminal region of mouse Spt6 that included the SH2 domain,

and the adjacent C-terminal region (residues 1295–1496, whereas the SH2 domain contains residues 1327–1427). We conclude that the SH2 domain alone and/or short CTD fragments are apparently insufficient for an interaction between Spt6 and Pol II. Since mutation of the invariant arginine residue abolished CTD binding, the SH2 domain is involved in the interaction. However, stable CTD binding apparently requires a flanking region C-terminal to the domain and/or multiple phospho-CTD repeats, which are not readily available in large quantities.

While this manuscript was under consideration, a publication appeared that supported these findings. In particular, a C-terminal fragment of murine Spt6 (residues 1162–1726) that contains the SH2 domain did not bind to a Ser2-phosphorylated tandem-repeat CTD peptide, but bound to the N-terminal half of the CTD, which contains a stretch of 15 consensus repeats.[66] Thus, interactions that are not predicted by our modeling (Fig. 4) may occur between the SH2 domain and a longer CTD peptide.

Functional architecture of Spt6

The central region of Spt6 shows homology to the bacterial Tex protein, which is involved in toxin gene expression.[28,29] The recently reported X-ray structure of Tex revealed the folds and relative position of domains HtH, YqgF, HhH, and S1 in the conserved central region.[29] Spt6 additionally contains an acidic N-terminal region, which might interact with nucleosomes, the C-terminal SH2 domain, and short regions flanking the SH2 domain. The previously published model of Spt6 can now be extended with our structure of the SH2 domain (Fig. 5).[29] To construct an updated model, we mapped 13 sequence insertions (Supplementary Data Fig. S1) in Spt6 onto surface regions within the Tex structure (Fig. 5), and added the structure of the SH2 domain. The model illustrates the relative sizes and locations of the various regions of the modular Spt6 protein, and the potential interaction surfaces of Spt6.
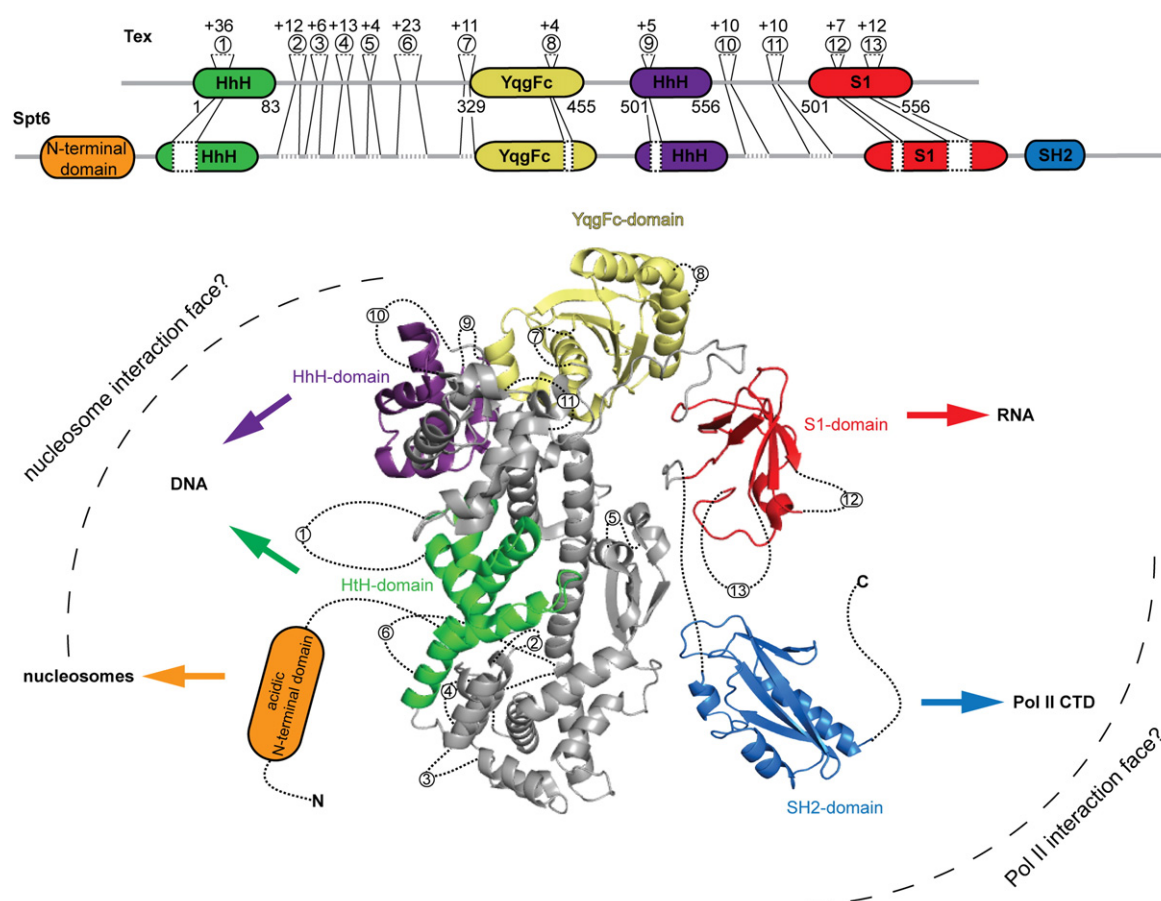

Fig. 5. Updated structural model of Spt6. At the top is a representation of the Tex domain organization. The relative positions of insertions in Spt6 (identified by aligning the *C. glabrata* Spt6 and the *P. aeruginosa* Tex sequences; see Supplementary Data Fig. S1) are indicated by numbers 1–13. Numbers above the insertions indicate their size in amino acid residues. Numbers below the domains indicate their boundaries in the Tex-protein. At the bottom is a ribbon-model for Spt6. The Spt6 core is represented by the Tex structure. Insertions 1–13 are indicated by broken lines. The positions for the SH2 domain and the putative nucleosome-binding domain relative to the rest of the model is chosen randomly. Potential interactions of the different domains are indicated by arrows.

