nature structural & molecular biology

The multifunctional human p100 protein 'hooks' methylated ligands

Neil Shaw[1,6], Min Zhao[2,6], Chongyun Cheng[1], Hao Xu[2], Juha Saarikettu[3], Yang Li[1], Yurong Da[4], Zhi Yao[4], Olli Silvennoinen[3], Jie Yang[4], Zhi-Jie Liu[1], Bi-Cheng Wang[2] & Zihe Rao[1,5]

The human p100 protein is a vital transcription regulator that increases gene transcription by forming a physical bridge between promoter-specific activators and the basal transcription machinery. Here we demonstrate that the tudor and SN (TSN) domain of p100 interacts with U small nuclear ribonucleoprotein (snRNP) complexes, suggesting a role for p100 in the processing of precursor messenger RNA. We determined the crystal structure of the p100 TSN domain to delineate the molecular basis of p100's proposed functions. The interdigitated structure resembles a hook, with a hinge controlling the movement and orientation of the hook. Our studies suggest that a conserved aromatic cage hooks methyl groups of snRNPs and anchors p100 to the spliceosome. These structural insights partly explain the distinct roles of p100 in transcription and splicing.

p100, encoded by *SND1*, is a ubiquitous, multifunctional protein that can interact with and modulate a broad spectrum of proteins involved in transcription[1–4]. Originally identified as a vital cellular component that enhances the transcription of EBNA-2–activated gene expression[5], p100 was later shown to have similar effects in activating additional transcription factors, such as STAT5 (ref. 6) and STAT6 (ref. 1). The underlying mechanism of p100-mediated coactivation seems to be similar among these factors and is dependent on p100's ability to interact with the basal transcription machinery. In all studies of this coactivation, p100 has been found to interact *in vitro* and *in vivo* with specific transcription activators and to form a physical bridge in their associations with components of the basal transcription machinery. Recently, p100 has been implicated in the pathogenesis of autosomal-dominant polycystic kidney disease (ADPKD)[7]. Notably, p100 is also a known component of the RNA-induced silencing complex (RISC), promoting cleavage of double-stranded RNA and hyperedited double-stranded RNA substrates[8]. These studies suggest that p100 may have several distinct roles.

The modular architecture of the p100 protein is well suited for participation in protein-protein interactions. Hydrophobic cluster analysis (HCA) of p100 has revealed multiple staphylococcal nuclease (SN)-like domains at the N terminus, and tudor and SN domains (which we term the TSN domain) at the C terminus[9,10]. It has been postulated that the SN-like domains of p100 have evolved into protein-protein–interacting domains. Experimental evidence seems to support this hypothesis, but structural evidence is still missing. The STAT5TAD, STAT6TAD, CBP, RNA polymerase II, RNA helicase A, Myb and Pim1 serine/threonine kinases have been shown to interact with the p100 protein through the SN-like domains[1–6]. The TSN domain of p100 shows similarity to the TSN domain of the survival of motor neurons (SMN) protein, which acts as a transcriptional coactivator and is also involved in splicing of precursor mRNA (pre-mRNA)[11,12].

Previously, we have shown that the SN-like domain of p100 alone is sufficient to enhance STAT6-mediated gene activity in response to interleukin-4 stimulation, whereas expression of the TSN domain does not affect transcriptional activity[1]. More recently, we found that the TSN domain of p100 interacts with U5 snRNP–specific proteins and promotes pre-mRNA splicing (J.Y. and O.S., unpublished data). This suggests that human p100 protein plays multiple roles via different functional domains. To delineate the molecular bases of the proposed functions, we determined the three-dimensional structure of the p100 TSN domain.

RESULTS

Overall structure of p100 TSN domain

The three-dimensional structure solution of the human p100 C-terminal TSN domain (residues 654–870) was determined at 2.0-Å resolution using X-ray crystallography. The crystallographic asymmetric unit contains two fragments of this C-terminal domain, probably originating from different truncations and crystal-packing preferences during crystallization. The larger fragment is composed of complete, interdigitated tudor and SN domains, with residues 654–870 clearly visible in the electron density map. A short fragment

[1]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China. [2]Southeast Collaboratory for Structural Genomics, Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia 30602, USA. [3]Institute of Medical Technology, University of Tampere and Tampere University Hospital, FIN-33014 Tampere, Finland. [4]Department of Immunology, Tianjin Medical University, Tianjin, 300070, China. [5]Laboratory of Structural Biology, Life Science Building, Tsinghua University, Beijing 100084, China. [6]These authors contributed equally to this work. Correspondence should be addressed to Z.-J.L. (zjliu@ibp.ac.cn) or J.Y. (yangj@tijmu.edu.cn).

