Crystal structure and RNA binding of the Rpb4/Rpb7 subunits of human RNA polymerase II

Hedije Meka, Finn Werner, Suzanne C. Cordell, Silvia Onesti and Peter Brick*

Department of Biological Sciences, Imperial College, London SW7 2BW, UK

ABSTRACT

The Rpb4 and Rpb7 subunits of eukaryotic RNA polymerase II (RNAP$_{II}$) form a heterodimer that protrudes from the 10-subunit core of the enzyme. We have obtained crystals of the human Rpb4/Rpb7 heterodimer and determined the structure to 2.7 Å resolution. The presence of putative RNA-binding domains on the Rpb7 subunit and the position of the heterodimer close to the RNA exit groove in the 12 subunit yeast polymerase complex strongly suggests a role for the heterodimer in binding and stabilizing the nascent RNA transcript. We have complemented the structural analysis with biochemical studies directed at dissecting the RNA-binding properties of the human Rpb4/Rpb7 complex and that of the homologous E/F complex from *Methanocaldococcus jannaschii*. A number of conserved, solvent-exposed residues in both the human Rpb7 subunit and the archaeal E subunit have been modified by site-directed mutagenesis and the mutants tested for RNA binding by performing electrophoretic mobility shift assays. These studies have identified an elongated surface region on the corresponding face of both subunit E and Rpb7 that is involved in RNA binding. The area spans the nucleic acid binding face of the OB fold, including the B4–B5 loop, but also extends towards the N-terminal domain.

INTRODUCTION

In the eukaryotic RNA polymerase II (RNAP$_{II}$), two subunits (Rpb4/Rpb7) form a heterodimer that associate with the 10-subunit core of the enzyme (1). The stoichiometry of RNAP$_{II}$ purified from *Saccharomyces cerevisiae* cells is dependent on growth conditions: in optimally growing cells only ∼20% of purified material includes the Rpb4/Rpb7 heterodimer, while in the post-log phase virtually all the polymerase molecules

contain the 12 subunits. RNAP$_{II}$ purified from *rpb4Δ* cells not only lacks the corresponding subunit Rpb4, but also contains no detectable Rpb7 (2). Whereas in *S.cerevisiae RPB7* is an essential gene, *RPB4* is dispensable under optimal growth conditions, but becomes essential during heat or cold shock, and under nutrient depletion. Overexpression of Rpb7 partly compensates for the *rpb4Δ* stress phenotypes and Rpb7 can interact with RNAP$_{II}$ independently of Rpb4, although this interaction is not very stable and can only be detected when the subunit is overexpressed (3,4). In the *S.cerevisiae* system, Rpb4/Rpb7 is necessary for promoter-directed transcription initiation but dispensable for transcription elongation and termination *in vitro* (2) and activated transcription *in vivo* (5). Rpb4 also recruits the CTD phosphatase Fcp1 (6). Both Rpb4 and Rpb7 have been shown to have additional specific functions in gene expression, such as regulating transcription-coupled repair through the Rad26-dependent pathway (7), in efficient mRNA export (8) and in transcription termination (9). The heterodimer seems therefore to play a role in the intricate network linking transcription, RNA processing and mRNA export.

Archaea contain a single RNAP enzyme with subunits which show considerable sequence homology to the subunits of the eukaryotic RNAP$_{II}$. The sequences of the Rpb7 subunits from eukaryotes and the homologous E subunits from archaea are well conserved while the sequence similarity between the eukaryotic Rpb4 subunits and the archaeal F subunits is very much lower (Figure 1). The archaeal RNAP is fully capable of promoter-directed transcription without subunits E/F (10). It is interesting to note that the stimulatory effect of the archaeal basal transcription factor TFE (homologous to eukaryotic TFIIEα) is dependent on the RNAP subunits E/F, also suggesting a role of E/F for transcription initiation in the archaea (11).

Since the heterogeneity in the yeast polymerase preparations caused by the variable stoichiometry of the Rpb4/Rpb7 heterodimer interfered with crystallization, crystals of RNAP$_{II}$ diffracting to high resolution were originally obtained for the 10-subunit core (12). However, the 3D structure of the complex between the archaeal homologues of E and F had been

*To whom correspondence should be addressed. Tel: +44 20 75947704; Fax: +44 20 75890191; Email: p.brick@imperial.ac.uk
Correspondence may also be addressed to Silvia Onesti. Tel: +44 20 75947704; Fax: +44 20 75890191; Email: s.onesti@imperial.ac.uk
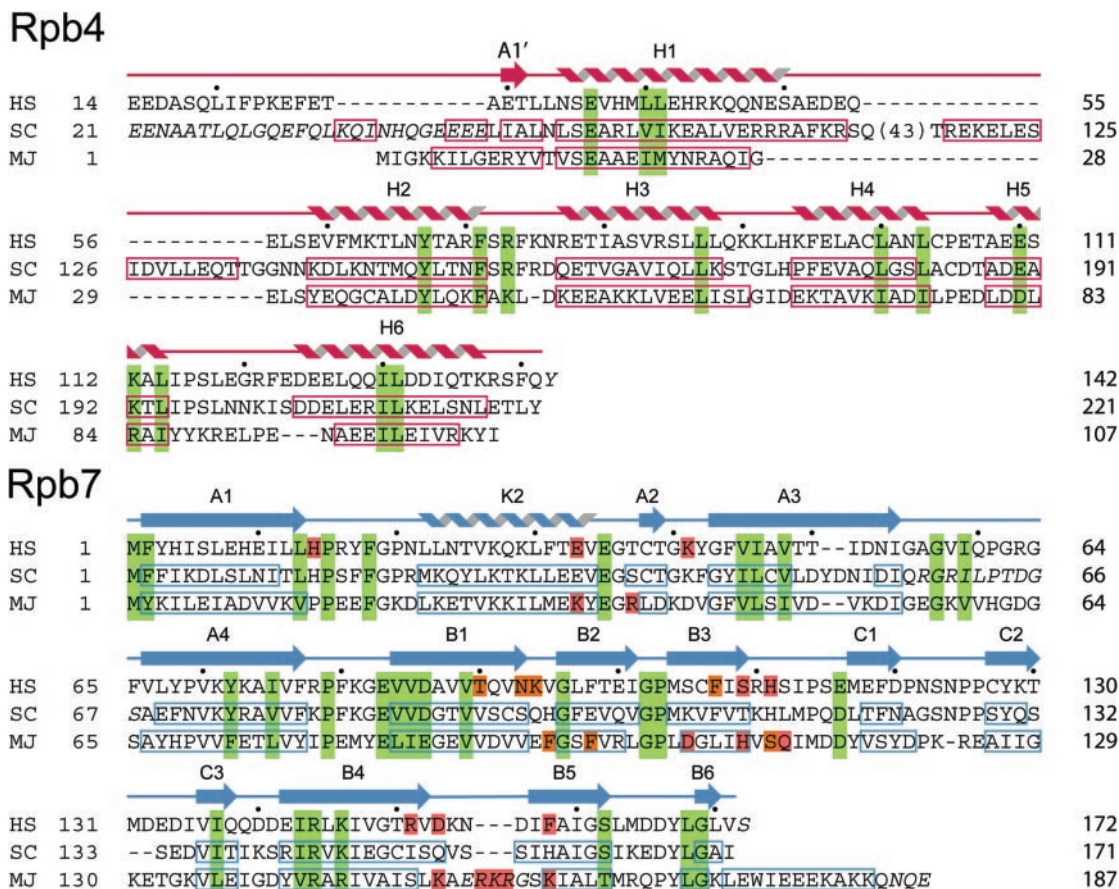
Figure 1. Structure based sequence alignment of the Rpb4 and Rpb7 subunits. Sequences of the human (HS), yeast (SC) and *M.jannaschii* (MJ) orthologues of Rpb4 and Rpb7 are included. The secondary structure of the human protein determined using DSSP (33) is indicated above the sequences while secondary structural elements in the yeast (PDB entry 1y14) and *M.jannaschii* (PDB entry 1go3) proteins are indicated by boxes. Amino acids not present in the high-resolution crystal structures are indicated by italics. Residues which are well conserved in archaeal and eukaryotic sequences are highlighted using a green background. In the Rpb7 alignment, a red background indicates residues mutagenized in this study which have a strong effect on the RNA-binding properties of the protein whereas an orange background indicates residues whose mutation has a weaker effect on RNA binding. Residues in strands B1–B6 of Rpb7 share sequence similarity with proteins containing an S1 motif and form a β-barrel structure called an OB fold.