Requirement of the Spt6 SH2 domain *in vivo*

To investigate the importance of the Spt6 SH2 domain *in vivo*, we carried out Affymetrix gene expression profiling (Affymetrix GeneChip Yeast Genome 2.0) with a yeast strain lacking the C-terminal region of Spt6 that includes the SH2 domain[30] (strain spt6ΔC-TAP, Materials and Methods). This mutant strain shows a significant slow-growth phenotype (Supplementary Data Fig. S3). Compared to the wild type strain, 204 out of 5665 genes that were present on the array showed significantly altered mRNA levels using a fold-change cut-off value of >+2.0 or <−2.0. Thus, the Spt6 SH2 domain is necessary for the correct expression of a subset of yeast genes. Of the significantly altered mRNA levels, 101 were up-regulated and 103 were down-regulated.

Western blotting revealed that deletion of the SH2 domain increases the Spt6 protein levels about fivefold *in vivo* (Fig. 6a). To exclude the possibility that the changes in gene expression result from different protein levels rather than deletion of the domain, we generated and analyzed a yeast strain in which full-length Spt6 levels were increased by replacing the endogenous promoter with the strong *adh1* promoter. In this strain, Spt6 protein levels were approximately 15-fold higher than in wild type cells (Fig. 6a), whereas the growth rate was similar to that of the isogenic wild type strain in YPD liquid culture and on YPD solid medium (Supplementary Data Fig. S3). Correlation studies and hierarchical cluster analysis of gene expression profiles from this strain indicated a very high similarity to the wild type strain (Fig. 6c; Supplementary Data Fig. S4). Thus, the changed expression levels observed for the
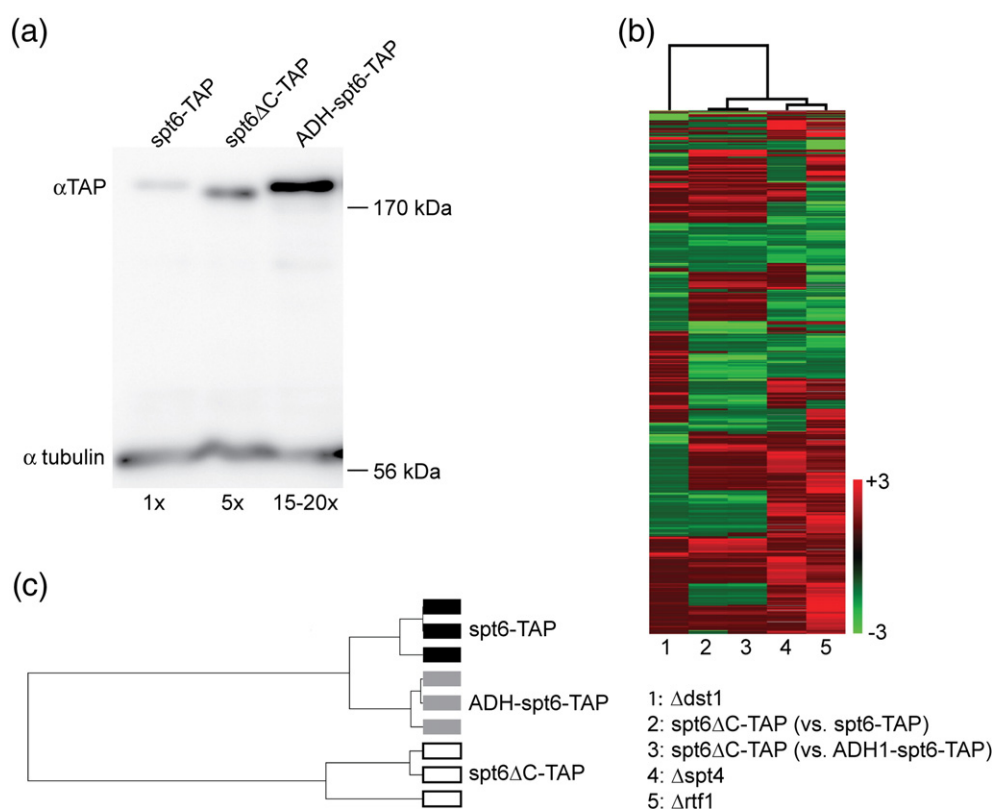
Fig. 6. Deletion of the SH2 domain alters Spt6 levels and the expression of a subset of genes. (a) Western blot showing the relative Spt6 levels in the three strains that were used in gene expression analysis. The approximate level of over-expression is indicated at the bottom (see Materials and Methods). (b) Hierarchical cluster analysis of genes that showed altered mRNA levels (fold change >2.0 or <−2.0) in at least one of the Pol II elongation factor gene deletion strains or the ADH-spt6-TAP strain. In total, the cluster diagram was calculated for 964 yeast genes of *dst1*Δ, spt6Δ C (as compared to a wild type and to an over-expression strain), *spt4*Δ and *rtf1*Δ mutant strains, given in lanes 1, 2, 3, 4 and 5, respectively. Each row corresponds to a particular gene and each column corresponds to a particular elongation factor mutant strain. Changes in mRNA levels compared with the isogenic wild type strain are depicted in red (increase), green (decrease), black (no change; see intensity bar), or gray (no corresponding fold-change value). Both rows and columns were clustered using a hierarchical cluster algorithm (Materials and Methods). The dendrogram for column clustering is shown. (c) Cluster analysis of microarray samples. Each box within the tree represents one microarray sample. The spt6-TAP (wild type), ADH-spt6-TAP (over-expressed wild type), and spt6ΔC-TAP (mutant) samples are indicated in black, gray and white, respectively. Samples with similar gene expression profiles are joined to form distinct clusters. Euclidean distance was used as the distance metric and average linkage as linkage method. The hierarchical clustering was performed with the Partek Genomics Suite 6.4 software package.

spt6ΔC-TAP strain resulted from deletion of the SH2 domain rather than the up-regulation of mutant Spt6 protein levels in this strain.