Received 24 December 2006; accepted 8 June 2007; published online 15 July 2007; doi:10.1038/nsmb1269

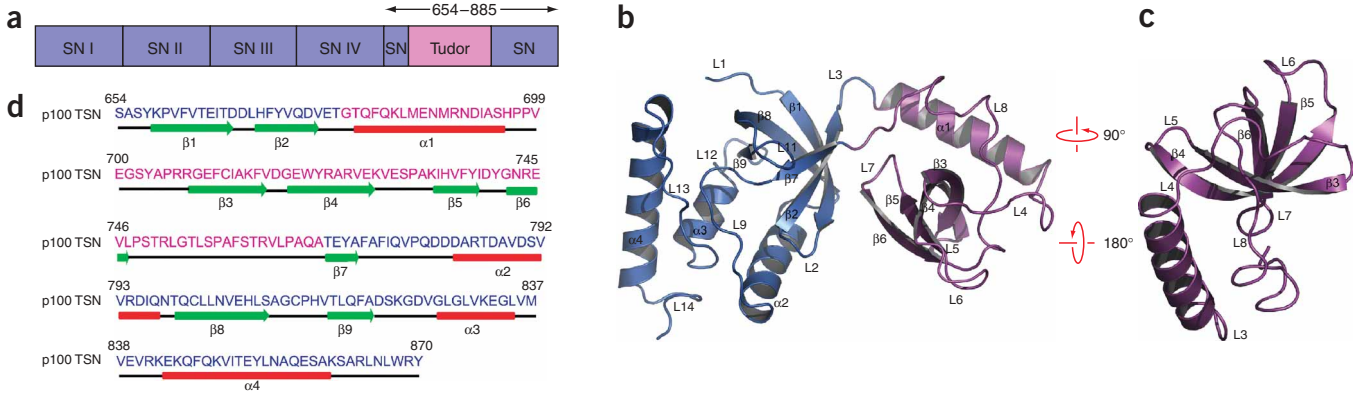

Figure 1 Overall structure of the p100 TSN domain. (**a**) Diagram of the human p100 protein architecture, showing the five SN-like domains and the tudor domain. (**b**,**c**) Cartoon illustration of the C-terminal human p100 TSN domain structure. Blue, SN domain; pink, tudor domain. (**d**) Amino acid sequence of interdigitated p100 TSN domain, with secondary structural annotations. Residues are colored by domain as in **b**.

made up of residues 680–770, representing the tudor domain alone, is also packed in the asymmetric unit along with the larger fragment (**Supplementary Fig. 1** online). This was confirmed by SDS-PAGE analysis of the protein obtained by harvesting and dissolving the crystals in buffer. Electron density for residues 870–885 is missing from the structure.

The human p100 protein has modular architecture and is predicted to be made up of four complete SN-like domains at the N terminus, and an incomplete SN-like domain and a tudor domain formed by residues 654–885 at the C terminus[9,10] (**Fig. 1a**). Our structure of the human p100 TSN domain shows four α-helices, nine β-sheets and 14 loops (**Fig. 1**). The SN domain is composed of two segments: residues 654–678, which form two sheets (β1 and β2), and residues 769–870, which form the remaining three sheets (β7, β8 and β9) and helix (α2). Residues 679–768 (β3, β4, β5 and β6) form a typical β-barrel tudor domain (**Fig. 1c**). In addition, a long helix (α1) and numerous loops (L3 through L8) are part of the tudor domain.

Although HCA analysis previously predicted an incomplete SN domain at the C terminus of p100, our TSN domain structure reveals a complete SN domain, including a typical oligosaccharide/oligonucleotide-binding (OB) fold[13]. The secondary structural elements of the SN domain are interdigitated with the tudor domain. The Protein Data Bank was searched using WU-BLAST (http://blast.wustl.edu/) to identify structures similar to the p100 TSN domain. Although no matches ($E < 1$) were found, the overall architecture of the human p100 SN domain is similar to the staphylococcal nuclease (SNase) structure[14] (PDB 1SNC), with an r.m.s. deviation of 1.86 Å over 118 of 135 SNase main chain Cα atoms (**Fig. 2a**). Differences are found in the loop regions implicated in DNA binding (loop L_{45} linking β3 with α1, and loop $L_{3\alpha}$ linking β4 with β5, according to the OB fold nomenclature of ref. 13). The amino acids necessary for catalysis are missing in the structure of the p100 SN domain. The positions of Asp21 and Asp40 in SNase are occupied by Leu669 and Gln777 in the p100 SN domain. Similarly, the catalytic amino acids Arg35, Glu43 and Arg87 in SNase are replaced by Ala772, Gln780 and Cys812, respectively, in the p100 SN domain. These substitutions may compromise the nuclease activity of the p100 protein, as reported[9,10].

Although the overall architecture of the secondary structural elements is identical between p100 SN and SNase, the nature of the surface residues differs. The p100 SN domain has a large stretch of negatively charged surface, predominantly occupied by aspartate residues. Residues 654–678 and 781–797, containing a total of nine aspartates and two glutamates, impart the negative charge to this region (**Fig. 2b**). Such negatively charged patches are absent in the SNase structure, and these patches could potentially mediate ionic protein-protein interactions of p100 SN.

Human p100's four-stranded β-barrel is similar in architecture to the SMN tudor domain[12]. Superimposition of the SMN tudor domain (PDB 1MHN) onto the p100 tudor domain reveals little difference between the two structures: the Cα atoms of 53 of the 59 SMN tudor domain residues overlap with an r.m.s. deviation of 1.2 Å (**Fig. 2a**). Differences between these two structures are found mainly at L5, the loop linking β3 with β4, and L6, the loop linking β4 with β5 (**Fig. 2a**; the secondary structural elements are named as described for the p100 TSN domain). The β-barrel of human p100 tudor is stabilized by a hydrophobic core, which is conserved in almost all proteins containing the tudor domain (**Fig. 3a**), composed of residues Phe715, Tyr721, Tyr738 and Tyr741. A similar hydrophobic core in the SMN tudor domain has been suggested to bind the dimethylated arginine/glycine-rich tail of the SN core protein[12].

