Published online January 2, 2004

Nucleic Acids Research, 2004, Vol. 32, No. 1 ***201–210***
DOI: 10.1093/nar/gkh163

Domainal organization of the lower eukaryotic homologs of the yeast RNA polymerase II core subunit Rpb7 reflects functional conservation

Sunanda R. Singh, Nambudiry Rekha[1], Beena Pillai, Vijender Singh, Aruna Naorem, Vinaya Sampath, N. Srinivasan[1] and Parag P. Sadhale*

Department of Microbiology and Cell Biology and [1]Molecular Biophysics Unit, Indian Institute of Science, Bangalore-12, India

Received October 22, 2003; Accepted November 12, 2003

ABSTRACT

The subcomplex of Rpb4 and Rpb7 subunits of RNA pol II in *Saccharomyces cerevisiae* is known to be an important determinant of transcription under a variety of physiological stresses. In *S.cerevisiae*, *RPB7* is essential for cell viability while *rpb4* null strains are temperature sensitive at low and high temperatures. The *rpb4* null strain also shows defect in sporulation and a predisposed state of pseudohyphal growth. We show here that, apart from *S.cerevisiae* Rpb7, the Rpb7 homologs from other lower eukaryotes like *Schizosaccharomyces pombe*, *Candida albicans* and *Dictyostelium discoideum* can complement for the absence of *S.cerevisiae RPB7*. This is the first report where we have shown that both the *C.albicans* and *D.discoideum* homologs are functional orthologs of the yeast *RPB7*. We also show that high expression levels of *S.cerevisiae RPB7* and its homologs rescue the sporulation defect of *rpb4* homozygous null diploids, but only some of them cause significant enhancement of the pseudohyphal phenotype. Structural modeling of Rpb7 and its homologs show a high degree of conservation in the overall structure. This study indicates a structural and functional conservation of different Rpb7 across species and also a conserved role of Rpb7 in the subcomplex with respect to nutritional stress.

INTRODUCTION

The core RNA polymerase II (RNA pol II) in eukaryotes comprises more than 10 different subunits. The most well studied core RNA polymerase II from *Saccharomyces cerevisiae* consists of 12 different subunits, Rpb1–Rpb12. Five of them, namely Rpb5, Rpb6, Rpb8, Rpb10 and Rpb12, are shared among the three RNA polymerases (1). All but Rpb4 and Rpb9 are essential for the viability of the cell. However, Rpb4 and Rpb9 are essential for growth at temperature extremes (2,3) and have recently been shown to mediate subpathways of transcription-coupled DNA repair in *S.cerevisiae* (4). The Rpb4–Rpb7 subcomplex is suggested to play an important role in stress response and stress survival (5,6). This subcomplex is easily dissociable and fails to co-purify with the core RNA polymerase II under mild denaturing conditions (7). The subcomplex is shown to bind to single-strand nucleic acid, most likely through a conserved domain of Rpb7, and mediates a post-recruitment step in transcription initiation (8). There are reports that Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4 (9). The role of Rpb4 seems to be to stabilize Rpb7 with the polymerase, but otherwise the function of these two subunits with respect to one another is less clear.

Among the pol II subunits Rpb7 turns out to be highly conserved at the sequence level. There are 13 entries in the seed alignment in the protein families (Pfam) database for eukaryotic homologs and 10 entries for the archaeal homologs of *RPB7* that have been identified through various genome-sequencing projects (10). The human *RPB7* ortholog can complement the *S.cerevisiae RPB7* deletion (6). The orthologs of Rpb4 and Rpb7 from *Arabidopsis thaliana* form a subcomplex tightly associated with the polymerase (11), similar to the ones from *Schizosaccharomyces pombe* (12) and humans (13). The archaeal RNA polymerase subunit RpoE1 was shown to be the bona fide homolog of *S.cerevisiae* Rpb7 (14). The structure of the subcomplex in the archaeal polymerase has been solved (15). While the human orthologs do not appear to interact strongly with their yeast counterparts, they can form heterodimers *in vitro* with archaeal orthologs (14). On the other hand, the *A.thaliana* and *S.pombe* Rpb4 orthologs can interact with yeast Rpb7 (11,12).

Interestingly, the counterpart of Rpb7 in yeast RNA polymerase III (pol III), called C25, was discovered some time ago as an essential protein (16). Two recent reports have further supported the notion that studying the complex is important in understanding the evolution of the basic

*To whom correspondence should be addressed. Tel: +91 80 3942292; Fax: +91 80 3602697; Email: pps@mcbl.iisc.ernet.in
Present address:
B. Pillai, Institute for Genomics and Integrative Biology, New Delhi-7, India

transcription machinery. The reports show that the paralogs of both the subunits as components of RNA polymerase I (pol I) and pol III exist in both *S.cerevisiae* and humans (17,18). In *S.cerevisiae*, A43 polypeptide forms a stable heterodimer with the A14 subunit of pol I and is required for the stabilization of both A14 and ABC23 within pol I. The N-terminal half of A43 is clearly related to Rpb7, and although A14 is different from Rpb4 by specific sequence and structure features, the A43–A14 pair appears to be the pol I counterpart of the Rpb7–Rpb4 heterodimer. The second report describes the C17 subunit of yeast RNA polymerase III (encoded by *RPC17*) that is conserved from yeasts to humans at the sequence level. C17 forms with C25 (encoded by *RPC25*) a heterodimer akin to Rpb4/Rpb7 in Pol II. That CGRP-RCP (the human ortholog of C17) can functionally replace C17 in Δ*rpc17* yeast cells establishes that human CGRP-RCP is the genuine ortholog of C17. These RNA polymerase subunits appear to have evolved to meet the distinct requirements of the multiple forms of RNA polymerases.

We also have recently uncovered many other phenotypes that are affected by Rpb4 and probably by Rpb7 as part of the subcomplex. Apart from the temperature sensitivity that *rpb4* null mutant cells exhibit, we find that *rpb4*Δ/*rpb4*Δ cells show a predisposed phenotype to form pseudohyphae under limiting nitrogen source and reduced sporulation levels under severe carbon starvation, as compared with a wild-type strain (19). We decided to test the conservation of the *RPB7* homologs from *C.albicans*, *S.pombe* and *D.discoideum* in two ways: (i) by modeling the orthologs using the archaeal structure and (ii) by testing complementation of the *RPB7* deletion mutant as well as rescue of the various phenotypes of *RPB4* deletion mutant. We also find that, apart from *S.cerevisiae* Rpb7, the conserved lower eukaryotic homologs of Rpb7 (selected in this study) can complement for the absence of *RPB7*, and their over-expression partially rescues the phenotypes associated with *S.cerevisiae rpb4*Δ/*rpb4*Δ cells mentioned above. These observations and the modeling results point toward a functional as well as structural conservation of Rpb7 across diverse eukaryotic species. A distinct two-domain structure of the protein is obvious, and our chimeric proteins with domains swapped between some of the homologs support this domainal structure.