We analyzed biological processes that were affected significantly by deletion of the SH2 domain with the Gene Ontology Enrichment Analysis Software Toolkit (GOEAST) and the web-based Gene Ontology (GO) tools (Table 2, Materials and Methods). This analysis showed that very diverse biological processes were over-represented, including genes involved in carbohydrate transport, copper ion import, and in the response to pheromones. A deconvolution of gene expression microarray data is generally difficult, as it represents the results of primary and secondary effects during gene expression. Indeed, nine genes of transcription factors[34] are contained in our list of differentially expressed genes (Supplementary Data Fig. S5b), suggesting that secondary effects play a role. We nevertheless aimed to detect a possible global, chromatin-related function of the SH2 domain in the transcriptome data with use of correlation analysis. We first investigated whether the deregulated genes correlate to genes described to show cryptic transcription initiation in a *spt6* mutant[33] that carried an internal deletion of amino acids 931–994 (corresponding to 930–993 in *C. glabrata* and comprising the HhH-domain that is purple in Fig. 5).[9] From the 204 significantly changed genes, only 26 (13%) were among the 960 open reading frames that showed cryptic transcription (Supplementary Data Fig. S5a).[33] We also investigated whether our set of differentially expressed genes shows any correlation with gene length or an unusual number of associated nucleosomes.[32] Again, we could not find significant correlations. These results are consistent with the proposed role of the SH2 domain in targeting Spt6 to transcribed genes, irrespective of the type of gene, and with a general function of Spt6 in chromatin transitions and transcription-coupled processes, rather than a role in the regulation of specific genes.

Cooperating elongation factors have distinct cellular functions

To analyze whether the expression of similar genes is affected by deletion or mutation of different Pol II elongation factor genes, we compared our gene expression data to data available for the yeast strains *dst1Δ* (*DST1* is the gene encoding TFIIS),[67] *spt4Δ*, and *rtf1Δ*.[34] We expected similarity between these data sets, since Spt6 interacts genetically with TFIIS,[5] and with the Rtf1-containing Paf1 complex,[35,36] and since Spt6 binds the Spt4-Spt5 complex.[37] An unsupervised hierarchical cluster analysis showed that the differential expression data from the *spt4Δ* and *rtf1Δ* strains form a distinct cluster within a dendrogram, indicating similarity of their gene expression profiles (Fig. 6b, lanes 4 and 5, Materials and Methods). However, the spt6ΔC-TAP mutant (as compared to the wild type or over-expression strain) exhibits a very different expression profile (Fig. 6b, lanes 2 and 3), suggesting that the function of the Spt6 SH2 domain is clearly distinct from the functions of Spt4-Spt5 and the Paf1 complex *in vivo*. This analysis additionally revealed that *dst1Δ* showed the most distinct expression profile (Fig. 6b, lane 1), perhaps because TFIIS is required during elongation and during initiation.[38,39] Additional correlation studies confirmed these results (Supplementary Data Fig. S4b). Taken together, deletion or mutation of various elongation factor genes results in different changes in the transcriptome, despite the observed genetic and physical interactions between these factors.

Discussion

We report the structure of the Spt6 SH2 domain, the only SH2 domain encoded in the genomes of lower eukaryotes, and demonstrate its importance for normal gene expression *in vivo*. Proteins that contain SH2 domains occur to various extents in the different eukaryotic taxa (Fig. 7a). *Dictyostelium discoideum*, an organism that can switch between single-celled and multicellular lifestyles, contains 12 different SH2 domain-containing polypeptides.[40,41] Human cells contain 120 SH2 domains, distributed over 110 proteins, that are generally involved in phosphotyrosine recognition during signaling.[42] The high abundance of SH2-containing proteins in animal cells likely reflects the need for fine-tuned intercellular signaling. Since Spt6 is found at the root of the eukaryotic tree, its SH2 domain can be regarded as an ancestor of the many structurally diverse SH2 domains in higher eukaryotes (Fig. 7a).

The Spt6 SH2 domain contains structural features from both structural families of SH2 domains that exist in higher eukaryotes. The relationship to Src-type domains in metazoan cells is revealed by the presence of a βE-βF sheet and a phylogenetic tree (Fig. 7b), in which Spt6 clusters together with Src-type family domains. This observation is strengthened by a structural comparison to other SH2 domains using the DALI server,[65] where the Spt6 SH2 domain shows a higher degree of similarity to Src-type than to STAT-type domains (Supplementary Data Table S1). On the other hand, the N-terminal helix αL could resemble part of a primitive

Table 2. GO biological processes over-represented among the differentially expressed genes

GO term	GO ID	p
Carbohydrate transport	0008643	0.0168
Monosaccharide transport	0015749	0.00131
Hexose transport	0008645	0.000212
Copper ion import	0015677	0.00226
Cytogamy	0000755	0.0102
Response to pheromone	0019236	0.0129

The significantly enriched GO terms, the GO identifiers, and the corresponding *p*-values are shown. Subgroups of GO terms (parent–child relationships) are indicated hierarchically. All GO analysis was performed with the web-based software toolkit GOEAST.[60]

(a)