Two antiparallel loops connect the SN-like domain with the tudor domain. One of the loops has a helix embedded inside. This L3-α1-L4 region works together with the long L8 loop like a hinge (**Fig. 1b**). The α1 helix helps maintain the distance and orientation of the tudor domain.

HCA of p100 and modeling of the p100 N terminus

Because HCA did not predict the interdigitation of the p100 SN domain's secondary structural elements, the complete set of residues involved in the formation of the SN domain was not identified in previous studies. Using the new information about these residues obtained from the structure of the C-terminal p100 SN domain, we carried out HCA again (**Supplementary Fig. 2** online) and compared the results with the previously reported HCA of the four N-terminal SN-like domains of human p100 (refs. 9,10). The hydrophobic core is similar overall, with similar secondary structural elements in all five SN-like domains of the p100 (data not shown). Next, we aligned the sequences of all five SN-like domains of p100 with that of the SNase. This revealed a number of conserved residues, including the absolutely conserved Leu835 (**Fig. 2b**). We then constructed a three-dimensional model of the four N-terminal SN-like domains of p100 using Geno 3D[15]. The crystal structure of SNase (PDB 1SNC) was used as

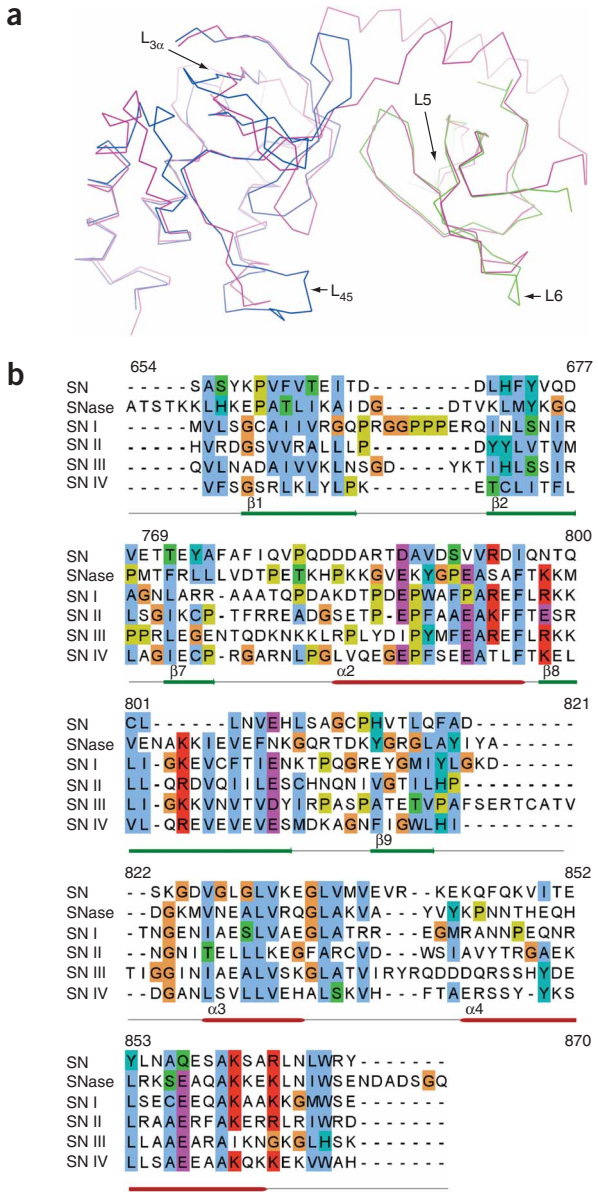

Figure 2 Analysis of the p100 TSN region. (**a**) Superimposition of Cα atoms of SNase (PDB 1SNC) and SMN tudor domain (PDB 1MHN) on the C-terminal p100 TSN domain structure. Main chains of p100 TSN, SNase and SMN tudor are colored magenta, blue and green, respectively. (**b**) Multiple sequence alignment of the C-terminal SN domain of p100 with SNase and the four N-terminal SN-like domains of p100. Conserved residues are highlighted in color according to the ClustalW convention (http://www.ebi.ac.uk/clustalw).

a template for model building. Although the overall topologies of the four SN-like domains are similar, the pattern of surface charge distribution is remarkably different for SN IV. Whereas SN I, SN II and SN III have positively charged surfaces, SN IV has a negatively charged surface, similar to that of the tudor domain (see below). In addition, there are differences among all the modeled SN domains at loop L_{45} (linking β3 with α1) and loop $L_{3\alpha}$ (linking β4 with β5).