determined by crystallography at 1.75 Å resolution (13). The crystal structure of the 12-subunit yeast RNAP$_{II}$ complex was initially determined at 4.2 Å resolution (14,15). As the low-resolution map did not allow an independent atomic model to be built for the eukaryotic Rpb4/Rpb7 complex, the structure of the archaeal E/F dimer was simply placed into the electron density. More recently the crystal structure of the *S.cerevisiae* Rpb4/Rpb7 complex has been determined (16), which has allowed the 12-subunit RNAP$_{II}$ complex to be refined to 3.8 Å resolution.

The archaeal subunit E (the Rpb7 homologue) has an elongated two-domain structure, containing two potential single-stranded RNA (ssRNA)-binding motifs: a truncated RNP motif at the N-terminus and a S1 motif at the C-terminus (13). The presence of an S1 motif in the sequence and structure suggests that its role is to bind single-stranded nucleic acids. Indeed, the eukaryotic Rpb4/Rpb7 complex has been shown to bind to both single-stranded DNA (ssDNA) and ssRNA with comparable affinities (17). Although proteins containing an OB fold can bind either ssRNA or ssDNA, the particular subgroup containing the sequence signature of the S1 motif bind ssRNA without sequence specificity.

We therefore proposed a model for the function of the heterodimer, in which the S1 motif of Rpb7 interacts with the nascent RNA transcript, possibly assisted by the truncated RNP domain (13).

We have now determined the crystal structure of the human Rpb4/Rpb7 complex. To gain insight into the contributions of different regions of the heterodimer to RNA binding, we have carried out structure-based mutational analysis of conserved, surface-exposed amino acid residues in both the archaeal E/F and the human Rpb4/Rpb7 protein complex and tested the ability of each mutant protein to bind RNA. We show that residues within an elongated patch on the surface of human Rpb7 contribute to RNA binding. The RNA-binding region includes residues forming the OB fold in the C-terminal domain of the protein as well as additional residues from the N-terminal domain. In the context of the yeast 12-subunit RNAP$_{II}$ complex, the RNA-binding region of human Rpb4/Rpb7 corresponds to the face of the heterodimer adjacent to the 10-subunit core. We also show that, in spite of relatively weak sequence identity, residues within a surface patch in a corresponding position on the surface of subunit E of the archaeal polymerase also contribute to RNA binding.

MATERIALS AND METHODS

Expression, purification and crystallization of human Rpb4/Rpb7

Co-expression in *Escherichia coli* of the human Rpb4/Rpb7 complex was achieved by using a bicistronic construct based upon pGEX-2TK (Amersham Biosciences) in which the Rpb4 subunit was fused to a glutathione-*S*-transferase tag that was cleavable by thrombin, whereas the Rpb7 subunit was expressed as an untagged protein (18). *E.coli* BL21(DE3) cells were transformed under the control of ampicillin and, after induction at OD_{600} between 0.6 and 0.9, cell cultures were grown at 28°C overnight. The cells were harvested by centrifugation and resuspended in 20 mM Tris–HCl, pH 7.9, 300 mM potassium acetate, 7 mM magnesium acetate, 1 mM EDTA, 10 mM DTT, 10% glycerol (buffer P300), supplemented with lysozyme at 1 mg/ml, Benzonase (0.1U/ml) and 1 mM 4-(2-aminoethyl)benzenesulphonyl fluoride. Cell lysis was performed by sonication and the resulting lysate clarified by centrifugation and filtration. All the chromatographic steps were performed at 4°C. The crude extract was bound in batch to glutathione sepharose beads (Amersham Biosciences) and the beads were extensively washed with buffer P300. The protein was recovered from the beads by cleavage with thrombin overnight at 4°C. In preparation for anion exchange chromatography, the supernatant was dialyzed against buffer A [20 mM Tris–HCl, pH 8.0, 1 mM EDTA, 10% glycerol and 10 mM β-mercaptoethanol] containing 100 mM NaCl. The dialyzed protein solution was loaded onto a Resource Q column (Amersham Biosciences) using an ÄKTA FPLC apparatus (Amersham Biosciences) and eluted with a linear gradient from 100 mM to 1 M NaCl in buffer A. After concentration, the sample was applied to a Superdex-200 10/30 gel filtration column (Amersham Biosciences) equilibrated with 20 mM MES, pH 6.5 and 100 mM NaCl. Pooled peak fractions were flash frozen in liquid nitrogen and stored at −80°C until further use.

For crystallization the protein was concentrated to 15 mg/ml in 20 mM PIPES, pH 6.8, 100 mM NaCl and 15% glycerol. Crystals were grown at 18°C by vapour diffusion in hanging drops, against a well solution containing 100 mM PIPES pH 7.4, 25% polyethylene glycol (MW 4000) and 15% glycerol. Most of the crystals grew as clusters of thin plates, but occasionally a few small chunky crystals could be obtained.

Structure determination and refinement

The crystals were flash-frozen directly from the drop and X-ray data collected at 100 K using a MAR image plate detector mounted on a Rigaku RU-H3R rotating anode X-ray generator equipped with OSMIC focusing mirrors (CuKα radiation; λ = 1.54 Å). The diffraction data were processed with MOSFLM and programs of the CCP4 suite (19). The crystals are in space group P1 with four heterodimers in the asymmetric unit (Table 1). A search model for molecular replacement was constructed from the yeast Rpb4/Rpb7 complex (29% sequence identity for Rpb4, 37% sequence identity for Rpb7) (16). The positions of the four copies of the heterodimer in the asymmetric unit were determined with the program PHASER (20), and alternating cycles of refinement with

Table 1. Data collection and refinement statistics

Space group	P1
Cell parameters (Å)	a = 45.69, b = 83.19, c = 99.31, α = 85.53°, β = 81.80°, γ = 81.77°
Resolution (Å)	43.3–2.7 (2.85–2.7)
R_{merge} (%)	6.9 (39.4)
Multiplicity	3.4 (3.4)
Completeness (%)	95.3 (93.8)
I/σ	14.5 (3.4)
R_{work} (%)	23.1
R_{free} (%)	26.6
R.m.s.d. bonds (Å)	0.008
R.m.s.d. angles (degrees)	1.4
Number of protein atoms	9236
Number of water molecules	54

Highest resolution shell values shown in parentheses.