MATERIALS AND METHODS

Yeast strains

The *S.cerevisiae* strains used in this study were as follows. SYD1011: *Mat a*/*Mat* α, *his3*Δ200/*his3*Δ200, *ura3-52*/*ura3-52*, *leu2-3,112*/*leu2-3,112*, *lys2*/*lys2*, *rpb4*Δ::*HIS3*/*rpb4*Δ::*HIS3*. SYD7: *Mat a*/*Mat* α, *ade2*/*ade2*, *his3*/*his3*, *leu2-3,112*/*leu2-3,112*, *RPB7*/*rpb7*Δ::*LEU2*. YGD-ts22W: *MAT a ura3-1 leu2-3,112 trp1-1 can1-100 ade2-1 his3-11,15 [psi+]*, $ess1^{H164R}$ (20). All the plasmid constructs have been transformed in these strains to generate the required strains. Yeast transformations were done using a modified lithium acetate method without heat shock treatment (21). Plasmids were transformed and amplified in *Escherichia coli* strain DH5α [sup*E44*Δ*lacU169*(ϕ*80lacZ*Δ*M15) hsdR17 recA1 endA1 gyrA96 thi-1relA1*] as per standard protocols (22).

Assay conditions

Complementation. To check for the complementation of *S.cerevisiae rpb7*Δ by other eukaryotic homologs, the heterozygous *RPB7*/*rpb7*Δ strain (SYD7) was transformed with plasmid harboring the homologous open reading frames (ORFs) expressed under the *S.cerevisiae TEF2* promoter. The strains were patched on YEPD plates containing 2% agar. Cells were re-patched on 1% potassium acetate + 2% agar plates for sporulation. Sporulated cells were treated with 0.5 mg/ml lyticase (Sigma Aldrich Ltd, USA) for 30 min and tetrads were dissected on YEPD plates. For random spore analysis (RSA), a suspension of sporulated strains treated with lyticase was made and an equal volume of mineral oil was added to it. The mixture was vortexed for 2 min and spun at 6000 r.p.m. for 3 min. A fixed volume (50 µl) of the mineral oil layer was plated on synthetic dropout (SD) plates lacking uracil and uracil + leucine.

Temperature sensitivity. All strains analyzed for temperature sensitivity were patched on SD plates + 2% glucose + 2% agar. Cells were spotted at equal densities onto plates containing SD medium + 2% glucose + 2% agar. The plates were incubated at 24, 34 and 37°C and photographed after 3 days for *rpb4*Δ *rpb4*Δ and after 5 days for the YGD-ts22W temperature-sensitive strain.

Pseudohyphal growth. Three transformants for each strain were grown on SD plates + 2% glucose + 2% agar and spotted on SLAD + 2% agar plates at suitable cell densities before they were photographed at 18 h using a camera attached to Olympus BX50 microscope. The pseudohyphal analysis with the chimeric proteins was carried out on SD media.

Sporulation. Three transformants for each of the strains were grown on YEPD plates and transferred to 1% potassium acetate plates. Sporulation counts were done using a hemocytometer 72–96 h after transfer to the sporulation medium. The number of tetrads was counted in a minimum of 500 cells and sporulation efficiency expressed in comparison with the wild-type sporulation level taken as 100%.

Construction of plasmids

The construction of *S.cerevisiae RPB4* gene in a CEN plasmid has been described previously (23). Since *rpb4* null mutants are defective for activated transcription while the constitutive transcription is largely unaffected (24), we have used a plasmid pPS189 containing a constitutive promoter, i.e. P_{TEF2} (25), to over-express all the *RPB7* homologs. The *C.albicans RPB7* ORF was amplified from the genomic DNA of *C.albicans CaI4* strain using the forward primer 5′-GCG GAT CCT CAG CAA CAA CAA CAA CAA AAT G-3′ and reverse primer 5′-AAC TGC AGT TAC ATA GGA CTT GGT CCC-3′, and the ~500 bp PCR product was cloned in pGEMT-easy vector (Promega, USA). The ORF was further subcloned in pPS189 at BamHI–XhoI sites. The *D.discoideum RPB7* ORF was cloned as a HindIII–XhoI fragment in pPS189 from the *Sal*I–*Not*I cDNA clone in SSK501 plasmid DNA (a kind gift from Takahiro Morio, Institute of Biological Sciences, Japan). The *S.pombe RPB7* ORF was cloned as a HindIII–SalI insert in pPS189 from the cDNA clone in pGEMT-easy vector

(a gift from Akira Ishihama). The *S.cerevisiae RPB7* ORF was subcloned from pPS41 as a BamHI–XhoI insert in pPS189. The *C.albicans* N-terminus was PCR amplified from pVM314 using the forward primer (above) and reverse primer 5′-GAT TTA AAT GGT TTC CAC ACA ACA G-3′ and cloned in pGEMT-easy vector, digested with BamHI–SwaI and ligated to the DraI–XhoI C-terminal fragment of the *S.cerevisiae RPB7* ORF from pSR262 to construct the *CaN-ScC* chimera. The underlined A residue was substituted instead of G to generate the SwaI site but without changing the corresponding amino acid phenylalanine (Fig. 6A). The *C.albicans* C-terminus was PCR amplified from pVM314 using the forward primer 5′-CTT TAA AGG TGA AGT GGT AGA TGC-3′and reverse primer (above) and cloned in pGEMT-easy, released with DraI–SalI and ligated to the BamHI–DraI N-terminal fragment of the *S.cerevisiae RPB7* ORF to construct the *ScN-CaC* chimera.

Sequence analysis of Rpb7 homologs

The Rpb7 sequences from *S.cerevisiae*, *S.pombe*, *C.albicans* and *D.discoideum* were aligned using CLUSTAL W (26) with the objective of understanding the nature of differences in amino acid residues in close homologs. The residues that are conserved or are conservatively substituted are marked in Figure 1. While conserved or conservatively substituted residues are located most often in the core of the structure, the solvent-exposed conserved residues might play an important role in the function of the protein. Conserved or conservatively substituted solvent-exposed residues are mapped on the modeled structures (Fig. 4). The solvent-accessible surface area has been calculated by running the PSA program as used by Overington *et al*. (27) which employs the algorithm of Lee and Richards (28).

Comparative modeling of *Saccharomyces cerevisiae* Rpb7 using an archaeal homolog

The RpoE1 subunit of *Methanococcus jannaschii* RNA polymerase is a homolog of Rpb7 from *S.cerevisiae* with known 3D structure (15). The structure of this archaeal protein is available in the Protein Data Bank (PDB ID, 1GO3) (15). We have used this structure as a template to model Rpb7 from *S.cerevisiae*. The sequence identity between the archaeal Rpb7 and *S.cerevisiae* Rpb7 is only about 25% (Table 1). However, the archaeal sequence identifies the *S.cerevisiae* sequence in a simple iterative blast search (29) with significant *E*-values of 3×10^{-28} at the round of convergence, which meant that the two proteins under consideration are remotely related. A sequence alignment between the archaeal and *S.cerevisiae* protein sequences was obtained using MALIGN (30). A 3D model of Rpb7 has been generated using a suite of programs encoded in COMPOSER (31,32) and incorporated in SYBYL (Tripos Inc., St Louis, MO). The structurally conserved regions are extrapolated to *S.cerevisiae* Rpb7 sequence. The rest of the regions (often loops) that show high divergence from template structure were modeled by identifying suitable loop regions from a database of non-homologous protein structures. This has been done by a template-matching approach, wherein a search is made for loop segments with the required number of residues that match with end-to-end distances of the structurally conserved regions of the three anchor Cα on either side of the loop. The results so obtained are ranked (33). The best ranking loop with no short contact with the rest of the structure was fitted (F.Eisenmenger, unpublished results). Side chains are modeled on the template structure at equivalent positions as seen in the template structure wherever appropriate or by using the rules from analysis of known structures (34). The model so obtained was subjected to energy minimization steps to relieve short contacts, if any. This was done using the AMBER force field (35) encoded in SYBYL software. The initial rounds of minimization were carried out by keeping backbone atoms fixed, and in subsequent cycles the backbone atoms were allowed to move. The final round of energy minimization was carried out keeping the electrostatics term on and all the atoms in the model were allowed to move. This approach ensured that the models generated are free of short contacts and bad geometries. Although, during the course of this work, structures of the 12-subunit complex of RNA polymerase from *S.cerevisiae* have been published (36,37), the structures are of low resolution (4.1 and 4.2 Å) and for Rpb7 and Rpb4 the positions of Cα atoms alone are available. Hence detailed analysis of the Rpb7 structure addressing specific residue variation is not possible. So we generated a model for Rpb7 using the archaeal complex structure and we ensured that the features of the model built are consistent with the features observed in the low-resolution structure of yeast Rpb7 in the complex. The models for Rpb7 sequences from other sources, like *S.pombe*, *C.albicans* and *D.discoideum*, whose sequences show significant sequence identity with that from *S.cerevisiae* was similarly carried out as discussed above by generating an alignment between the modeled *S.cerevisiae* sequence and sequences from other sources under consideration. All models were energy minimized as discussed.