p100 tudor domain binds methylated ligands

Proteins have evolved intricate mechanisms for recognition of covalent modifications such as methylation of lysine and arginine residues. Gene expression is regulated precisely and efficiently by physical rearrangement of the chromatin structure in response to the methylation marks found on histone proteins[16]. Arginine methylation is also routinely used as a signal for recognition of partners and assembly of the spliceosome[17,18]. It is likely that the modules involved in recognition and binding of methylated amino acid residues use a common mechanism, in which methyl groups are trapped inside a cage comprising at least three aromatic amino acid residues (**Fig. 3**). The methyl carbon is highly polarized because of an adjacent, strongly electron-withdrawing nitrogen, and this protonated carbon is localized in space by ionic interactions with the ψ electrons, effectively securing the methyl group. The peptide binding specificity is determined by the nature of the residues surrounding the cage. In the structures of proteins known to bind methylated ligands, including SMN (PDB 1MHN), HP1 (1GUW), JMJDA (2GFA), 53BP1 (1XNI) and Polycomb (1PDQ)[19–23], such an aromatic cage is implicated in recognition and binding of methylation marks; the human p100 tudor domain structure reveals an identical rectangular aromatic cage. In p100, three tyrosine residues (Tyr721, Tyr738 and Tyr741) and a phenylalanine residue (Phe715) form the cage, whereas in the SMN tudor domain three tyrosines (Tyr109, Tyr127, Tyr130) and a tryptophan (Trp102) enclose a dimethylated arginine ligand, and in the Polycomb chromodomain the trimethylated lysine ligand is enclosed by Tyr4, Trp47, Trp50 and Tyr54 (ref. 23) (**Fig. 3a**). A recent study describes a similar evolutionarily conserved mechanism for recognition of methylated ligands by an aromatic cage in the 53BP1 tudor domain[24]. Our structure of the p100 tudor domain also reveals the fortuitous caging of the Leu808 dimethyl group by another C-terminal fragment of the same protein packed in the crystal. The dimethyl group of Leu808 is caged by the same aromatic residues predicted to bind methylated ligands (**Fig. 3c,d**).

A deep, negatively charged cleft is evident at the interface of p100's tudor and SN domains, composed of the carboxyl oxygens of Glu664, the main chain carbonyl oxygens of Gln767, Thr663, Val673, Ala768 and Tyr741, and the hydroxyl oxygen of Tyr672. The close proximity of this negatively charged patch to the aromatic cage implicated in binding of methyl groups suggests a role for the charged region in determining ligand binding specificity (**Fig. 3b**). Proteins known to bind methylated ligands have similar negatively charged surfaces[19–23]. The SMN tudor domain binds the positively charged arginine/glycine-rich tails of Sm proteins through negatively charged surface residues; similarly, the JMJD2A double tudor domain uses negatively charged groups surrounding the aromatic cage to bind positively charged histone peptides (**Fig. 3b**).

p100 interacts with snRNP particles

We have previously shown that the SN-like domains of p100 protein recruit CBP histone acetyl transferase (HAT) activity to STAT6, facilitating access of the STAT6–p100 complex to the basal transcriptional machinery[1]. Recently, we identified a group of U5 snRNP–specific proteins, including the 220-kDa, 200-kDa and 116-kDa proteins, associated with the TSN domain of p100 in an *in vitro* pull-down assay (J.Y. and O.S., unpublished data). To confirm the interaction of p100 protein and U5 snRNP protein, we carried out pull-down assays with glutathione *S*-transferase (GST)-fused p100-SN and p100-TSN proteins. Equal amounts of GST, GST–p100-TSN or GST–p100-SN fusion proteins were bound to glutathione-coupled beads and incubated with nuclear extracts of HeLa cells. Western blotting showed that GST–p100-TSN precipitated PRP8, whereas

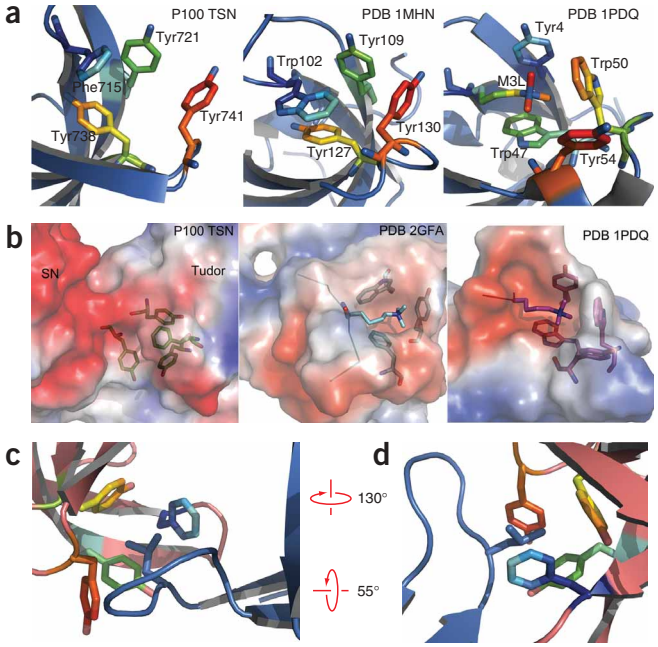

Figure 3 Human p100 binds methylated ligands. (**a**) Aromatic cage implicated in binding of methylated ligands in human p100 TSN domain, SMN tudor domain (PDB 1MHN) and Polycomb chromodomain (PDB 1PDQ). Trimethylated ligand and aromatic residues forming the cage are represented as sticks. (**b**) Environment surrounding binding sites for methylated ligands, characterized by negatively charged surface patches that specifically bind positively charged methylated ligands via ionic interactions. Shown are surface electrostatic potential of human p100 TSN domain, JMJD2A double TSN domain (PDB 2GFA) and Polycomb chromodomain (PDB 1PDQ). Trimethylated substrate residues and aromatic residues forming the cage are represented as sticks. Potentials were calculated with APBS[37] and contoured from $-1\ k_bTe^{-1}$ (red) to $+1\ k_bTe^{-1}$ (blue). (**c**,**d**) Caging of methyl groups within human p100 tudor domain. Methyl groups of Leu808 from chain A (residues 654–870) of the p100 tudor domain are bound by the aromatic cage formed in the hydrophobic core of the TSN domain from chain B (residues 680–770). Leu808 and the aromatic cage residues are shown as sticks.