CNS (21) and model building with O (22) resulted in the final model. The four copies of the Rpb7 subunit make different crystal contacts and adopt two slightly different conformations. The largest differences are located in the A1–K2 and the A3–A4 loops (Figures 1 and 2) with residues 60–63 differing in position by >5 Å. Tight restraints between pairs of heterodimers were applied during the refinement procedure. The refined model has 98.9% of the residues in allowed or additionally allowed regions of the Ramachandran plot and no residues in disallowed regions. Residue 172 of Rpb7 and residues 1–13 and 142 of Rpb4 have no electron density and are presumed disordered. Crystallographic statistics are summarized in Table 1. The coordinates have been deposited in the Protein Data Bank (entry 2C35).

Expression and purification of the wild-type and mutant protein for RNA binding

The wild-type human Rpb4/Rpb7 complex was expressed and purified as described above. For expression of the *Methanocaldococcus jannaschii* heterodimer, subunits E and F were subcloned generating a bicistronic expression construct in pET21a(+) (Novagen) and co-expressed as untagged proteins in BL21(DE3) cells by inducing exponentially growing cells with 0.5 mM Isopropyl-β-D-thiogalactopyranoside for 3–5 h at 37°C. Cells were lysed in P300 as described above and the lysate held at 75°C for 40 min. The thermally denatured *E.coli* proteins were removed by centrifugation and the supernatant incubated with DEAE-sepharose in order to remove nucleic acids. The resulting lysate preparation was loaded onto a HiLoad Superdex-75 26/60 gel filtration column (Amersham Biosciences) and equilibrated with 20 mM Tris–HCl, pH 8.0 and 100 mM NaCl. Pooled peak fractions were concentrated to between 10 and 15 mg/ml, flash frozen in liquid nitrogen and stored at −80°C.

Site-directed mutagenesis was performed using the QuikChange (Stratagene) mutagenesis kit according to the manufacturer's instructions. After confirmation of each mutation by sequencing, the *M.jannaschii* and human complexes were expressed in *E.coli* as described above and purified following essentially the same strategy as used for the corresponding wild-type proteins, with the exception that the anion exchange chromatography step was omitted in the purification of the human Rpb4/7 mutants.
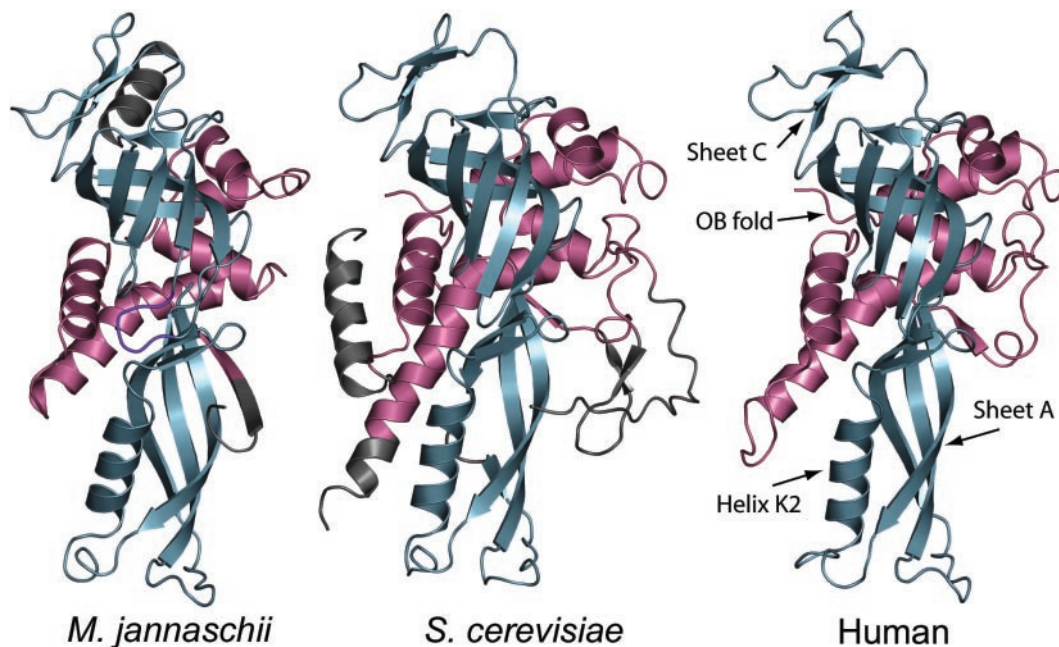
Figure 2. Structures of the *M.jannaschii* E/F complex (PDB entry 1go3), yeast Rpb4/Rpb7 complex (PDB entry 1wcm) and human Rpb4/Rpb7 complex. The complexes are viewed from the side of the heterodimer which faces the RNA exit groove of the RNAP$_{II}$ core. The structurally conserved regions of the Rpb4 subunits are coloured pink while the structurally conserved regions of the Rpb7 subunits are coloured violet. In the *M.jannaschii* crystal structure no electron density is visible for the loop connecting strands B4 and B5; this loop has been drawn for the sake of clarity and is coloured in dark blue. Regions of the yeast and archaeal structures that differ from the human protein are shown in grey. The yeast structure is from the 12-subunit RNAP$_{II}$ complex obtained at 3.8 Å (PDB entry 1wcm) rather than the higher resolution but less complete structure of the isolated heterodimer. This figure and all others representing molecular structures were generated with PyMOL (http://pymol.sourceforge.net).

RNA binding

The probe template was a gift from S. Curry (Imperial College) and directs transcription in pGEM4Z (Promega) of a 54-ribonucleotide RNA species. The RNA probe was produced by *in vitro* transcription using T7 RNA polymerase and ^{32}P-UTP according to standard procedures. Following transcription, the template was digested with 1 U of RNase-free DNase I and unincorporated nucleotides removed using a MicroSpin G-25 desalting column (Amersham Biosciences); the RNA was recovered by precipitation with 1 vol of 7.5 M NH$_4$OAc and 4 vol of ethanol. Probe at a concentration of 2500 c.p.m. was used in a 10 µl reaction containing 10 mM MOPS, pH 6.8, 100 mM NaCl, 10% glycerol, 40 U of RNasin (Promega) and varying amounts of protein. The samples were incubated at room temperature for 15 min before loading onto a native 4–20% acrylamide (Novex), 0.5× TBE gel. The complexes were separated electrophoretically on 4–20% TBE gels (Invitrogen) at 100 V for 2.5 h at 4°C and quantified using a phosphorimager (Fuji).