RESULTS

Sequence conservation of Rpb7 homologs and their functional complementation in *S.cerevisiae*

Rpb7 is essential for cell viability in *S.cerevisiae* (38). It is a highly conserved protein and its conserved counterparts have been reported in archaea and in lower and higher eukaryotes. The sequences of the homologs from three other lower eukaryotic model systems, namely the pathogenic yeast *C.albicans*, the fission yeast *S.pombe* and the cellular slime mold *D.discoideum* were compared with the *S.cerevisiae* sequence (Table 1 and Fig. 1). Besides the central stretch of about 20 amino acid residues which has been known to be essential for the function of *RPB7*, the similarity between the sequences extends to both the N-terminal as well the C-terminal portion of the homologs (16). The high sequence similarity of this protein from different systems prompted us to check the extent to which other eukaryotic homologs of *S.cerevisiae* Rpb7 can functionally complement the absence of *RPB7* in *S.cerevisiae.* The tetrad analysis as well as random spore analysis gave comparable results with all the homologs (Fig. 2A and B). We find that these eukaryotic counterparts can functionally complement the absence of *RPB7* in *S.cerevisiae.*

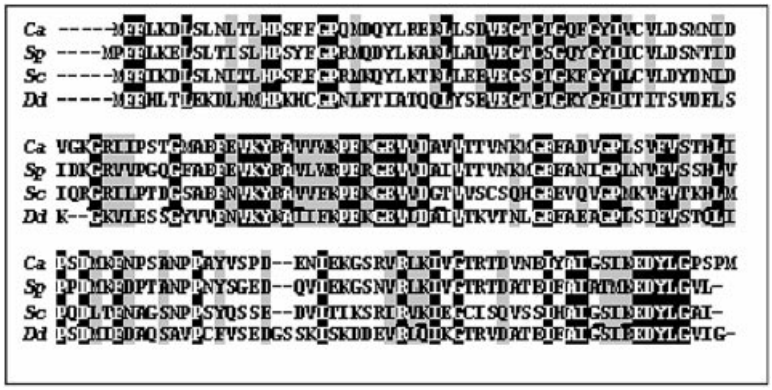

Figure 1. Multiple sequence alignment of the lower eukaryotic homologs of *S.cerevisiae* Rpb7 obtained using CLUSTAL W. Black bars indicate identical residues and gray bars indicate conservative substitutions. The solvent-exposed residues (shown as black bars under the alignment) between *S.cerevisiae* and its homologs are conserved or conservatively substituted in all identified homologs of Rpb7.

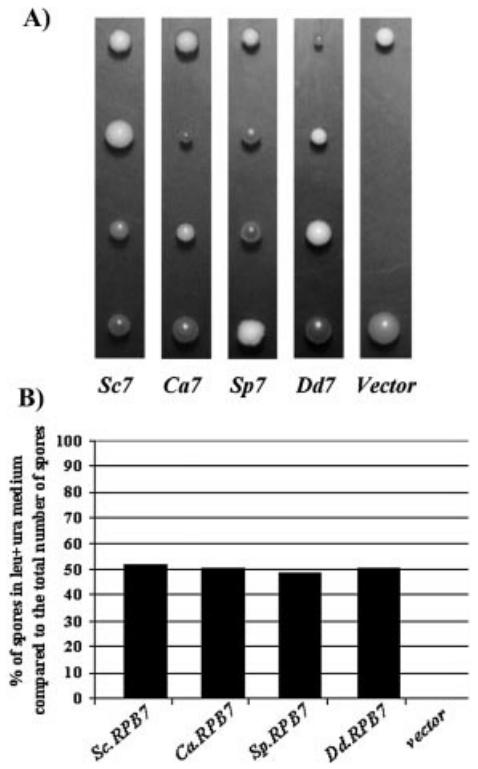

Figure 2. Complementation of *S.cerevisiae RPB7* by lower eukaryotic homologs. The representative tetrads dissected from each of the heterozygous *rpb7*Δ diploid strain transformed with either the vector or the respective homolog in the same vector are shown in (**A**). The corresponding strains were also analyzed by random spore analysis and the ratio of colonies formed on leucine and uracil dropout medium to the number of colonies on the uracil dropout medium is shown in (**B**). Vector transformants do not survive on the leucine/uracil dropout medium.

Saccharomyces cerevisiae Rpb7 and its homologs partially rescue the phenotypes associated with *rpb4* deletion in *S.cerevisiae*

Temperature sensitivity. Previous reports have shown that *S.cerevisiae* Rpb4 is required for cell growth at both low and high temperatures (2). This subunit is also required to be associated with the core enzyme for transcription at high temperatures (39,40). Studies in our laboratory and by other workers have shown that over-expression of its interacting partner Rpb7 can rescue the growth at lower temperatures and at 34°C (9,23). In order to test the functional conservation in rescuing this growth defect, we over-expressed these homologs in *rpb4* deletion diploid and found that they could rescue growth up to 34°C like the *S.cerevisiae* counterpart (Fig. 3A).

Pseudohyphal growth. The morphogenesis of pseudohyphal growth that *S.cerevisiae* diploid cells undergo under limiting nitrogen starvation conditions is well known (41). We have reported earlier that *rpb4*Δ*/rpb4*Δ cells mimic such a starvation response even under non-starved conditions, exhibiting a predisposition to form pseudohyphae (19). We find that over-expression of *RPB7* in this background exaggerates this phenotype of pseudohyphal growth (Fig. 3B). While the exaggerated response is less with the homologous proteins from *S.pombe* and *D.discoideum*, it is most impressive with Rpb7 from *C.albicans*, which incidentally is known to undergo yeast to hyphal form transition under a variety of stress conditions (42).

Sporulation. Under extreme nitrogen starvation but in the presence of a non-fermentable carbon source, *S.cerevisiae* diploid cells undergo meiosis resulting in four ascospores that are enclosed in an ascus. The *rpb4*Δ*/rpb4*Δ cells are defective for sporulation suggesting a requirement for Rpb4 for efficient sporulation (19). We find that this defect can be partially overridden by the over-expression of *RPB7* and its homologs, although the *S.cerevisiae* protein rescues the defect marginally better than the rest (Fig. 3C).