GST–p100-SN or GST protein alone did not (**Fig. 4**). To provide additional evidence that the p100 TSN domain associates with snRNP particles, we performed GST pull-down assays with nuclear extract of HeLa cells and tested for the presence of snRNAs by northern blotting with ^{32}P-labeled probes specific to U1, U2, U4, U5 or U6 snRNAs. Notably, GST–p100-TSN fusion protein precipitated not only U5 snRNA but also U1, U2, U4 and U6 snRNAs, whereas GST–p100-SN and GST protein did not (**Fig. 4**). This result indicates that p100 protein is associated with snRNP particles through its TSN domain.

Finally, mutagenesis studies support the idea that the p100 TSN domain's conserved aromatic cage binds snRNPs. Double mutants of the aromatic cage (Y738A Y741A or Y721A F715A) did not bind snRNPs, and mutation of either Tyr721, Tyr738, Tyr741 or Phe715 to alanine diminished binding (**Fig. 4d**).

DISCUSSION

The modeled three-dimensional structures of the N-terminal SN-like domains, in combination with the C-terminal TSN structure determined by X-ray crystallography, indicate that the overall structure of full-length p100 resembles a stick with a hook (**Fig. 5a**,**b**). The SN-like domains form the stick, and the tudor domain makes up the hook. The aromatic cage of the tudor domain's hook traps methyl groups of ligands and anchors the p100 protein in the ligand–protein complex. This could be the mechanism for p100's participation in splicing reactions. We have shown that the GST–p100-TSN fusion protein binds snRNPs, such as U1, U2, U4, U5 and U6. The snRNAs associated with these snRNPs are known to contain covalently modified guanosine nucleotides[25]: the 5′ guanosine of snRNA is hypermethylated by the Tgs1 methyltransferase, and the mature snRNAs have a 2,2,7-trimethylated guanosine cap. The p100 TSN domain may hook the methyl caps of snRNAs to anchor the protein to the spliceosomal complex. It is possible that the interactions of p100's N-terminal SN-like domains with the basal transcription machinery components RNA polymerase II, CBP and RNA helicase A form a complex that couples transcription with splicing events. Alternatively, the SN-like domains may have unknown partners in the spliceosome that remain to be detected.

In addition, the structure of full-length p100 reveals a distinct role in transcription and DNA replication independent of splicing events. It is likely that p100 recognizes specific histone methylation marks, binds methylated histones in the aromatic cage of the tudor domain and recruits HAT to unfold the nucleosomes. Once p100 bridges promoter-specific transcription factors and the basal transcription machinery, its DNA-binding ability might serve to anchor them on the DNA. There is also evidence linking the assembly of the mRNA export machinery with transcription and splicing[26]. Thus, p100 may function in transcription, splicing and export of mRNA from the nucleolus. Further experiments are necessary to determine the exact role of p100 in these processes.

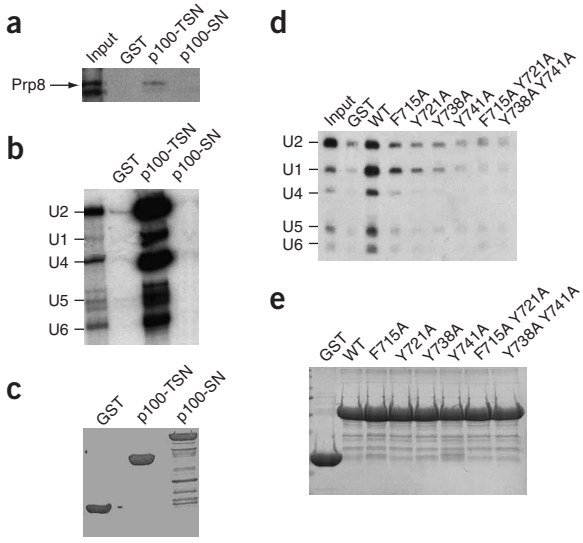

Figure 4 TSN domain of p100 interacts with the U snRNP complex. Nuclear lysates of HeLa cells were incubated with GST alone, or with GST–p100-SN or GST–p100-TSN fusion protein. (**a**) Bound proteins were resolved by SDS-PAGE and immunoblotted with PRP8-specific antibody. (**b**) The coprecipitated RNAs were fractionated on a 7 M urea 6% PAGE gel, analyzed by northern blotting with U1, U2, U4, U5 and U6 snRNA probes and visualized by autoradiography. Positions of snRNAs are indicated on the left. (**c**) Expression of different GST fusion proteins measured by western blotting. (**d**) Coprecipitation of snRNAs with wild-type (WT) or mutant GST–p100-TSN or GST–p100-TSN. Mutants contain alanine substitutions of aromatic residues in the putative methyl-binding cage. (**e**) Expression of WT and mutated GST–p100-TSN proteins measured by western blotting.