RESULTS

Crystal structure of the human Rpb4/Rpb7 heterodimer

The human Rpb4/Rpb7 complex crystallizes with four copies of the heterodimer in the crystallographic asymmetric unit. The structure was solved at 2.7 Å resolution by molecular replacement using the yeast Rpb4/Rpb7 complex as a search model (16) and refined to a free *R*-factor of 26.6% (working *R*-factor of 23.1%) with good geometry (for more statistics see Table 1).

The Rpb7 polypeptide folds into two separate domains forming an elongated structure (Figure 2, blue). The six helices of Rpb4 (Figure 2, pink) pack around the central region of Rpb7 at the interface between the two domains. The N-terminal domain of Rpb7 (lower blue domain in Figure 2) folds into a four-stranded anti-parallel β-sheet (strands A1–A4) that wraps around an α-helix (K2). An additional fifth strand of this β-sheet (strand A1′) is formed from residues near the N-terminus of Rpb4. The C-terminal domain of Rpb7 folds into a β-barrel structure known as an OB fold (strands B1–B6) which is capped by a small 3-stranded antiparallel β-sheet (strands C1–C3).

As expected, the structure of the human Rpb4/Rpb7 heterodimer is similar to the yeast and *M.jannaschii* orthologues (Figure 2). Optimally superimposition of the human Rpb4/Rpb7 heterodimer onto the yeast heterodimer in the 12-subunit RNAP$_{II}$ complex obtained at 3.8 Å (PDB entry 1wcm) results in 282 equivalent Cα atoms that are within 4 Å (r.m.s. separation of 0.66 Å). A similar superposition onto the higher resolution (2.3 Å) but less complete crystal structure of the isolated yeast heterodimer (PDB entry 1y14) results in 249 Cα atoms within 4 Å (rmsd 0.61 Å). Superposition onto the archaeal heterodimer (PDB entry 1go3) results in 252 equivalent Cα atoms that are within 4 Å (rmsd of 0.99 Å).

The archaeal E/F and the human Rpb4/Rpb7 heterodimers bind RNA with comparable affinities

To test for ssRNA-binding activity, we performed electrophoretic mobility shift assays (EMSA). The archaeal E/F heterodimer was incubated with a radiolabelled RNA probe
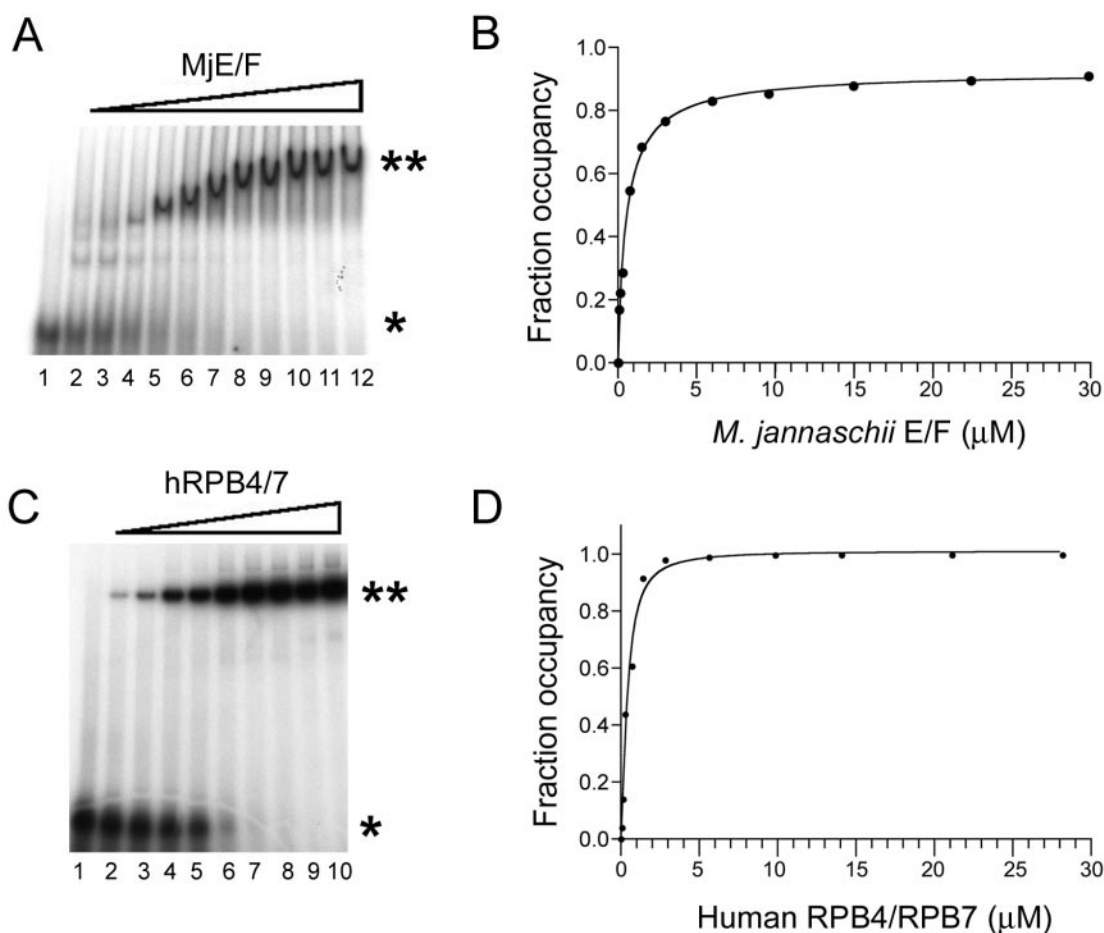
Figure 3. EMSAs using RNAP subunits E/F from *M.jannaschii* and human Rpb4/Rpb7. Archaeal E/F and human Rpb4/7 give rise to complexes with the RNA probe in a protein-concentration dependent manner. Asterisks indicate the position of the free probe, while double asterisks indicate the position of the protein–RNA complexes. The protein concentrations were 25 ng, 50 ng, 100 ng, 250 ng, 500 ng, 1 µg, 2 µg, 3.5 µg, 5 µg, 7.5 µg and 10 µg. (**A**) Lane 1, free probe; lanes 2–12: increasing concentration of *M.jannaschii* E/F complex. (**B**) Quantification of the EMSA experiment for the *M.jannaschii* complex has been carried out by plotting the fractional occupancy of the probe bound to the protein, corrected for the background, against the protein concentration (**C**) Lane 1, free probe; lanes 2–10: increasing concentration of human Rpb4/Rpb7 complex. (**D**) Quantification of the EMSA experiment for the human complex has been carried out as described in (B).

using increasing protein concentrations (25 ng to 10 µg) and the mixture analysed by PAGE (Figure 3A) to monitor the formation of protein–RNA complexes. A quantitative analysis of the RNA binding gave a value of 540 nM for the dissociation constant (Figure 3B). A similar procedure was carried out with the recombinant human Rpb4/Rpb7 heterodimer and gave an RNA dissociation constant of 420 nM (Figure 3C and D). Protein–RNA complexes with comparable affinities were obtained using various RNA probes (data not shown) indicating that the binding is not sequence specific.