Comparison of the modeled structures shows a high degree of structural conservation between the lower eukaryotic homologs of Rpb7

The archaeal structure of RpoE1 (PDB ID, 1GO3) was used as template for modeling the structure of *S.cerevisiae* Rpb7. The model generated is shown in Figure 4. The modeled structure shows an N-terminal RNP domain and a C-terminal OB-fold region as seen in the archaeal structure. Our analysis of the sequences of Rpb7 shows that the residue level variation among the homologs reported here is concentrated at the C-terminal OB-fold region as indicated in the sequence

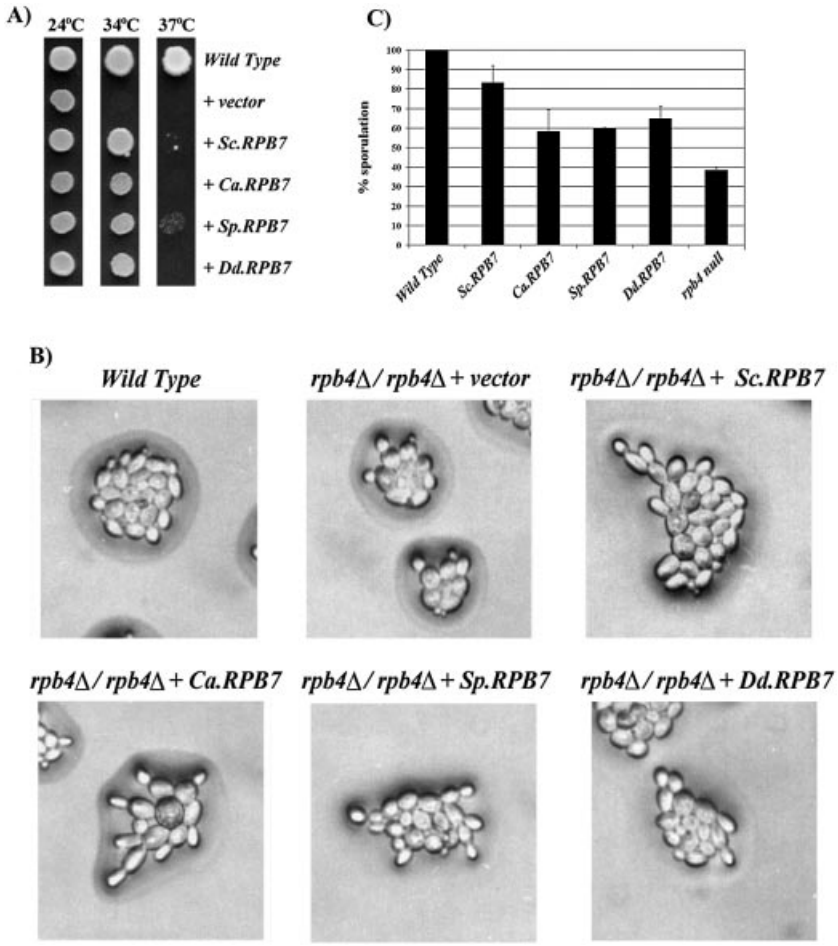

Figure 3. *S.cerevisiae RPB7* and its eukaryotic homologs on over-expression rescue the phenotypes associated with *rpb4* homozygous diploid null strain. (**A**) Cells of transformants expressing each of the homolog were spotted at equal density and plates incubated at temperatures indicated above each panel. (**B**) Individual representative colonies of each of the above transformants were photographed following 18 h incubation on SLAD medium. Photographs were taken at 400× magnification. (**C**) Sporulation percentage was determined as described in Materials and Methods and represented in comparison to the wild-type sporulation level taken as 100%.

alignment (Fig. 1). The aromatic residues in β-strand 3, the loop regions connecting the β-strands 1 and 2 (L12) and the loop between β-strands 4 and 5 (L45) present the binding site for the oligonucleotide (43). These regions are marked in Figure 4 in the context of the present study. In all the proteins that have the OB fold, the L45 region is known to show large sequence variation and is implicated in RNA binding (43,44). Our analysis shows that most of the sequence variations are concentrated at the L3α region and the L45 region of the OB fold. One of the most conserved regions among the four sequences under consideration is the β3 and L12 region of the protein. The RNA binding property of β-strand 3 and L12 is believed to contribute towards stabilizing the RNA through stacking interactions (45) brought about by the aromatic conserved Phe (Phe 99 and Phe 109) present in this region. The other surface-exposed conserved regions map to the Rpb4 binding region and N-terminal RNP domain as shown in Figure 4. These regions are identified by aligning sequences of the Rpb7 homologs available in public databases. Most of the substitutions are in the surface-exposed regions of the structure and not in the proposed Rpb4 interaction regions (marked by the green patches on the structure). When the extent of residue level conservation at the surface-exposed residues is analyzed, we pick up two conserved patches, one in the Rpb4 binding site and the other located at the 'tip' of the S1 domain. The surface-exposed nature of a residue is extrapolated from the archaeal structure. The sequence identity between the Rpb7 proteins across eukaryotic species is remarkable (>60%). The Pfam database considers all the bona fide members of a sequence family while constructing seed alignment. This seed alignment has 23 sequences, each from a different species, and 43 identifiable homologs to the bona fide 23 members. We analyzed the residue level variation in the sequences between these homologs, and when these variations were mapped onto the Rpb7 fold (Fig. 5; variation indicated as red patches) we observe that the variations are not at the sites of interaction with Rpb4 and are concentrated in the OB-fold region.

Rpb7 homologs show significant functional differences in the N- and C-terminal domains

As discussed above the structure modeling shows two clear domains in the Rpb7 protein structure. In an attempt to understand the role of the different homologs of Rpb7 and its

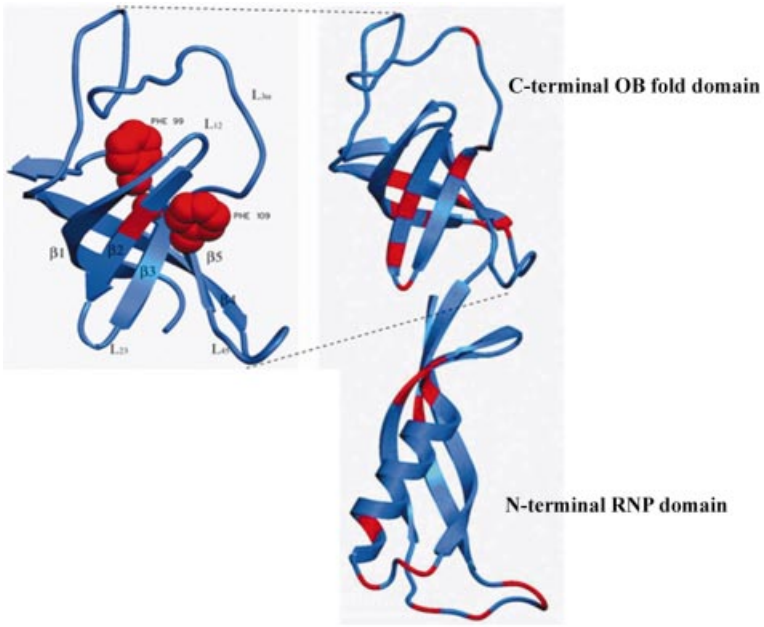

Figure 4. Molecular model of *S.cerevisiae* Rpb7 generated based on the archaeal structure. The N-terminal RNP domain and the C-terminal OB-fold region are indicated. The regions marked β2β3L12L3αL45 forms the RNA binding region in the OB fold. Surface-exposed patches in Rpb7 conserved across organisms are shown in red. The figure was generated using SETOR (53).