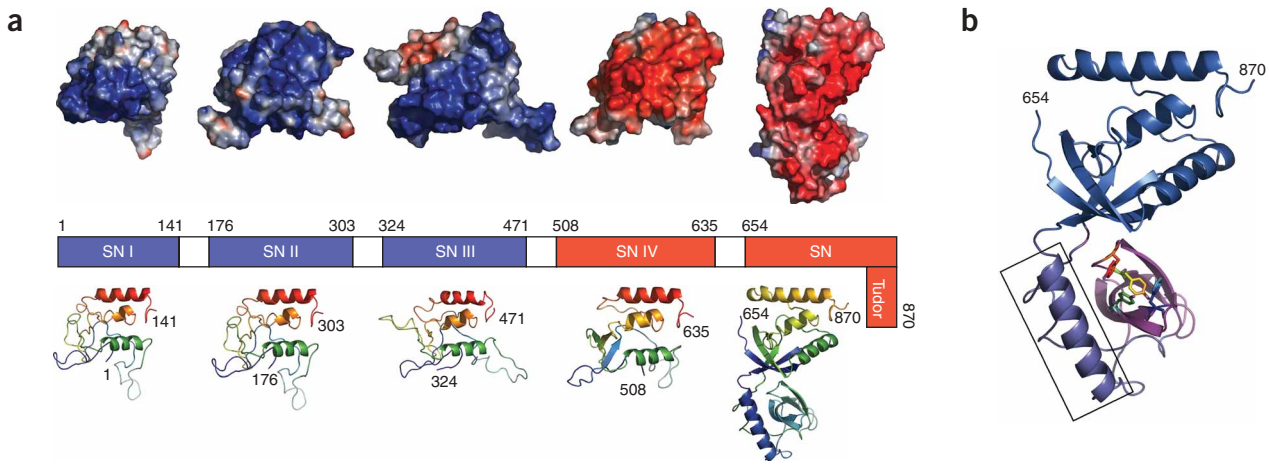

Figure 5 Structure of full-length p100. (**a**) The four N-terminal SN-like domains were modeled using Geno 3D. Surface electrostatic potential plots (as in **Fig. 3b**) of SN domains I–III show positively charged surfaces, in contrast to SN IV and tudor domain, which have negatively charged surfaces. (**b**) Tudor domain structure solved by X-ray crystallography resembles a hook with a hinge (boxed). The hook's conserved aromatic cage implicated in binding ligand methyl groups is represented as sticks.

Experimental evidence collected in the last decade has firmly established the importance of the ubiquitous p100 protein in transcription. The domain boundaries of p100 defined by HCA analysis have formed the basis for interpretation of all the functional studies published so far. Our crystal structure of p100 TSN domain redefines these domain boundaries, indicates the molecular basis of some of p100's proposed functions and establishes a new role for p100 in splicing.

METHODS

Protein preparation. The C-terminal region of human p100 protein containing the tudor and SN domains (residues 654–885) was cloned into the pGEX-4T-1 vector (Amersham Biosciences) and expressed in *Escherichia coli* strain BL21-DE3 (Invitrogen) to produce N-terminal GST-tagged protein with a thrombin cleavage site for removal of the GST tag. Cells were grown in LB medium until the culture reached an A_{600} of 0.8. Recombinant protein was produced by lowering the temperature to 16 °C and inducing the culture with 0.2 mM IPTG for 20 h. Cells were harvested and lysed by sonication. The clarified lysate containing soluble recombinant protein was subjected to GST affinity chromatography followed by thrombin treatment to remove the tag. The protein was further purified using anion-exchange and size-exclusion chromatography. The purified protein was chemically modified using a reductive methylation protocol as described[27]. Native or methylated protein (15–20 mg ml^{-1}) in 20 mM Tris-HCl (pH 7.4), 200 mM NaCl and 1 mM DTT was used for crystallization.

Crystallization and data collection. Crystallization screening was done using commercially available sparse matrix screens (Hampton Research). Trials for native and methylated protein were set up in 2-μl hanging drops containing equal amounts of protein and mother liquor equilibrated over 300 μl of reservoir solution. After 7 d of incubation at 16 °C, the methylated protein crystallized in a mother-liquor solution containing 0.1 M phosphate-citrate buffer (pH 4.2), 20% (w/v) PEG 8,000 and 0.2 M NaCl.

For data collection, the crystals were harvested and frozen with liquid nitrogen. We derivatized the crystals by adding a small grain of potassium tetrachloroplatinate(II) to the drop and soaking for 2 h. All data sets were collected at cryogenic temperatures (100 K) using the frozen crystals. Both native and derivatized data were collected using X-rays generated with a copper target on an FR-E+ SuperBright generator (Rigaku) and focused through VariMax HR optics (Osmic). Both native and derivatized data sets consisting of a single-axis ϕ scan with 360 oscillation images of 1° each were recorded on a R-AXIS IV^{2+} detector (Rigaku) using a crystal-to-detector distance of 200 mm and 240 s exposure time per image. The higher-resolution native data set was collected using synchrotron radiation (beamline 22-ID, Southeast Regional Collaborative Access Team (SER-CAT), Advanced Photon Source, Argonne National Laboratory). The data were processed with HKL2000 (ref. 28).

The crystals belong to the space group $P2_12_12_1$, with unit cell parameters of $a = 49.93$ Å, $b = 93.41$ Å and $c = 95.28$ Å. The asymmetric unit contains two different fragments of the p100 C-terminal domain (identified after the structures were determined): a large fragment containing the TSN domain (residues 654–870) and a short fragment containing the tudor domain (residues 680–770).