Dissecting the RNA-binding region in the archaeal E/F complex

To identify the RNA-binding surface of the *M.jannaschii* E/F complex we mutated residues of the E subunit, tested the RNA-binding properties of the mutant protein by a gel-shift assay and compared the binding with the wild-type protein. Targets for the mutagenesis were selected based on their surface-exposure (in the context of RNAP$_{II}$), amino acid conservation (in the archaeal domain) and chemical nature.

Because we expected a reasonably large interaction area between protein and nucleic acid, we reasoned that substitution of one of the contact residues to alanine was probable to have a minor effect on RNA binding owing to the large number of remaining interactions stabilizing the complex. We therefore opted for charge-reversal substitutions which are more probable to disrupt the interaction.

The atomic model of the E subunit shows the presence of two distinct domains forming an elongated structure (Figure 2). The sequence of the C-terminal domain (Figure 2, upper blue domain) indicated the presence of a S1 motif (Figure 1, strands B1–B6). The S1 motif folds into the β-barrel structure which characterizes OB folds. A cluster of conserved surface-exposed amino acids is located on the exposed face of the β-barrel, and includes charged or polar residues (Asp105, His109, Ser111, Gln112, Lys152 and Lys160) and aromatic residues (Phe95 and Phe98) (Figure 4). To determine whether any of these surface residues were functionally important, each of them was replaced by either a reverse-charge mutation or by alanine. In some OB folds, the loop connecting strand B4 and B5 has been shown to play a central role in DNA or RNA
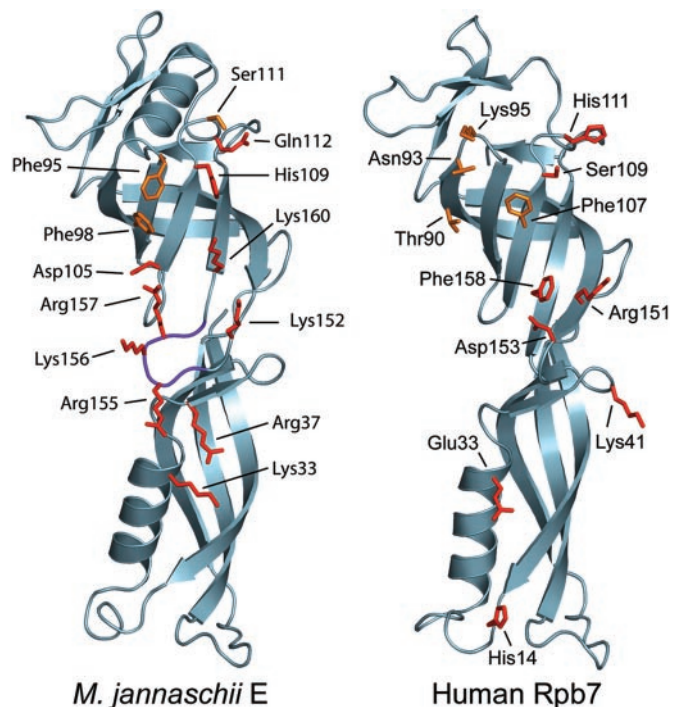
Figure 4. Ribbon diagrams showing the position of the residues mutated in this study on the crystal structures of *M.jannaschii* subunit E and human Rpb7. The subunits are viewed in the same orientation as that used in Figure 2. The amino acid side chains whose mutations have a large effect on the RNA-binding properties of the complex are shown in red, while the amino acids that have a weaker effect are shown in orange.

binding by closing over the nucleic acids. The sequence of the loop in *M.jannaschii* includes four positively charged residues and is disordered in the crystal structure. To examine the role of the B4–B5 loop in nucleic acid binding, we have created a 'loop mutant' in which Arg155, Lys156 and Arg157 were mutate to alanine, serine and alanine, respectively. Based on the same criteria of surface exposure and sequence conservation, a further set of residues were identified in the N-terminal domain of the E subunit, which folds into a truncated RNP domain. These comprised Lys33, Arg37, Lys40 and Asp41.

Figure 4 shows the position of the mutated residues on the crystal structure of the *M.jannaschii* E/F complex. The expression levels and purity of the mutants were comparable with that of the wild-type protein. Moreover, all mutants were thermostable (75°C) and behaved in a similar manner to the wild-type heterodimer on gel filtration, demonstrating that the structural integrity of the mutant E/F complexes were not compromised by the amino-acid substitutions. The two exceptions were the mutants involving Lys40 and Asp41, which could not be purified to the same standards and were therefore not tested for binding. The RNA-binding properties of the mutants were then compared with those of the wild-type protein. Examples of RNA-binding experiments for a selection of mutants are reported in Figure 5A and B, while all the results are summarized in Figure 5C.

While mutation of Phe95, Phe98 and Ser111 only mildly affect the interaction with the nucleic acids, most of the mutants showed a decrease in RNA binding when compared

with the wild-type protein. The loop mutant, as well as mutation of a number of positively charged residues (Lys33, Arg37, Lys152 and Lys160), virtually abolish RNA binding. When the residues whose mutation has have a strong effect in the interaction are displayed on the crystal structure of the *M.jannaschii* E/F heterodimer (Figure 4), a putative RNA-binding path emerges involving both the N-terminal and C-terminal domains.

Dissecting the RNA-binding region in the human Rpb4/Rpb7 complex

We applied a similar approach to study the RNA-binding properties of the human Rpb4/Rpb7 complex. The sequence identity between subunit E and Rpb7 is only ∼21% of the amino acids that are conserved in the S1 motif of the archaeal homologue, only Phe 98 is also well conserved in the eukaryotic homologues. However, a similar pattern of exposed, conserved residues in the C-terminal domain of the eukaryotic Rpb7 subunits can be inferred. These residues include Thr90, Asn93, Lys94, Phe98, Phe107, Ser109, His111, Arg151, Asp153 and Phe158. Another set of conserved, surface-exposed residues can be found in the N-terminal domain, comprising His14, Glu33 and Lys41. These residues were mutated to either alanine, lysine or glutamic acid and are shown on Figure 4.

As was the case with the E/F mutant complexes described above, the expression levels and purity of the human Rpb4/Rpb7 mutants were comparable with the wild-type protein. Unfortunately we have been unable to express any protein with a mutation at position 98; attempts to change the phenylalanine residue to either alanine or glutamic acid seriously affected the solubility of the protein. The RNA-binding properties of the mutants were then compared with those of the wild-type protein. Examples of RNA-binding experiments for a selection of mutants are reported in Figure 5D and E, while all the results are summarized in Figure 5F.

As in the archaeal protein, residues can be classified into two categories, based on the impact of the mutation on the RNA-binding properties. Thr90, Asn93, Lys94 and Phe107 decrease the binding by about a factor of two, while the other residues have a stronger impact on the interaction. When the residues which have a strong effect are displayed on the crystal structure of the human Rpb4/Rpb7 heterodimer, a pattern similar to that observed in the archaeal system emerges, with the putative RNA-binding path spanning both domains of the Rpb7 subunit.