Table 1. Sequence comparison of the three Rpb7 homologs with RpoE1 of *M.jannaschii* and *S.cerevisiae* Rpb7

Rpb7p homologs	Sequence length	% comparison with RpoE1		% comparison with *S.cerevisiae*	
		Identity	Similarity	Identity	Similarity
S.cerevisiae	171	25	51	-	-
C.albicans	173	22	47	62	79
S.pombe	172	26	48	52	75
D.discoideum	172	22	43	39	60

domainal organization, we chose to compare the domains of *S.cerevisiae* and the *C.albicans* homologs. The choice of these homologs was based on the fact that the *C.albicans* homolog shows distinct phenotypic differences as reported above and has also been shown to rescue temperature sensitive phenotype of *ess1* alleles (20) while the *S.cerevisiae* Rpb7 does not. We created the chimeric Rpb7 proteins of *S.cerevisiae* with *C.albicans* by swapping the N-terminal RNP domain and C-terminal OB-fold region of the protein (Fig. 6A). We chose the invariable region PF-KGEVVD, which links the N- and C-termini, as the demarcation for the N- and C-terminal domains (see Materials and Methods). The two chimeric fusion proteins (*ScN-CaC* and *CaN-ScC*) were analyzed for all the phenotypes mentioned above.

Both the chimera can complement for the absence of *S.cerevisiae* Rpb7, suggesting that the fundamental transcriptional requirement of Rpb7 can be compensated by such chimeric proteins (data not shown). We tested if these constructs can rescue the phenotypes associated with the *rpb4* diploid null strain. We find significant differences in phenotypes contributed by either of the domains from the two homologs (Fig. 6B). Interestingly, the chimera with the C-terminal domain of the *C.albicans* homolog, namely *ScN-CaC*, can rescue the temperature sensitivity and the sporulation defect of *rpb4* diploid null strain, while the other chimera *CaN-ScC* shows extensive exaggeration of pseudohyphal growth (Fig. 6B). This observation suggests that the former two phenotypes may be co-regulated while the pseudohyphal difference might be regulated through interactions of proteins with the N-terminal domain. Similarly, exaggeration of pseudohyphal morphology was found to be associated with the *HsN-ScC* chimera between human and the yeast homologs (data not shown).

It was reported earlier that the Rpb7 homolog of *C.albicans*, and not the *S.cerevisiae* homolog, can rescue temperature sensitivity of certain alleles of Ess1, a peptidyl prolyl isomerase, at 37°C in *S.cerevisiae* (20). Ess1 is known to be involved in modification of CTD of Rpb1 and regulation of gene expression in *C.albicans* as well as *S.cerevisiae* (20,46). We chose the $Ess1^{H164R}$ temperature-sensitive allele to test if the Rpb7 homologs under study and the chimera can suppress the temperature sensitivity. We find that, besides *C.albicans* Rpb7, the *S.pombe* and *D.discoideum* homologs can also suppress temperature sensitivity at 37°C in *S.cerevisiae*.

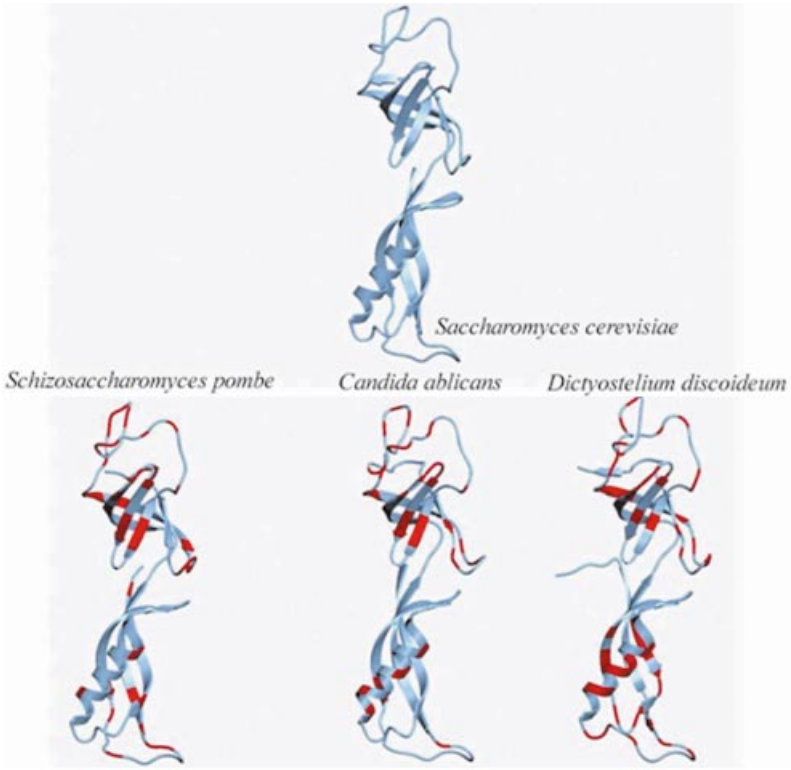

Figure 5. Molecular models of Rpb7 from the lower eukaryotic homologs studied. Residue level variations seen in each homolog when compared with *S.cerevisiae* Rpb7 are shown in red. The figure was generated using SETOR (53).

However, only the *ScN-CaC* chimera can suppress the temperature sensitivity in this background, confirming the functional differences in the domains of these homologs (Fig. 6C).

DISCUSSION

RNA polymerase II is perhaps one of the most crucial molecular assemblies that dictates various fates of the genomic information. Almost all of the subunits of this complex enzyme at the core of transcription show conservation from archaea to mammalian cells. We have chosen to analyze the functional conservation of two subunits, Rpb4 and Rpb7, of this enzyme that have been shown to form a subcomplex in several eukaryotic species (11–13). It has also been shown to be the case in the archaeal enzyme (14). The fact that Rpb7, a core RNA pol II subunit, is essential for cell viability in *S.cerevisiae* suggests that it plays an important role in eukaryotic transcription (38). The requirement of the Rpb4 subunit by the cell under stress conditions like extreme temperatures and stationary phase has been shown by various groups (2,5,47). Apart from this requirement for Rpb4, we have shown that diploid *rpb4* null strains show considerable reduction in sporulation levels and a predisposed state of pseudohyphal growth, evidence of the role played by this subunit under nutritional stress conditions (19). Ability to monitor multiple phenotypes in *S.cerevisiae* allows us to carry out more refined analysis of the structure and role of Rpb4.

These two RNA polymerase subunits appear to have evolved to meet the distinct requirements of the multiple forms of RNA polymerases. *HsRPB7* interacts with the amino terminus of the human transforming factor EWS-Fli and enhances its transcriptional activity (48). In fact even when *hsRPB7* is fused to the Fli1 DNA binding domain it leads to the same effect, thus establishing that it plays an important role in promoter selectivity through protein–protein interaction, while in *S.pombe* and *S.cerevisiae* it was shown recently that the Rpb7 protein interacts with Nrd1 and the corresponding *S.pombe* homolog Seb1 in the respective systems (49). These proteins link the RNA processing activities to the Rpb4/7 subcomplex. Recently, HsRpb7 has been shown to interact with the VHL protein (pVHL), a component of the ubiquitin ligase E3 complex (50). This interaction targets hsRpb7 for ubiquitination and subsequent proteasomic degradation leading to altered gene expression. A protein called Ess1, which modulates CTD structure and plays a role in transcription elongation/termination, has also been genetically shown to interact with Rpb7, while the RNA binding ability of Rpb7 has been implicated in the stability of the transcript emerging from the transcribing polymerase (20).