Table 1 Data collection and refinement statistics

	Native 1	Native 2	Platinum derivative
Data collection			
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	49.93, 93.41, 95.28	49.81, 93.79, 95.37	49.80, 93.10, 95.02
Resolution (Å)	44.0–2.0 (2.1–2.0)	15.0–3.1 (3.3–3.1)	20.0–2.8 (2.9–2.8)
R_{sym}	0.078 (0.40)	0.138 (0.379)	0.079 (0.349)
$I / \sigma I$	25.8 (4.3)	22.9 (8.2)	41.1 (8.8)
Completeness (%)	99.7 (99.5)	100.0 (100.0)	99.5 (100.0)
Redundancy	8.9 (8.7)	14.4 (14.3)	13.9 (13.4)
Refinement			
Resolution (Å)	44.0–2.0		
No. reflections	29102		
R_{work} / R_{free}	0.233 / 0.249		
No. atoms			
Protein	2,437		
Water	156		
B-factors			
Protein	27.9		
Water	34.0		
R.m.s. deviations			
Bond lengths (Å)	0.009		
Bond angles (°)	1.080		

A total of three crystals were used to solve the structure. Values in parentheses are for highest-resolution shell.

Phasing and structure refinement. The initial phases were determined by SIRAS in SHARP[29] with native and platinum-derivatized data sets collected using in-house FR-E+ X-rays. Several rounds of semiautomated refinement were done using ARP/wARP[30], REFMAC[31] and manual revisions of the models in XFIT[32]. The refinement converged to give the statistics presented in **Table 1**. The final model was validated using MolProbity[33] and PROCHECK[34] before submission to the Protein Data Bank[35].

GST fusion protein constructs and GST pull-down assay. GST–p100-SN and GST–p100-TSN constructs were made and GST pull-down assays done as described[1].

Mutagenesis. GST–p100-TSN mutant constructs were generated using the QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommendations.

RNA extractions and northern blotting. The bead-bound GST fusion proteins were incubated first with HeLa cell nuclear lysates, then in 300 µl PK buffer containing 10 mM Tris-HCl (pH 8.0), 2 mM EDTA, 200 mM NaCl, 0.5% (w/v) SDS and 200 µg ml^{-1} proteinase K (Invitrogen) at 65 °C for 60 min. After phenol-chloroform extraction, RNA was precipitated in ethanol, separated by denaturing 6% (w/v) PAGE, transferred to a nylon filter with semidry blotter (Owl Scientific) in 0.5× Tris-EDTA buffer using a constant 3 mA cm^{-2} current for 1.5–2 h, and cross-linked with a Stratalinker (Stratagene). Hybridization conditions for snRNA blots were as described[36]. Radiolabeled probes for U1, U2, U4, U5 and U6 snRNAs were synthesized by *in vitro* transcription of the linearized snRNA plasmids as described[36].

Accession codes. Protein Data Bank: Coordinates have been deposited with accession code 2O4X.

Note: Supplementary information is available on the Nature Structural & Molecular Biology website.

ACKNOWLEDGMENTS

This work was funded by the 863 (grant 2006AA02A316) and 973 (grant 2006CB910901) projects of the Ministry of Science and Technology of China, the National Natural Science Foundation of China (grants 30670427, 30670441 and 30300070), the US National Institutes of Health (grant 1P50 GM62407), the University of Georgia Research Foundation, the Georgia Research Alliance, Program for New Century Excellent Talents in University (grant NCET-04-0245), Tianjin Municipal Science and Technology Commission (grant 07JCZDJC07300) and the Institute of Biophysics, Chinese Academy of Sciences. Supporting institutions for the SER-CAT 22-ID beamline at the Advanced Photon Source may be found at http://www.ser-cat.org/members.html. Use of the Advanced Photon Source was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract number W-31-109-Eng-38.