DISCUSSION

In the yeast 12-subunit RNAP$_{II}$ complex, the Rpb4/Rpb7 heterodimer is positioned with the Rpb7 subunit orientated towards the 10-subunit core so that the Rpb4 helices are on the outside of the complex (Figure 6) (16). The large insertion between helices H1 and H2 in yeast Rpb4 is therefore positioned on the outer face of the polymerase molecule, well away from the core subunits. However, the very variable N-terminal segment of Rpb4 is in close proximity to the polymerase core. The 13 N-terminal amino acids of human Rpb4 are not conserved in eukaryotes and are disordered in the crystal structure. In contrast, the solvent exposed Glu14 and Glu15 in
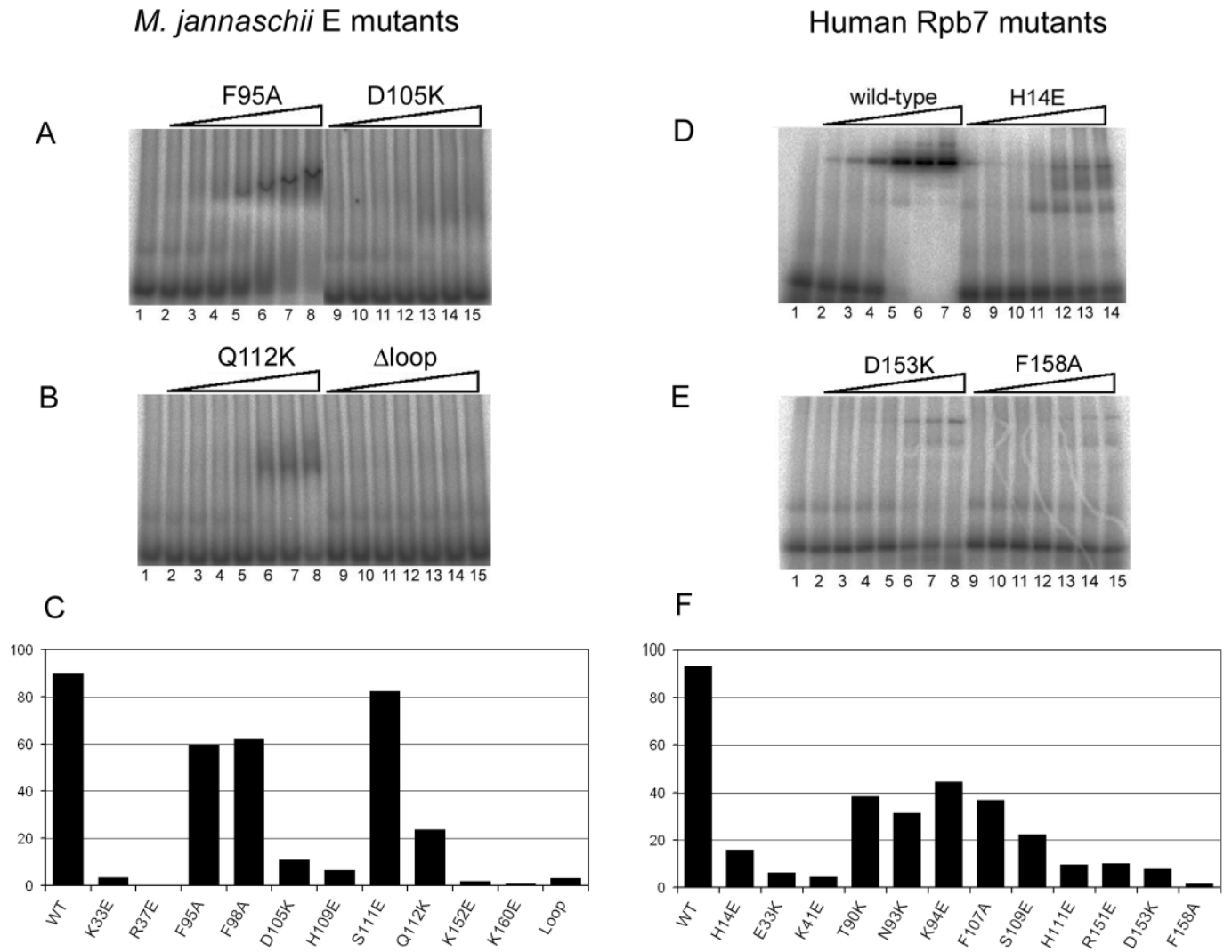
Figure 5. Results of EMSAs performed using a radiolabelled RNA probe. (**A**) Example of a mobility assay using the *M.jannaschii* E/F heterodimer. Lane 1, free probe; lanes 2–8, increasing concentrations of the E/F complex with a F95A mutation in the E subunit; lanes 10–15, increasing concentrations of the E/F complex with a D105K mutation in the E subunit. (**B**) Example of a mobility assay using the *M.jannaschii* E/F heterodimer. Lane 1, free probe; lanes 2-8, increasing concentrations of the E/F complex with a Q112K mutation in the E subunit; lanes 10–15, increasing concentrations of the E/F complex with the Δloop mutation in the E subunit. (**C**) Summary of the effect of mutations in the *M.jannaschii* E subunit on the RNA-binding properties. For each mutant protein, the percentage of RNA probe bound to the protein, after background correction, is shown. (**D**) Example of a mobility assay performed using the human Rpb4/Rpb7 heterodimer. Lane 1, free probe; lanes 2–7, increasing concentrations of wild-type Rpb4/Rpb7 (ranging from 125 ng to 10 µg); lanes 9–14, increasing concentrations of Rpb4/Rpb7 with a H14E mutation in Rpb7. (**E**) Example of a mobility assay performed using the human Rpb4/Rpb7 heterodimer. Lane 1, free probe; lanes 2-8, increasing concentrations of Rpb4/Rpb7 with a D153K mutation in Rpb7; lanes 10–15, increasing concentrations of Rpb4/Rpb7 with a F158A mutation in Rpb7. (**F**) Summary of the effect of mutations in the human Rpb7 subunit on the RNA-binding properties, where experimental data were treated as described for the *M.jannaschii* E subunit.

the human structure (residues 21 and 22 in yeast, Figure 1) are well conserved and residues 14–28 at the N-terminus of the human Rpb4 overlap with the corresponding residues (21–35) of the yeast structure. The conformation of this loop is probable to be conserved in eukaryotes as it is stabilized by the interactions of conserved hydrophobic side chains: Phe23 and Phe26 of Rpb4 pack against Phe95 of Rpb4 and Phe80 of Rpb7. In the crystal structure of the human protein, the loop is directly followed by strand A1′ in which the polypeptide chain forms hydrogen bonds with strand A1 of the Rpb7 subunit. The yeast Rpb4 sequence has a 11-residue insertion at this position (residues 36–46) that folds to form a pair of short antiparallel β-strands before entering strand A1′.

In the structure of the 12-subunit yeast polymerase, the loop between strands A3 and A4 of Rpb7 (the 'tip loop') contacts the core subunits of the polymerase. This loop region appears to be flexible as it is disordered in the high-resolution yeast crystal structure (16) and adopts slightly different conformations in the different copies of the Rpb7 subunit within the crystallographic asymmetric units of both the archaeal and human structures.