(b)
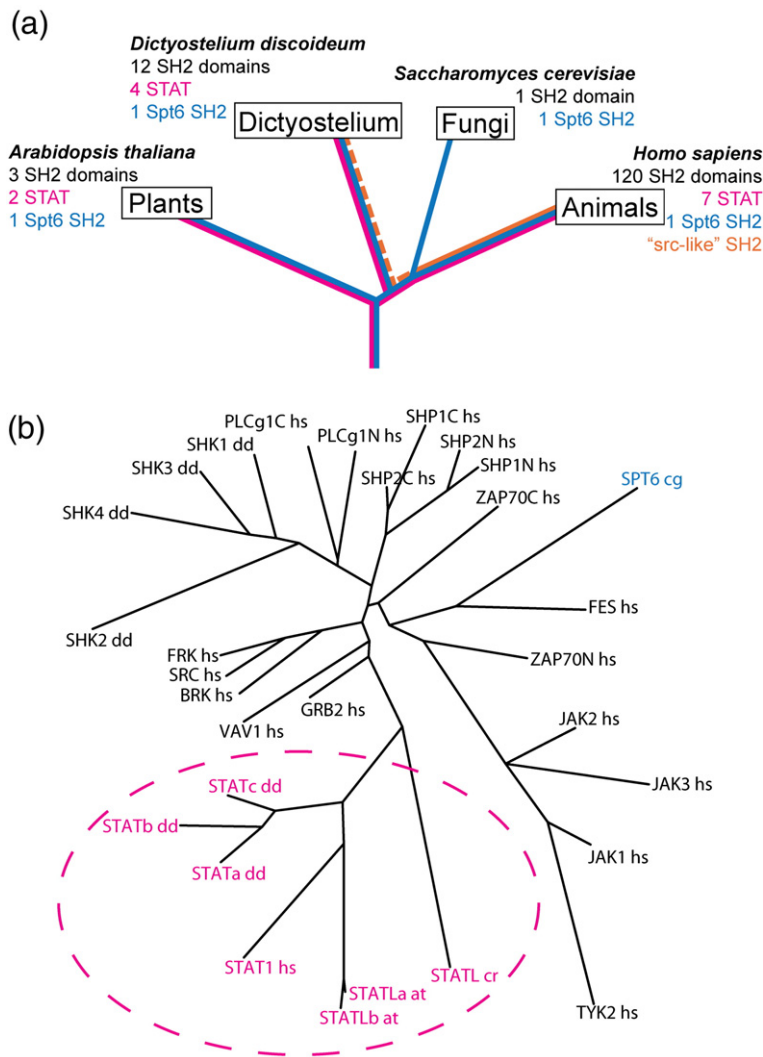
Fig. 7. Evolution of SH2 domains. (a) Unscaled phylogenetic tree of eukaryotic organisms.[41] For every taxon, a representative model organism is indicated together with the number of SH2 domain-containing proteins and STAT proteins characterized so far. Colors of the branches of the tree indicate the distribution of SH2 subfamily-types and Spt6 in the different taxa: blue for Spt6, magenta for STAT-type, orange for Src-like. The broken orange line in the *Dictyostelium* branch shows that the structural nature of some of the SH2-domains in this organism is not completely clear. (b) The Spt6 SH2-domain (blue) clusters with Src-type domains in a phylogenetic analysis. The cluster of STAT sequences is shown in purple. Organisms are indicated as in Fig. 2.

or degenerated linker domain, which was characterized in STAT proteins (Fig. 2). In human STAT-1,[24] and *D. discoideum* STATa,[43] this α-helix is named α11 and αN, respectively (yellow in Fig. 2). In the Spt6 SH2 domain, the N-terminal helix αL is in a fixed position, as it is anchored on the core fold by hydrophobic contacts and a salt bridge between R1251 and D1330. Apparently, helix αL is part of the ancient SH2 domain fold and may have been extended into a helical linker domain during the evolution of STAT proteins. A specific feature of the Spt6 SH2 domain is an unusually long C-terminal helix αB, which has be seen only in the SH2 domain of the adaptor protein APS, a substrate of the insulin receptor.[25] These long C-terminal helices are not observed in Src-type domains (Supplementary Data Table S1) and probably are related to the BB′ helical structures in STAT-type domains.

The Spt6 SH2 domain contains a unique peptide-binding site. Like other SH2 domains, the Spt6 SH2 domain has two surface pockets, pocket 1 and pocket 2, which are likely involved in binding the phosphorylated residue and a residue C-terminal of the phospho-amino acid, respectively. Pocket 1 contains an invariant arginine that can bind the phospho group, but is shallower and lacks residues that could

recognize an aromatic ring, inconsistent with phosphotyrosine binding, but consistent with phosphoserine binding. Pocket 2 is highly negatively charged, in contrast to most other SH2 domains, where it is either positively charged or neutral. These unusual features of the structure strongly suggest the basis of the unique phosphoserine (rather than phosphotyrosine) specificity of this SH2 domain. Conformation of this model awaits a complex structure that we were not able to obtain due to the experimental limitations of the system.

A genomic deletion of the Spt6 C-terminal region that comprises the SH2 domain decreased cell growth rate, increased Spt6 protein levels, and altered the expression levels of a subset of yeast genes. The deregulation of these genes was generally due to the deletion of the SH2 domain, since over-expression of wild type Spt6 did not lead to any significant changes in the transcriptome. The deregulated genes did not correlate with the gene length, the number of associated nucleosomes, or the previously reported set of genes that show cryptic transcription upon Spt6 mutation. These results support the view that Spt6 has multiple functions, including nucleosome assembly, RNA splicing, and mRNA export. Taken together, our

results provide structural information on the unique SH2 domain in yeast, demonstrate the relevance of this domain *in vivo*, and help us to understand the unusual target specificity and evolutionary implications of this domain.