AUTHOR CONTRIBUTIONS

N.S., M.Z., C.C. and H.X. contributed to the structural studies. J.Y., J.S., Y.D. and O.S. contributed to the mutagenesis and functional characterization of the p100 TSN domain. Z.J.-L., Y.L. and Z.Y. contributed to data collection and analysis. Z.-J.L., J.Y., Z. R. and B.-C.W. conceived the study and participated in its design and coordination. N.S., Z.-J.L., O.S. and J.Y. drafted the manuscript.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/nsmb/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Yang, J. *et al.* Identification of p100 as a coactivator for STAT6 that bridges STAT6 with RNA polymerase II. *EMBO J.* **21**, 4950–4958 (2002).
2. Valineva, T., Yang, J., Palovuori, R. & Silvennoinen, O. The transcriptional co-activator protein p100 recruits histone acetyltransferase activity to STAT6 and mediates interaction between the CREB-binding protein and STAT6. *J. Biol. Chem.* **280**, 14989–14996 (2005).
3. Leverson, J.D. *et al.* Pim-1 kinase and p100 cooperate to enhance c-Myb activity. *Mol. Cell* **2**, 417–425 (1998).
4. Valineva, T., Yang, J. & Silvennoinen, O. Characterization of RNA helicase A as component of STAT6-dependent enhanceosome. *Nucleic Acids Res.* **34**, 3938–3946 (2006).
5. Tong, X., Drapkin, R., Yalamanchili, R., Mosialos, G. & Kieff, E. The Epstein-Barr virus nuclear protein 2 acidic domain forms a complex with a novel cellular coactivator that can interact with TFIIE. *Mol. Cell. Biol.* **15**, 4735–4744 (1995).
6. Paukku, K., Yang, J. & Silvennoinen, O. TSN and nuclease-like domains containing protein p100 function as coactivators for signal transducer and activator of transcription 5. *Mol. Endocrinol.* **17**, 1805–1814 (2003).
7. Low, S.H. *et al.* Polycystin-1, STAT6, and P100 function in a pathway that transduces ciliary mechanosensation and is activated in polycystic kidney disease. *Dev. Cell* **10**, 57–69 (2006).
8. Caudy, A.A. *et al.* A micrococcal nuclease homologue in RNAi effector complexes. *Nature* **425**, 411–414 (2003).
9. Callebaut, I. & Mornon, J.P. The human EBNA-2 coactivator p100: multidomain organization and relationship to the staphylococcal nuclease fold and to the TSN protein involved in *Drosophila melanogaster* development. *Biochem. J.* **321**, 125–132 (1997).
10. Ponting, C.P. P100, a transcriptional coactivator, is a human homologue of staphylococcal nuclease. *Protein Sci.* **6**, 459–463 (1997).
11. Selenko, P. *et al.* SMN tudor domain structure and its interaction with the Sm proteins. *Nat. Struct. Biol.* **8**, 27–31 (2001).
12. Sprangers, R., Groves, M.R., Sinning, I. & Sattler, M. High-resolution X-ray and NMR structures of the SMN TSN domain: conformational variation in the binding site for symmetrically dimethylated arginine residues. *J. Mol. Biol.* **327**, 507–520 (2003).
13. Murzin, A.G. OB (oligonucleotide/oligosaccharide binding)-fold: common structural and functional solution for non-homologous sequences. *EMBO J.* **12**, 861–867 (1993).
14. Hynes, T.R. & Fox, R.O. The crystal structure of Staphylococcal nuclease refined at 1.7 Å resolution. *Proteins Struct. Funct. Genet.* **10**, 92–105 (1991).
15. Combet, C., Jambon, M., Deleage, G. & Geourjon, C. Geno3D: automatic comparative molecular modeling of protein. *Bioinformatics* **18**, 213–214 (2002).
16. Eissenberg, J.C. & Elgin, C.R. Antagonizing the neighbours. *Nature* **438**, 1090–1091 (2005).
17. Brahms, H., Meheus, L., Brabandere, V., Fischer, U. & Luhrmann, R. Symmetrical dimethylation of arginine residues in spliceosomal Sm protein B/B0 and the Sm-like protein LSm4, and their interaction with the SMN protein. *RNA* **7**, 1531–1542 (2001).
18. Friesen, W.J., Massenet, S., Paushkin, S., Wyce, A. & Dreyfuss, G. SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol. Cell* **7**, 1111–1117 (2001).
19. Nielsen, P.R. *et al.* Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. *Nature* **416**, 103–107 (2002).
20. Jacobs, S.A. & Khorasanizadeh, S. Structure of the HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080–2083 (2002).
21. Huang, Y., Fang, J., Bedford, M.T., Zhang, Y. & Xu, R-M. Recognition of histone H3 lysine-4 methylation by the double TSN domain of JMJD2A. *Science* **312**, 748–751 (2006).
22. Huyen, Y. *et al.* Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature* **432**, 406–411 (2004).
23. Min, J., Zhang, Y. & Xu, R.M. Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **17**, 1823–1828 (2003).
24. Botuyan, M.V. *et al.* Structural basis for the methylation state-specific recognition of histone H4–K20 by 53BP1 and Crb2 in DNA repair. *Cell* **127**, 1361–1373 (2006).
25. Kiss, T. Biogenesis of small nuclear RNPs. *J. Cell Sci.* **117**, 5949–5951 (2004).
26. Reed, R. Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* **15**, 326–331 (2003).
27. Rayment, I. Reductive alkylation of lysine residues to alter crystallization properties of proteins. *Methods Enzymol.* **276**, 171–179 (1997).
28. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
29. de La Fortelle, E. & Bricogne, G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* **276**, 472–494 (1997).
30. Perrakis, A., Morris, R. & Lamzin, V.S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6**, 458–463 (1999).
31. Murshudov, G.N., Vagin, A.A. & Dodson, E.J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol. Crystallogr.* **53**, 240–255 (1997).
32. McRee, D.E. XtalView/Xfit — a versatile program for manipulating atomic coordinates and electron density. *J. Struct. Biol.* **125**, 156–165 (1999).
33. Davis, I.W., Murray, L.W., Richardson, J.S. & Richardson, D.C. MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. *Nucleic Acids Res.* **32**, W615–W619 (2004).
34. Laskowski, R.A., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.* **26**, 283–291 (1993).
35. Berman, H.M. *et al.* The Protein Data Bank and the challenge of structural genomics. *Nat. Struct. Biol.* **7**, 957–959 (2000).
36. Frilander, M.J. & Steitz, J.A. Initial recognition of U12-dependent introns requires both U11/5' splice-site and U12/branchpoint interactions. *Genes Dev.* **13**, 851–863 (1999).
37. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. & McCammon, J.A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. USA* **98**, 10037–10041 (2001).