While yeast RNA pol III has a counterpart of the Rpb4/7 subcomplex made up of C17/C25 proteins, the pol I has the A14/A43 subcomplex. The C17 human homolog, CGRP-RCP, is known to be involved in the signal transduction cascade and is functionally conserved (18). A43, which is related to Rpb7, interacts with Rrn3 which is a general transcription factor required for transcription by RNA pol I (51), while C17, the paralog of Rpb4, is shown to interact with the pol III transcription factor TFIIIB (52). Thus the subcomplex in all the three eukaryotic polymerases and in

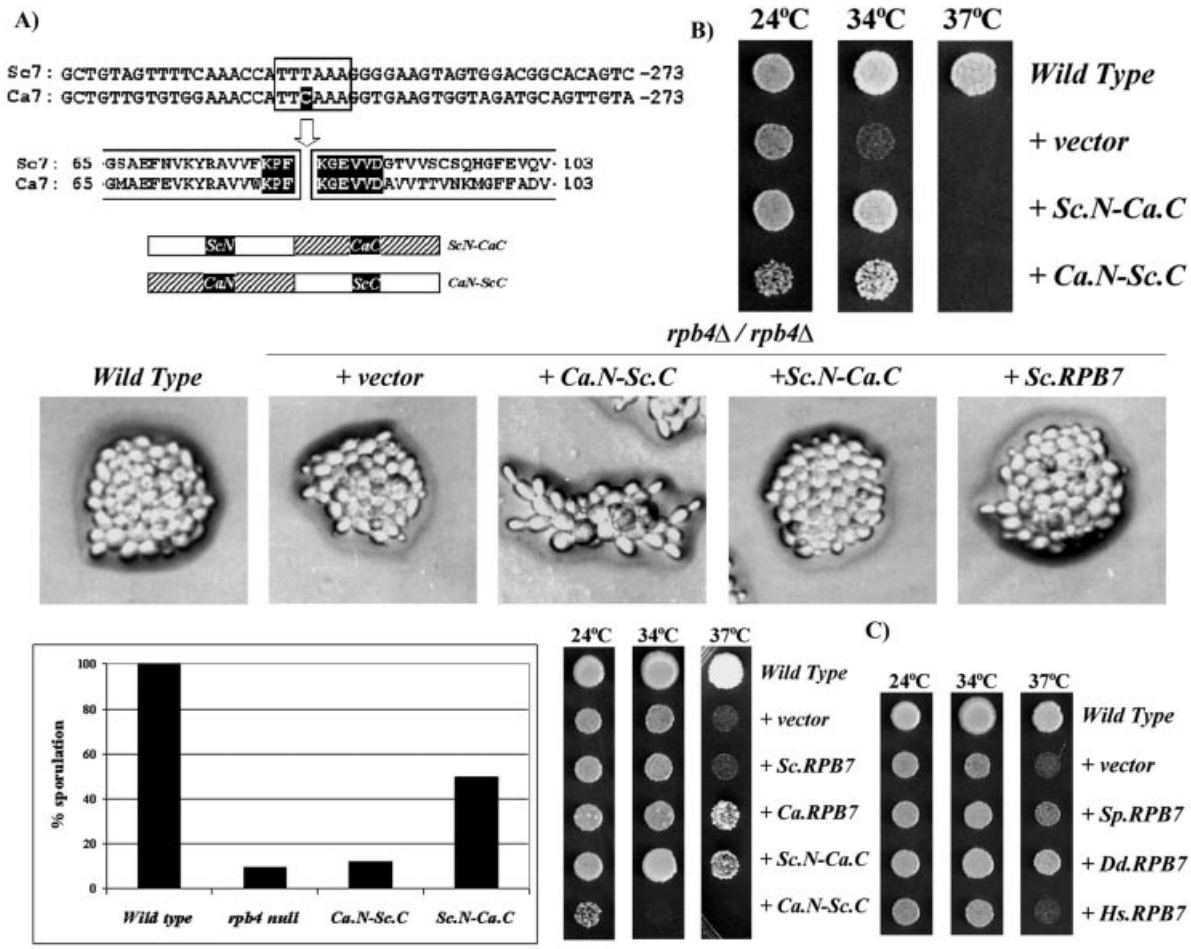

Figure 6. (**A**) Scheme showing the construction of chimeric Rpb7 proteins of *S.cerevisiae* with *C.albicans*. (**B**) Rescue of the phenotypes associated with *rpb4* homozygous diploid null strain by *S.cerevisiae* and *C.albicans* chimera. (**C**) Suppression of Ess1 temperature-sensitive allele $ess1^{H164R}$ at 37°C by *S.cerevisiae* with *C.albicans* chimera (*Sc.N-Ca.C*), *S.pombe* and *D.discoideum* Rpb7 homologs.

transcription in general plays an important role through interactions with proteins outside the respective core RNA polymerases.

We have reported here how the high sequence similarity of *RPB7* from different systems allowed us to test functional conservation between the lower eukaryotic homologs of Rpb7 in *S.cerevisiae*. The homologs from three other lower eukaryotic model systems, namely the pathogenic yeast *C.albicans*, the fission yeast *S.pombe* and the cellular slime mold *D.discoideum*, not only have high sequence similarity but can also functionally complement the absence of Rpb7 in *S.cerevisiae* (Figs 1 and 2).

Previous reports have shown that in *S.cerevisiae* the over-expression of *RPB7* can partially rescue the growth defect associated with the deletion of its interacting partner Rpb4 (9,23,40). Rescue of these defects by over-expression of the homologs from other systems implies conservation of the structural aspects of the protein that are not involved in interaction with Rpb4 but with the other conserved components of the transcription machinery. Our observation that these homologs on over-expression can rescue the growth defect of *RPB4* deletion at 34°C (Fig. 3A) and the defect in sporulation phenotype (Fig. 3C) supports the above speculation.

While we have reported earlier that the human homolog of *RPB7* exaggerates the pseudohyphal morphology of a strain (Σ1278b) predisposed to forming pseudohyphae (6), we have shown here that the over-expression of *S.cerevisiae RPB7* causes exaggeration of the pseudohyphal phenotype of the *rpb4* null homozygous diploid. Among the lower eukaryotic homologs, the *C.albicans* homolog leads to the most impressive enhancement of the pseudohyphal morphology in *RPB4* deletion (Fig. 3B). Taking into account the fact that *C.albicans* shows profound hyphal conversion under many stress conditions as well as the fact that signal transduction pathways leading to pseudohyphal morphology in *S.cerevisiae* and *C.albicans* are well conserved (42) allows us to speculate that some of the interactions involved in regulation of this morphogenesis may be conserved as well and may be mediated by Rpb7 either directly or indirectly.

The solved crystal structure of the subunit E of *M.jannaschii* RNA polymerase shows an N-terminal RNP domain and a C-terminal OB-fold region (15). Since it was observed earlier that the *M.jannaschii* subunit E is the bona

fide homolog of eukaryotic Rpb7 (14), we modeled the structures of *S.cerevisiae* Rpb7 and three other lower eukaryotic homologs based on this structure. The models generated suggest high similarity in the structural features of the eukaryotic variants (Fig. 5). In all the proteins that have the OB fold, the L45 region is known to show large sequence variation and is implicated in RNA binding (43,44). However, it has been observed that the variation does not compromise RNA binding activity of the OB-fold-containing proteins (44). The Rpb7 homologs in this study also reveal that most of the sequence variation is concentrated at the L3α region and the L45 region of the OB fold in these homologs (Fig. 5).