Materials and Methods

Sample preparation

DNA containing the SH2-domain of Spt6 (residues 1250–1348) was amplified by polymerase chain reaction from *C. glabrata* genomic DNA and was cloned into the pET28b expression vector (Novagen). This results in an open reading frame that encodes the SH2-domain with an N-terminal His$_6$ tag connected by a thrombin cleavage site. The sequence was modified by introduction of codons for methionine residues (L1309 and L1317) by the overlap extension method. The resulting variant was used for selenomethionine-labeling as described.[44,45] The SH2 domain was expressed overnight at 18°C in *Escherichia coli* BL23 (DE3) CodonPlus RIL cells (Stratagene). Cells were harvested, resuspended in 50 mM Tris–HCl at pH 8.0, 1 M NaCl, 10 mM β-mercaptoethanol and lysed by sonication. The cell extract was cleared by centrifugation and subjected to affinity chromatography on a nickel-nitrilotriacetic acid column (Qiagen), followed by an overnight cleavage of the His$_6$ tag with thrombin at 4°C and dialysis. The tag-free protein was collected by passage through a second Ni-NTA column. The flowthrough was concentrated and subjected to size-exclusion chromatography (Superose12, GE Healthcare) equilibrated with 50 mM Tris pH 8.0, 200 mM NaCl, 5 mM dithiothreitol. Peak fractions were collected and pure proteins were concentrated to 15 mg/ml.

Limited proteolysis

For trypsin and chymotrypsin treatment of purified protein samples, 100 μl of purified Spt6 SH2-domain at a concentration of 1 mg/ml was mixed with 1 μg of the corresponding protease. The reaction was carried out in the size-exclusion buffer, supplemented with 1 mM CaCl$_2$. The reaction mixture was incubated at 37°C. Aliquots of 10 μl were taken at different timepoints (usually after 1, 3, 5, 10, 30, and 60 min) and the reaction was stopped immediately by the addition of 5 × SDS sample buffer and incubation at 95°C for 5 min.

All samples were analyzed by SDS-PAGE. The N-termini of digestion products were analyzed by Edman-sequencing.

Crystal structure analysis

Crystals of form A were grown at 20°C by the hanging-drop, vapor-diffusion method using a reservoir solution containing 100 mM Hepes pH 7.0, 1 M succinic acid, 1% (v/v) polyethylene glycol 2000 monomethyl ether (MME), 5 mM Tris(2-carboxyethyl) phosphine hydrochloride. Crystals grew to a maximum size of 0.5 mm × 0.1 mm × 0.1 mm within one week and were transferred into reservoir solution containing 13% polyethylene glycol 400 or 15% (v/v) glycerol and flash-cooled in liquid nitrogen. For crystals of form B, a reservoir solution of 100 mM bicine pH 8.0, 4.3 M NaCl and 5 mM Tris(2-carboxyethyl) phosphine hydrochloride was used. Crystals grew to a maximum size of 0.3 mm × 0.15 mm × 0.15 mm within three days and were flash-cooled in liquid nitrogen. Synchrotron radiation data were collected at Berliner Elektronen speicherring - Gesellschaft für Synchrotron-strahlung m.b.H. (BESSY) BL 14.1. Multiwavelength anomalous diffraction data were collected from a seleno-methionine-labeled crystal form A to a resolution of 1.9 Å. For a selenomethionine-labeled crystal of form B, single wavelength anomalous diffraction data were collected to a resolution of 2.4 Å. Data were processed with HKL[46] or XDS.[47] Selenium sites in crystal form A were located with HKL2MAP,[48] and multiwavelength anomalous diffraction phases were calculated with SHARP.[49] A model of the SH2 domain was built with Coot,[50] and refined with CNS.[51] For crystal form B, phases were calculated by molecular replacement using PHASER,[52] with one molecule of the asymmetric unit of crystal A as a search model. Model building was done as for crystal A, and Refmac5 was used for refinement.[53] The X-ray and refinement statistics are given in Table 1.

Phylogenetic analysis

Phylogenetic studies were based on published work.[17] The amino acid sequences of the proteins were downloaded from *Swiss-Prot/TrEMBL*.[54] The corresponding SH2 domains were predicted by *SMART* database,[55] and aligned with *ProbCons*.[56] The re-formatting of the alignment file to PHYLIP format was performed with *Reformat* (Bioinformatics Toolkit Max-Planck Institute for Developmental Biology). The unrooted phylogenetic tree was calculated with *Proml* (Protein maximum likelihood) and drawn by *Drawtree*, which are tools of the PHYLIP software package version 3.67.[57]

Yeast strains and cell growth for gene expression analysis

S. cerevisiae strain S288C (MATa; SUC2; gal2; mal; mel; flo1; flo8-1; hap1; ho; bio1; bio6) was used for all genetic manipulations . A C-terminal TAP-tag was introduced to full-length Spt6 or to a truncated version (deletion of residues 1250–1451) as described,[68] yielding the strains spt6-TAP and spt6ΔC-TAP. The full-length spt6-TAP strain was used for replacement of the Spt6-promoter with an ADH1 promoter,[69] yielding the strain ADH-spt6-TAP. Expression of the TAP-tagged proteins was verified by resolving the total protein by SDS-PAGE (8% (w/v) polyacrylamide gel) and blotting onto a PVDF membrane. The membrane was probed with antibodies against the TAP-tag (PAP, Sigma) and tubulin (3H3087, Santa Cruz Biotechnology) as a loading control. Bound antibodies were detected by chemiluminescence (ECL Plus Western Blotting detection system, GE Healthcare). For tubulin, peroxidase-conjugated AffiniPure Rabbit Anti-Rat IgG (H+L) (Jackson ImmunoResearch) was used for detection. Detection was carried out with the LAS-4000 mini detection system (Fuji) and signals were quantified in relation to the tubulin loading control using the Multi Gauge software (Fuji). All TAP-tagged strains were grown at 25°C in YPD medium. Pre-cultures were grown to an absorbance at 600 nm (A_{600}) of 10–12 in the case of Spt6-Tap and the ADH1-spt6-TAP strains and to an A_{600} of ~7 in the case of the spt6ΔC-TAP strain. These pre-cultures were used to inoculate 30 ml of YPD to reach an A_{600} of 0.1. Cells were grown at 25°C to mid-log phase (A_{600} 0.8) and then harvested. The final A_{600}

values were consistently reached after 7 h or 13 h for the wild type and mutant strains, respectively. Cells were pelleted by centrifugation (2 min at 4000 rpm and 20°C, Rotana 460R, Hettich). The supernatant was discarded, and the pellets were flash-frozen in liquid nitrogen, and stored at –80°C. Biological triplicates were grown from three distinct yeast colonies.