The S1 motif, present in the C-terminal domain of the E subunits, is found in a number of other proteins including bacterial and eukaryotic translation initiation factors, proteins involved in mRNA degradation, proteins involved in mRNA release from the spliceosome as well as bacterial transcription factors. Some of these proteins have been shown to bind
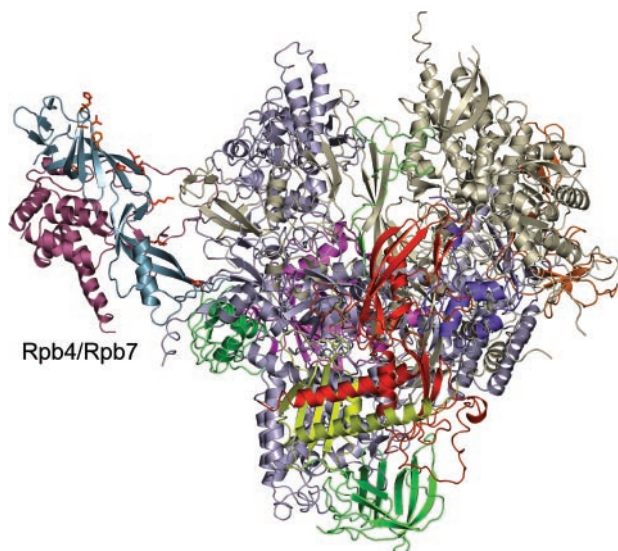
Figure 6. Ribbon diagram showing the location of the mutations and their effect on RNA binding, within the context of the RNAP$_{II}$. The crystal structure of the entire 12-subunits RNAP$_{II}$ from yeast (PDB entry: 1wcm) is shown, with the 10 core subunits colour-coded as in Cramer *et al*. (12). The Rpb4 and Rpb7 subunits are shown on the left in pink and blue, respectively. The yeast residues structurally equivalent to the residues that have been mutated in the human Rpb7 are shown in red and yellow. As in Figure 4, residues that have a strong effect on RNA binding are shown in red, while orange indicates residues that have a weaker effect on binding.

ssRNA *in vitro*. Despite a high degree of sequence similarity, no residue is absolutely conserved throughout the entire motif. From a structural viewpoint, the motif folds into a 5-stranded β-barrel known as OB fold, a common and ubiquitous structural motif, with no conserved sequence pattern, often involved in oligonucleotide binding. To date, no direct structural information is available on the interaction of S1 domains with nucleic acids, but the structures of several OB folds have been determined in the presence of RNA or DNA (23). Although the specific details of the interactions differ, in all these proteins oligonucleotide binding occurs on the face of the barrel containing the conserved residues in the S1 motif.

Gel mobility shift assays have confirmed that the yeast Rpb4/Rpb7 heterodimer binds single-stranded nucleic acid, with an apparent dissociation constant of 0.7 μM for DNA and 1.2 μM for RNA (17). We have subsequently demonstrated that both the archaeal E/F complex and the yeast A14/A43 complex (its homologue in RNAP$_I$) are able to bind nucleic acid with comparable affinities (24). This indicates that the RNA-binding function is evolutionary conserved in the archaeal and eukaryotic RNA polymerases. We have now quantified the binding of RNA to both the archaeal E/F heterodimer and the human Rpb4/Rpb7 heterodimer and in both cases determined a dissociation constant of 400–500 nM. This relatively low-binding affinity is consistent with the proposed role of the complex in the non-specific, transient binding of the newly synthesized RNA transcript. We have used the structure of both the archaeal *M.jannaschii* E/F complex and the human Rpb4/Rpb7 complex to identify residues potentially important for RNA recognition.

In the crystal structure of the *M.jannaschii* subunit, no electron density is visible for part of the loop connecting strand B4

and B5. In a number of OB folds this loop has been shown to play a central role in DNA or RNA binding (using arginine, serine and lysine residues to interact with the sugar-phosphate backbone) and often undergoes a conformational change to trap the nucleic acid (23). The sequence of the loop in *M.jannaschii* includes four positively charged residues that may play a role in non-sequence-specific nucleic acid binding. Although the type and order of these residues are not conserved in archaeal sequences, charged and polar residues are usually present. We have created a triple mutant in the E subunit, in which three positively charged residues (Arg155, Lys156 and Arg157) were eliminated. The mutation of the loop has a remarkable effect, completely abolishing RNA binding. Mutating two adjacent positively charged residues (Lys152 and Lys160) to glutamic acids also completely abolishes RNA binding. In the human Rpb7 sequence the loop is shorter, but mutation of the surrounding residues (Arg151, Asp153 and Phe158) also nearly abolish RNA binding, confirming the importance of this region in protein–nucleic acid interaction.

In deletion mutagenesis studies of *S.cerevisiae* Rpb7 (17), a mutant lacking the B4–B5 loop was shown to retain nucleic acid binding but was found to be inactive in transcription. Mutants in this earlier study were designed based on the homology to the S1 motif, without knowledge of the actual structure. The B4–B5 loop deletion (residues 151–158) included part of β-strands B4 and B5, but the fact that such deletion had no effect on nucleic acid binding may be correlated to the lack of positively charged residues in this region of the yeast sequence.

The S1 motif in the C-terminal domain of subunit E was the primary candidate for mutagenesis studies. However, the topology of the N-terminal domain of subunit E/Rpb7 resembles a 'truncated', non-canonical RNP fold. RNP domains are the most widely found and best characterized ssRNA-binding motifs and fold into a β1−α1−β2−β3−α2−β4 structure, with the four antiparallel β-strands packing against the two helices. A truncated RNP motif, lacking α2 and β4, is present in the structure, and resembles most closely the anticodon-binding domain of phenylalanine tRNA synthetase (25). However it was not clear whether this was simply a structural module, or whether it may be involved in RNA binding.

When the criteria for mutagenesis were applied to the N-terminal domain, only a handful of residues for mutagenesis could be identified. A number of well-conserved residues are located on two loops at the bottom of this domain (loop A1–K2 and A3–A4; Figure 2), including a number of proline and glycine residues, as well as hydrophobic amino acids. However the crystal structure of the 12-subunit complex shows that these loops are involved in the interaction with the RNAP core, in particular with helix 1 of Rpb6, three short regions of Rpb1 (including the N-terminus and the linker leading to the CTD) and the anchor region of Rpb2 (16). The conserved residues are therefore not available for RNA binding within the polymerase context, and have not therefore been tested. No reasonably conserved, surface-exposed residue could be identified on the canonical RNA-binding surface of the RNP domain. However, a few residues that fulfilled the criteria for mutagenesis were found on the face of the RNP domain which includes helix K2, and were shown to have an effect on RNA binding.

When the residues that are involved in RNA binding are plotted on the crystal structures of the archaeal and the human heterodimers, a similar pattern emerges, with residues located in similar positions having comparable effects on the interaction with nucleic acids, even when the chemical nature of the residues differ. The RNA-binding residues are clustered along the vertical axis of the elongated Rpb7/E subunit, spanning the length of the molecule from the N-terminal to the C-terminal domain. Within the OB fold, the residues that have the strongest impact on RNA binding are located on strands B3, B4 and B5 of the β-barrel.