Rpb7 has been shown to bind to single-strand nucleic acid *in vitro* by virtue of the OB-fold region, and deletion of the region 108–113 of the Rpb7 *S.cerevisiae* sequence (which corresponds to β-strand 3 of the OB fold) abrogated any nucleic acid binding and *in vitro* transcriptional activity (8). One of the most conserved regions among the four lower eukaryotic sequences under consideration is indeed the β-strand 3 and L12 region of the protein. That the homologs of Rpb7 complement the absence of Rpb7 in *S.cerevisiae* could be explained based on the fact that the transcriptional activity requires RNA binding property of β-strand 3 and L12. The other surface-exposed conserved regions that map to the Rpb4 binding region and N-terminal RNP domain (Fig. 4) are identified by multiple alignment of a number of Rpb7 sequences from other sources. The high degree of conservation of the latter regions might support an important role such as binding to RNA pol II. In the recently released crystal structures of wild-type RNA pol II complex (36,37) it has been observed that the N-terminal 'tip' region extensively contacts with the base of the clamp of the RNA pol II.

It was obvious from the above structural analysis that the Rpb7 homologs have two distinct domains and sequence comparison suggested that there are short patches of differences between the homologs (Fig. 5) and that maximum variation in the residues is seen at the C-terminus in each case. We employed one more phenotype reported earlier, namely the temperature sensitivity of *ess1* mutant, as a differential phenotype since it was alleviated by the *C.albicans* Rpb7 homolog but not by the *S.cerevisiae* homolog. It was interesting to see that, while the other two lower eukaryotic homologs rescued the phenotype, observations with the chimeras showed that the rescue was dependent on the differences between the two homologs in the C-terminal half. The chimeric proteins further showed that the pseudohyphae and sporulation phenotype could easily be segregated, supporting the notion of functionally independent domains.

We have shown that the structural conservation of the essential subunit of RNA pol II, Rpb7, reflects in its functional conservation. That the lower eukaryotic homologs showed variations mainly in certain restricted positions suggested that these changes are tolerated without significantly disturbing the structure. These subtle 'allowed' variations in the sequence might dictate various interactions of this subunit with other components of the transcriptional regulatory machinery affecting complex phenotypes like pseudohyphae formation and sporulation. Thus our studies now suggest regions of Rpb7 that could be manipulated to understand the finer details of how they mediate these complex interactions and achieve transcriptional regulation.

ACKNOWLEDGEMENTS

P.S. and N.S. thank the members of their respective laboratories for discussions and critical reading of the manuscript. We thank Takahiro Morio for the SSK501 plasmid and *Dictyostelium* cDNA project in Japan with support by Japan Society for the Promotion of Science (RFTF96L00105) and Ministry of Education, Science, Sports and Culture of Japan (08283107). Major work reported here was funded by grants from the Department of Biotechnology, the Department of Science and Technology and the Council for Scientific and Industrial Research India to P.S. N.S. is supported by the award of a Senior Fellowship in Biomedical Sciences by the Wellcome Trust, UK. N.R. is supported by a fellowship from Council of Scientific and Industrial Research, Government of India.

REFERENCES

1. Woychik,N.A. and Hampsey,M. (2002) The RNA polymerase II machinery: structure illuminates function. *Cell*, **108**, 453–463.
2. Woychik,N.A. and Young,R.A. (1989) RNA polymerase II subunit RPB4 is essential for high- and low-temperature yeast cell growth. *Mol. Cell. Biol.*, **9**, 2854–2859.
3. Woychik,N.A., Lane,W.S. and Young,R.A. (1991) Yeast RNA polymerase II subunit RPB9 is essential for growth at temperature extremes. *J. Biol. Chem.*, **266**, 19053–19055.
4. Li,S. and Smerdon,M.J. (2002) Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in *Saccharomyces cerevisiae*. *EMBO J.*, **21**, 5921–5929.
5. Choder,M. and Young,R.A. (1993) A portion of RNA polymerase II molecules has a component essential for stress responses and stress survival. *Mol. Cell. Biol.*, **13**, 6984–6991.
6. Khazak,V., Sadhale,P.P., Woychik,N.A., Brent,R. and Golemis,E.A. (1995) Human RNA polymerase II subunit hsRPB7 functions in yeast and influences stress survival and cell morphology. *Mol. Biol. Cell.*, **6**, 759–775.
7. Edwards,A.M., Kane,C.M., Young,R.A. and Kornberg,R.D. (1991) Two dissociable subunits of yeast RNA polymerase II stimulate the initiation of transcription at a promoter *in vitro*. *J. Biol. Chem.*, **266**, 71–75.
8. Orlicky,S.M., Tran,P.T., Sayre,M.H. and Edwards,A.M. (2001) Dissociable Rpb4-Rpb7 subassembly of RNA polymerase II binds to single-strand nucleic acid and mediates a post recruitment step in transcription initiation. *J. Biol. Chem.*, **276**, 10097–10102.
9. Sheffer,A., Varon,M. and Choder,M. (1999) Rpb7 can interact with RNA polymerase II and support transcription during some stresses independently of Rpb4. *Mol. Cell. Biol.*, **19**, 2672–2680.
10. Bateman,A., Birney,E., Cerruti,L., Durbin,R., Etwiller,L., Eddy,S.R., Griffiths-Jones,S., Howe,K.L., Marshall,M. and Sonnhammer,E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.*, **30**, 276–280.
11. Larkin,R.M. and Guilfoyle,T.J. (1998) Two small subunits in *Arabidopsis* RNA polymerase II are related to yeast RPB4 and RPB7 and interact with one another. *J. Biol. Chem.*, **273**, 5631–5637.
12. Sakurai,H., Mitsuzawa,H., Kimura,M. and Ishihama,A. (1999) The Rpb4 subunit of fission yeast *Schizosaccharomyces pombe* RNA polymerase II is essential for cell viability and similar in structure to the corresponding subunits of higher eukaryotes. *Mol. Cell. Biol.*, **19**, 7511–7518.
13. Khazak,V., Estojak,J., Cho,H., Majors,J., Sonoda,G., Testa,J.R. and Golemis,E.A. (1998) Analysis of the interaction of the novel RNA polymerase II (pol II) subunit hsRPB4 with its partner hsRPB7 and with pol II. *Mol. Cell. Biol.*, **18**, 1935–1945.
14. Werner,F., Eloranta,J.J. and Weinzierl,R.O. (2000) Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res.*, **28**, 4299–4305.
15. Todone,F., Brick,P., Werner,F., Weinzierl,R.O. and Onesti,S. (2001) Structure of an archaeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell.*, **8**, 1137–1143.