RNA preparation

Total RNA was extracted with the RiboPure-Yeast Kit (Ambion/Applied Biosystems) according to the manufacturer's instructions. Any remaining DNA was removed by the treatment with DNase I that was part of the RiboPure-Yeast Kit. RNA quality was assessed by 2% (w/v) agarose gel electrophoresis (RNAse-free ready-to-use gels, Invitrogen). RNA concentrations and purity were determined by measuring the absorbance at 260 nm and at 280 nm (ND-1000 spectrophotometer, NanoDrop Technologies). The A_{260}/A_{280} ratios were in the range 2.10–2.20 for all RNA samples. The RNA yield was usually 80–150 µg.

Microarray handling

RNA was hybridized to GeneChipYeast Genome 2.0 arrays (Affymetrix) as described in the users' manual, using the GeneChip Expression 3′ Amplification One-Cycle Target Labeling and Control Reagents kit. Briefly, 1 µg of total RNA was reverse transcribed using a bacteriophage T7 oligo(dT) primer. Following second-strand cDNA synthesis, the double-stranded cDNA was purified as a template for the subsequent *in vitro* transcription reaction. Linearly amplified, biotin-labeled complementary RNA (cRNA) was synthesized in the presence of a biotinylated nucleotide analog. The labeled cRNA was purified and the quantity of each cRNA sample was determined by measurement of absorbance at 260 nm and at 280 nm (ND-1000 spectrophotometer, NanoDrop Technologies). The quantities were >90 µg of cRNA for all samples, and 15 µg of each cRNA sample was fragmented by metal ion-catalyzed hydrolysis. The success of the cRNA fragmentation process was evaluated by 4% agarose gel electrophoresis (ready-to-use gels, Invitrogen). A 5 µg sample of the labeled and fragmented cRNA was hybridized to the arrays at 45°C for 16 h with constant rotation at 60 rpm in a GeneChip Hybridization Oven 640 (Affymetrix, Santa Clara, CA). Washing and staining of the arrays was done with a FS450_0003 script of the Affymetrix GeneChip Fluidics Station 450. The arrays were scanned using an Affymetrix GeneChip Scanner 3000 7G. To minimize errors, samples were processed in parallel and arrays were scanned the same day.

Gene expression data analysis

Raw signal intensities for each probe set in the .*CEL* files were analyzed using version 6.4 of the PARTEK GENO-MICS SUITE software (Partek Inc.). Data were filtered by application of an expanded mask file that was based on the *s_cerevisiae.msk* file of Affymetrix, to mask the *S. pombe* probe sets, unspecific probe sets, and replicate probe sets of *S. cerevisiae*. The robust multiarray average (RMA) normalization method[58] was used for RMA background correction, quantile normalization and medianpolish probe set summarization. Expression values were transformed to $\log_2$ before statistical analysis. A sample intensity plot was calculated, showing that the data

were normally distributed for all samples, as well as the homogeneity of variance (no outlier), which also indicated the great similarity of biological replicates. A hierarchical cluster analysis of samples supported the latter point, since the biological triplicates of the mutant strain and of the wild type strain form two highly distinct clusters (Fig. 6b). Genes that were expressed differentially between wild type and mutant strains were detected with one-way analysis of variance (ANOVA), a statistical technique used to compare means of two or more samples implemented in PARTEK. A linear contrast was used to compare mutant samples with baseline wild type or over-expression samples, respectively. The recovered *P*-values of the comparisons were then corrected using a step-up false discovery rate value of 5%.[59] The resulting list of significantly differentially expressed genes was filtered to include only genes that demonstrated twofold or greater up- or down-regulation. For gene ontology (GO) analysis of significantly over-represented biological processes, the list of Affymetrix probe set identification numbers of differentially expressed genes was analyzed with the Gene Ontology Enrichment Analysis Software Toolkit GOEAST,[60] which is based on the GO database.[61] Only over-represented biological process terms with a recovered *P*-value <0.02 were considered.

Hierarchical cluster and correlation analysis

Hierarchical cluster analysis was performed for gene expression data of strains dst1Δ (deletion of the gene encoding TFIIS), spt4Δ, rtf1Δ, and spt6ΔC-TAP (as compared to wild type as well as to the over-expression strain) yeast strains. The Spt4 as well as the Rtf1 data are from Ref. 34 and were downloaded from the Longhorn Array Database (ExptID: *5649*, *5360*, *5387*, and *5618*).[62] The log-fold change values of the averaged data sets were anti-logged and values <1 were transformed ($-1/x$) to yield symmetrically distributed fold change values. In total, the hierarchical cluster analysis was performed for yeast genes that showed a fold change of >2.0 or <–2.0 in at least one of the mutants (a total of 964 genes). Hierarchical cluster analysis was calculated using the TIGR MeV application,[63] thereby choosing average linkage as the linkage method and Euclidean distance as the distance metric. Pearson's correlation was calculated in R version 2.8.1. and Microsoft Excel. The respective correlation coefficient (*r*-value) was calculated for each pair of elongation factor mutant strains and was based on the respective lists of significantly altered genes.

Protein Data Bank and ArrayExpress accession numbers

Coordinates and structure factors have been deposited in the Protein Data Bank with accession numbers *3GXW* and *3GXX* for crystal forms A and B, respectively. Microarray data were deposited in the ArrayExpress data bank with accession number s *E-MEXP-2123*.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.04.016

References

1. Winston, F., Chaleff, D. T., Valent, B. & Fink, G. R. (1984). Mutations affecting Ty-mediated expression of the HIS4 gene of Saccharomyces cerevisiae. *Genetics*, **107**, 179–197.
2. 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.
3. Swanson, M. S., Carlson, M. & Winston, F. (1990). SPT6, an essential gene that affects transcription in Saccharomyces cerevisiae, encodes a nuclear protein with an extremely acidic amino terminus. *Mol. Cell Biol.* **10**, 4935–4941.
4. Swanson, M. S. & Winston, F. (1992). SPT4, SPT5 and SPT6 interactions: effects on transcription and viability in Saccharomyces cerevisiae. *Genetics*, **132**, 325–336.
5. Hartzog, G. A., Wada, T., Handa, H. & Winston, F. (1998). Evidence that Spt4, Spt5, and Spt6 control transcription elongation by RNA polymerase II in Saccharomyces cerevisiae. *Genes Dev.* **12**, 357–369.
6. Andrulis, E. D., Guzman, E., Doring, P., Werner, J. & Lis, J. T. (2000). High-resolution localization of Drosophila Spt5 and Spt6 at heat shock genes in vivo: roles in promoter proximal pausing and transcription elongation. *Genes Dev.* **14**, 2635–2649.
7. 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.
8. Bortvin, A. & Winston, F. (1996). Evidence that Spt6p controls chromatin structure by a direct interaction with histones. *Science*, **272**, 1473–1476.
9. Kaplan, C. D., Laprade, L. & Winston, F. (2003). Transcription elongation factors repress transcription initiation from cryptic sites. *Science*, **301**, 1096–1099.
10. 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.
11. 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.
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. Maclennan, A. J. & Shaw, G. (1993). A yeast SH2 domain. *Trends Biochem. Sci.* **18**, 464–465.
14. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. (1991). SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science*, **252**, 668–674.
15. Pawson, T., Gish, G. D. & Nash, P. (2001). SH2 domains, interaction modules and cellular wiring. *Trends Cell Biol.* **11**, 504–511.
16. 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.
17. Gao, Q., Hua, J., Kimura, R., Headd, J. J., Fu, X. Y. & Chin, Y. E. (2004). Identification of the linker-SH2 domain of STAT as the origin of the SH2 domain using two-dimensional structural alignment. *Mol. Cell Proteomics*, **3**, 704–714.
18. Kuriyan, J. & Cowburn, D. (1997). Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288.
19. Pendergast, A. M., Muller, A. J., Havlik, M. H., Maru, Y. & Witte, O. N. (1991). BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. *Cell*, **66**, 161–171.
20. Muller, A. J., Pendergast, A. M., Havlik, M. H., Puil, L., Pawson, T. & Witte, O. N. (1992). A limited set of SH2 domains binds BCR through a high-affinity phosphotyrosine-independent interaction. *Mol Cell Biol*, **12**, 5087–5093.
21. Maignan, S., Guilloteau, J. P., Fromage, N., Arnoux, B., Becquart, J. & Ducruix, A. (1995). Crystal structure of the mammalian Grb2 adaptor. *Science*, **268**, 291–293.
22. Nolte, R. T., Eck, M. J., Schlessinger, J., Shoelson, S. E. & Harrison, S. C. (1996). Crystal structure of the PI 3-kinase p85 amino-terminal SH2 domain and its phosphopeptide complexes. *Nature Struct. Biol.* **3**, 364–374.
23. Becker, S., Groner, B. & Muller, C. W. (1998). Three-dimensional structure of the Stat3beta homodimer bound to DNA. *Nature*, **394**, 145–151.
24. Chen, X., Vinkemeier, U., Zhao, Y., Jeruzalmi, D., Darnell, J. E., Jr & Kuriyan, J. (1998). Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. *Cell*, **93**, 827–839.
25. Hu, J., Liu, J., Ghirlando, R., Saltiel, A. R. & Hubbard, S. R. (2003). Structural basis for recruitment of the adaptor protein APS to the activated insulin receptor. *Mol. Cell*, **12**, 1379–1389.
26. Waksman, G., Kominos, D., Robertson, S. C., Pant, N., Baltimore, D., Birge, R. B. *et al.* (1992). Crystal structure of the phosphotyrosine recognition domain SH2 of v-src complexed with tyrosine-phosphorylated peptides. *Nature*, **358**, 646–653.
27. Kimber, M. S., Nachman, J., Cunningham, A. M., Gish, G. D., Pawson, T. & Pai, E. F. (2000). Structural basis

for specificity switching of the Src SH2 domain. *Mol. Cell*, **5**, 1043–1049.

28. 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.

29. 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*.

30. Youdell, M. L., Kizer, K. O., Kisseleva-Romanova, E., Fuchs, S. M., Duro, E., Strahl, B. D. & Mellor, J. (2008). Roles for Ctk1 and Spt6 in regulating the different methylation states of Histone H3 lysine 36. *Mol. Cell Biol.* **28**, 4915–4926.

31. 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.

32. Lee, W., Tillo, D., Bray, N., Morse, R. H., Davis, R. W., Hughes, T. R. & Nislow, C. (2007). A high-resolution atlas of nucleosome occupancy in yeast. *Nature Genet.* **39**, 1235–1244.

33. Cheung, V., Chua, G., Batada, N. N., Landry, C. R., Michnick, S. W., Hughes, T. R. & Winston, F. (2008). Chromatin- and transcription-related factors repress transcription from within coding regions throughout the Saccharomyces cerevisiae genome. *PLoS Biol.* **6**, e277.

34. Hu, Z., Killion, P. J. & Iyer, V. R. (2007). Genetic reconstruction of a functional transcriptional regulatory network. *Nature Genet.* **39**, 683–687.