16. Sadhale,P.P. and Woychik,N.A. (1994) C25, an essential RNA polymerase III subunit related to the RNA polymerase II subunit RPB7. *Mol. Cell. Biol.*, **14**, 6164–6170.
17. Peyroche,G., Levillain,E., Siaut,M., Callebaut,I., Schultz,P., Sentenac,A., Riva,M. and Carles,C. (2002) The A14–A43 heterodimer subunit in yeast RNA pol I and their relationship to Rpb4–Rpb7 pol II subunits. *Proc. Natl Acad. Sci. USA*, **99**, 14670–14675.
18. Siaut,M., Zaros,C., Levivier,E., Ferri,M.L., Court,M., Werner,M., Callebaut,I., Thuriaux,P., Sentenac,A. and Conesa,C. (2003) An Rpb4/Rpb7-like complex in yeast RNA polymerase III contains the orthologue of mammalian CGRP-RCP. *Mol. Cell. Biol.*, **23**, 195–205.
19. Pillai,B., Verma,J., Abraham,A., Francis,P., Kumar,Y., Tatu,U., Brahmachari,S.K. and Sadhale,P.P. (2003) Whole genome expression profiles of yeast RNA polymerase II core subunit, Rpb4, in stress and nonstress conditions. *J. Biol. Chem.*, **278**, 3339–3346.
20. Wu,X., Wilcox,C.B., Devasahayam,G., Hackett,R.L., Arevalo-Rodriguez,M., Cardenas,M.E., Heitman,J. and Hanes,S.D. (2000) The Ess1 prolyl isomerase is linked to chromatin remodeling complexes and the general transcription machinery. *EMBO J.*, **19**, 3727–3738
21. Sherman,F., Fink,G.R. and Lawrence,C.W. (1983) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
22. Ausubel,F.M., Brent,R., Kingston,R., Moore,D., Seidman,J., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. John Wiley, New York, NY.
23. Sharma,N. and Sadhale,P.P. (1999) Overexpression of the gene for Rpb7 subunit of yeast RNA polymerase II rescues the phenotypes associated with the absence of the largest, nonessential subunit Rpb4. *J. Genet.*, **78**, 149–156.
24. Pillai,B., Sampath,V., Sharma,N. and Sadhale,P. (2001) Rpb4, a non-essential subunit of core RNA polymerase II of *Saccharomyces cerevisiae* is important for activated transcription of a subset of genes. *J. Biol. Chem.*, **276**, 30641–30647.
25. Mumberg,D., Muller,R. and Funk,M. (1995) Yeast vectors for the controlled expression of heterologous proteins in different genetic backgrounds. *Gene*, **156**, 119–122.
26. Thompson,J.D., Higgins,D.G. and Gibson,T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.*, **22**, 4673–4680.
27. Overington,J., Johnson,M.S., Sali,A. and Blundell,T.L. (1990) Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. R. Soc. Lond. B Biol. Sci.*, **241**, 132–145.
28. Lee,B. and Richards,F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379–400.
29. Altschul,S.F., Madden,T.L., Schaffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.
30. Johnson,M.S. and Overington,J.P. (1993) A structural basis for sequence comparisons. An evaluation of scoring methodologies. *J. Mol. Biol.*, **233**, 716–738.
31. Blundell,T., Carney,D., Gardner,S., Hayes,F., Howlin,B., Hubbard,T., Overington,J., Singh,D.A., Sibanda,B.L. and Sutcliffe,M. (1988) 18th Sir Hans Krebs lecture. Knowledge-based protein modelling and design. *Eur. J. Biochem.*, **172**, 513–520.
32. Srinivasan,N. and Blundell,T.L. (1993) An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng.*, **6**, 501–512.
33. Topham,C.M., McLeod,A., Eisenmenger,F., Overington,J.P., Johnson,M.S. and Blundell,T.L. (1993) Fragment ranking in modelling of protein structure. Conformationally constrained environmental amino acid substitution tables. *J. Mol. Biol.*, **229**, 194–220.
34. Sutcliffe,M.J., Hayes,F.R. and Blundell,T.L. (1987) Knowledge based modelling of homologous proteins, Part II: Rules for the conformations of substituted sidechains. *Protein Eng.*, **1**, 385–392.
35. Weiner,S.J., Kollman,P.A., Case,D.A., Singh,U.C., Ghio,C., Alagona,G., Profeta,S. and Weiner,P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.*, **106**, 765–784.
36. Armache,K.J., Kettenberger,H. and Cramer,P. (2003) Architecture of initiation-competent 12-subunit RNA polymerase II. *Proc. Natl Acad. Sci. USA*, **100**, 6964–6968.
37. Bushnell,D.A. and Kornberg,R.D. (2003) Complete, 12-subunit RNA polymerase II at 4.1-Å resolution: implications for the initiation of transcription. *Proc. Natl Acad. Sci. USA*, **100**, 6969–6973.
38. McKune,K., Richards,K.L., Edwards,A.M., Young,R.A. and Woychik,N.A. (1993) RPB7, one of two dissociable subunits of yeast RNA polymerase II, is essential for cell viability. *Yeast*, **9**, 295–299.
39. Rosenheck,S. and Choder,M. (1998) Rpb4, a subunit of RNA polymerase II, enables the enzyme to transcribe at temperature extremes *in vitro*. *J. Bacteriol.*, **180**, 6187–6192.
40. Maillet,I., Buhler,J.M., Sentenac,A. and Labarre,J. (1999) Rpb4p is necessary for RNA polymerase II activity at high temperature. *J. Biol. Chem.*, **274**, 22586–22590.
41. Gimeno,C.J., Ljungdahl,P.O., Styles,C.A. and Fink,G.R. (1992) Unipolar cell divisions in the yeast *S.cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell*, **68**, 1077–1090.
42. Liu,H. (2001) Transcriptional control of dimorphism in *Candida albicans*. *Curr. Opin. Microbiol.*, **4**, 728–735.
43. Murzin,A.G. (1993) OB (oligonucleotide/oligosaccharide binding) fold: common structural and functional solution for non-homologous sequences. *EMBO J.*, **12**, 861–867.
44. Arcus,V. (2002) OB-fold domains: a snapshot of the evolution of sequence, structure and function. *Curr. Opin. Struct. Biol.*, **12**, 794–801.
45. Bycroft,M., Hubbard,T.J., Proctor,M., Freund,S.M. and Murzin,A.G. (1997) The solution structure of the S1 RNA binding domain: a member of an ancient nucleic acid-binding fold. *Cell*, **88**, 235–242.
46. Devasahayam,G., Chaturvedi,V. and Hanes,S.D. (2002) The Ess1 prolyl isomerase is required for growth and morphogenetic switching in *Candida albicans*. *Genetics*, **160**, 37–48.
47. Choder,M. (1993) A growth rate-limiting process in the last growth phase of the yeast life cycle involves RPB4, a subunit of RNA polymerase II. *J. Bacteriol.*, **175**, 6358–6363.
48. Petermann,R., Mossier,B.M., Aryee,D.N., Khazak,V., Golemis,E.A. and Kovar,H. (1998) Oncogenic EWS-Fli1 interacts with hsRPB7, a subunit of human RNA polymerase II. *Oncogene*, **17**, 603–610.
49. Mitsuzawa,H., Kanda,E. and Ishihama,A. (2003) Rpb7 subunit of RNA polymerase II interacts with an RNA-binding protein involved in processing of transcripts. *Nucleic Acids Res.*, **31**, 4696–4701.
50. Na,X., Duan,H.O., Messing,E.M., Schoen,S.R., Ryan,C.K., di Sant'Agnese,P.A., Golemis,E.A. and Wu,G. (2003) Identification of the RNA polymerase II subunit hsRPB7 as a novel target of the von Hippel–Lindau protein. *EMBO J.*, **22**, 4249–4259.
51. Peyroche,G., Milkereit,P., Bischler,N., Tschochner,H., Schultz,P., Sentenac,A., Carles,C. and Riva,M. (2000) The recruitment of RNA polymerase I on rDNA is mediated by the interaction of the A43 subunit with Rrn3. *EMBO J.*, **19**, 5473–5482.
52. Ferri,M.L., Peyroche,G., Siaut,M., Lefebvre,O., Carles,C., Conesa,C. and Sentenac,A. (2000) A novel subunit of yeast RNA polymerase III interacts with the TFIIB-related domain of TFIIIB70. *Mol. Cell. Biol.*, **20**, 488–495.
53. Evans,S.V. (1993) SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.*, **11**, 127–128, 134–138.