35. Costa, P. J. & Arndt, K. M. (2000). Synthetic lethal interactions suggest a role for the Saccharomyces cerevisiae Rtf1 protein in transcription elongation. *Genetics*, **156**, 535–547.

36. Mueller, C. L. & Jaehning, J. A. (2002). Ctr9, Rtf1, and Leo1 are components of the Paf1/RNA polymerase II complex. *Mol. Cell Biol.* **22**, 1971–1980.

37. Krogan, N. J., Kim, M., Ahn, S. H., Zhong, G., Kobor, M. S. & Cagney, G. (2002). RNA polymerase II elongation factors of Saccharomyces cerevisiae: a targeted proteomics approach. *Mol. Cell Biol.* **22**, 6979–6992.

38. Kim, B., Nesvizhskii, A. I., Rani, P. G., Hahn, S., Aebersold, R. & Ranish, J. A. (2007). The transcription elongation factor TFIIS is a component of RNA polymerase II preinitiation complexes. *Proc. Natl Acad. Sci. USA*, **104**, 16068–16073.

39. Guglielmi, B., Soutourina, J., Esnault, C. & Werner, M. (2007). TFIIS elongation factor and Mediator act in conjunction during transcription initiation in vivo. *Proc. Natl Acad. Sci. USA*, **104**, 16062–16067.

40. Williams, J. G., Noegel, A. A. & Eichinger, L. (2005). Manifestations of multicellularity: Dictyostelium reports in. *Trends Genet.* **21**, 392–398.

41. Eichinger, L., Pachebat, J. A., Glockner, G., Rajandream, M. A., Sucgang, R. & Berriman, M. (2005). The genome of the social amoeba Dictyostelium discoideum. *Nature*, **435**, 43–57.

42. Liu, B. A., Jablonowski, K., Raina, M., Arce, M., Pawson, T. & Nash, P. D. (2006). The human and mouse complement of SH2 domain proteins-establishing the boundaries of phosphotyrosine signaling. *Mol. Cell*, **22**, 851–868.

43. Soler-Lopez, M., Petosa, C., Fukuzawa, M., Ravelli, R., Williams, J. G. & Muller, C. W. (2004). Structure of an activated Dictyostelium STAT in its DNA-unbound form. *Mol. Cell*, **13**, 791–804.

44. Budisa, N., Steipe, B., Demange, P., Eckerskorn, C., Kellermann, J. & Huber, R. (1995). High-level biosynthetic substitution of methionine in proteins by its analogs 2-aminohexanoic acid, selenomethionine, telluromethionine and ethionine in Escherichia coli. *Eur. J. Biochem.* **230**, 788–796.

45. Meinhart, A., Blobel, J. & Cramer, P. (2003). An extended winged helix domain in general transcription factor E/IIE alpha. *J. Biol. Chem.* **278**, 48267–48274.

46. Otwinowski, Z. & Minor, W. (1996). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.

47. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallogr.* **26**, 795–800.

48. Pape, T. & Schneider, T. R. (2004). HKL2MAP: a graphical user interface for phasing with SHELX programs. *Acta Crystallogr. D*, **37**, 843–844.

49. Terwilliger, T. C. (2002). Automated structure solution, density modification and model building. *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1937–1940.

50. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D*, **60**, 2126–2132.

51. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P. & Grosse-Kunstleve, R. W. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D*, **54**, 905–921.

52. McCoy, A. J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr D*, **63**, 32–41.

53. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D*, **53**, 240–255.

54. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A. & Gasteiger, E. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365–370.

55. 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.

56. Do, C. B., Mahabhashyam, M. S., Brudno, M. & Batzoglou, S. (2005). ProbCons: probabilistic consistency-based multiple sequence alignment. *Genome Res.* **15**, 330–340.

57. Felsenstein, J. (1989). PHYLIP - phylogeny inference package (version 3.2). *Cladistics*, **5**, 164–166.

58. Irizarry, R., Hobbs, B., Collin, F., Beazer-Barclay, Y., Antonellis, K., Scherf, U. & Speed, T. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, **4**, 249–264.

59. Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. B*, **57**, 289–300.

60. Zheng, Q. & Wang, X. J. (2008). GOEAST: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Res.* **36**, W358–W363.

61. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H. & Cherry, J. M. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genet.* **25**, 25–29.

62. Killion, P. J., Sherlock, G. & Iyer, R. (2003). The Longhorn Array Database (LAD): an open-source, MIAME compliant implementation of the Stanford

Microarray Database (SMD). *BMC Bioinformatics*, **4**, 32.

63. Saeed, A. I., Sharov, V., White, J., Li, J., Liang, W. & Bhagabati, N. (2003). TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, **34**, 374–378.

64. Labarga, A., Valentin, F., Anderson, M. & Lopez, R. (2007). Web services at the European bioinformatics institute. *Nucleic Acids Res.* **35**, W6–W11.

65. Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138.

66. 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.

67. Koschubs, T., Seizl, M., Lariviere, L., Kurth, F., Baumli, S., Martin, D. E. & Cramer, P. (2009). Identification, structure, and functional requirement of the Mediator submodule Med7N/31. *EMBO J.* **28**, 69–80.

68. Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E. & Bragado-Nilsson, E. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, **24**, 218–229.

69. Janke, C., Magiera, M. M., Rathfelder, N., Taxis, C., Reber, S. & Maekawa, H. (2004). A versatile toolbox for PCR-based tagging of yeast genes: new fluorescent proteins, more markers and promoter substitution cassettes. *Yeast*, **21**, 947–962.

70. Ardehali, M. B., Yao, J., Adelman, K., Fuda, N. J., Petesch, S. J., Webb, W. W. & Lis, J. T. (in press). Spt6 enhances the elongation rate of RNA polymerase II in vivo. EMBO J. 2009 Mar 12. [Epub ahead of print].