In summary, combining structural information with mutational analysis and EMSAs on both the archaeal E subunit and the eukaryotic Rpb7 subunit, we have identified a putative RNA-binding path along the surface of the Rpb7/E subunit. Most of the residues that have an impact on RNA binding are located on the canonical nucleic acid binding face of the OB fold. Interestingly, the residues contributing to RNA binding in the N-terminal domain are not situated on the front surface of the β-sheet of the truncated RNP domain, the surface normally involved in nucleic acid binding.

Recent results have shown that yeast Rpb7 can be UV-crosslinked in a strong and reproducible manner to both strands of promoter DNA, between the TATA box and the start site (26), suggesting that in addition to its function in binding and stabilizing the nascent RNA, Rpb7 has an additional DNA-binding function during transcription initiation. In this scenario, it is possible that the RNA-binding surface we have identified is also involved in binding DNA during the initiation stage.

The human Rpb7 has been shown to interact with a number of transcription factors, some of which are involved in cancer development. Human Rpb7 interacts with the Ewings Sarcoma oncogene DNA-binding domain (27–29) and with the Nephoblastoma Overexpressed proto-oncogene, which in non-malignant cells is probably involved in the differentiation of several cell types (30). The retinoic acid receptor has been shown to bind to Rpb7 in its unliganded form and repress transcription of a number of transcriptional activators, such as AP-1 (31). Human Rpb7 has also been identified as a target of the von Hippel-Lindau protein, a component of a ubiquitin ligase E3 complex that acts as potent tumour suppressor and regulates the transcription and secretion of the vascular endothelial growth factor (32). All these data suggest that Rpb4/Rpb7 may mediate the interaction between a number of enhancer transcription factors and the RNA-polymerase core, to regulate gene expression. A high-resolution model of the human complex provides a structural framework to understand the multiple interactions between oncogenic transcription factors and Rpb7.

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Choder,M. (2004) Rpb4 and Rpb7: subunits of RNA polymerase II and beyond. *Trends Biochem. Sci.*, **29**, 674–681.
2. 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.
3. 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.
4. Tan,Q., Li,X., Sadhale,P.P., Miyao,T. and Woychik,N. (2000) Multiple mechanisms of suppression circumvent transcription defects in RNA polymerase mutant. *Mol. Cell. Biol.*, **20**, 8124–8133.
5. 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 activate transcription of a subset of genes. *J. Biol. Chem.*, **276**, 30641–30647.
6. Kimura,M., Suzuki,H. and Ishihama,A. (2002) Formation of a carboxy-terminal domain phosphatase (Fcp1)/TFIIF/RNA polymerase II (pol II) complex in *Schizosaccharomyces pombe* involves direct interaction between Fcp1 and the Rpb4 subunit of pol II. *Mol. Cell. Biol.*, **22**, 1577–1588.
7. 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.
8. Farago,M., Nahari,T., Hammel,C., Cole,C.N. and Choder,M. (2003) Rpb4p, a subunit of RNA polymerase II, mediates mRNA export during stress. *Mol. Biol. Cell*, **14**, 2744–2755.
9. 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.
10. Werner,F. and Weinzierl,R.O. (2002) A recombinant RNA polymerase II-like enzyme capable of promoter-specific transcription. *Mol. Cell*, **10**, 635–646.
11. Ouhammouch,M., Werner,F., Weinzierl,R.O. and Geiduschek,E.P. (2004) A fully recombinant system for activator-dependent archaeal transcription. *J. Biol. Chem.*, **279**, 51719–51721.
12. Cramer,P., Bushnell,D.A. and Kornberg,R.D. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science*, **292**, 1863–1876.
13. Todone,F., Brick,P., Werner,F., Weinzierl,R.O.J. and Onesti,S. (2001) Structure of an archaeal homologue of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell*, **8**, 1137–1143.
14. 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.
15. Bushnell,D.A. and Kornberg,R.D. (2003) Complete, 12-subunit RNA polymerase II at 4.1-A resolution: implications for the initiation of transcription. *Proc. Natl Acad. Sci. USA*, **100**, 6969–6973.
16. Armache,K.J., Mitterweger,S., Meinhart,A. and Cramer,P. (2005) Structures of complete RNA polymerase II and its subcomplex, Rpb4/7. *J. Biol. Chem.*, **280**, 7131–7134.
17. Orlicky,S.M., Trans,P.T., Sayre,M.H. and Edwards,A.M. (2001) Dissociable Rpb4–Rpb7 subassembly of RNA polymerase II binds to single-stranded nucleic acid and mediates a post-recruitment step in transcription initiation. *J. Biol. Chem.*, **276**, 10097–10102.
18. Werner,F., Eloranta,J.J. and Weinzierl,R.O.J. (2000) Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RPB4 and RPB12. *Nucleic Acids Res.*, **28**, 4299–4305.
19. Collaborative Computational Project Number 4 (1994), The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D*, **50**, 760–763.
20. Storoni,L.C., McCoy,A.J. and Read,R.J. (2004) Likelihood-enhanced fast rotation functions. *Acta Crystallogr. D*, **60**, 432–438.
21. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallographic and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D*, **50**, 760–763.
22. Jones,T.A., Zou,J.-Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A*, **47**, 110–119.
23. Theobald,D.L., Mitton-Fry,R.M. and Wuttke,D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 115–133.

24. Meka,H., Daoust,G., Bourke-Arnvig,K., Werner,F., Brick,P. and Onesti,S. (2003) Structural and functional homology between the RNAPI subunits A14/A43 and the archaeal RNAP subunits E/F. *Nucleic Acids Res.*, **31**, 4391–4400.

25. Goldgur,Y., Mosyak,L., Reshetnikova,L., Ankilova,V., Lavrik,O., Khodyreva,S. and Safro,M. (1997) The crystal structure of phenylalanyl-tRNA synthetase from *Thermus thermophilus* complexed with cognate tRNAPhe. *Structure*, **5**, 59–68.

26. Chen,B.S., Mandal,S.S. and Hampsey,M. (2004) High-resolution protein–DNA contacts for the yeast RNA polymerase II general transcription machinery. *Biochemistry*, **43**, 12741–12749.

27. Bertolotti,A., Melot,T., Acker,J., Vigneron,M., Delattre,O. and Tora,L. (1998) EWS, but not EWS-FLI-1, is associated with both TFIID and RNA polymerase II: interactions between two members of the TET family, EWS and hTAFII68, and subunits of TFIID and RNA polymerase II complexes. *Mol. Cell. Biol.*, **18**, 1489–1497.

28. Zhou,H. and Lee,K.A. (2001) An hsRPB4/7-dependent yeast assay for *trans*-activation by the EWS oncogene. *Oncogene*, **20**, 1519–1524.

29. 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.

30. Perbal,B. (1999) Nuclear localisation of NOVH protein: a potential role for NOV in the regulation of gene expression. *Mol. Pathol.*, **52**, 84–91.

31. Shen,X.Q., Bubulya,A., Zhou,X.F., Khazak,V., Golemis,E.A. and Shemshedini,L. (1999) Ligand-free RAR can interact with the RNA polymerase II subunit hsRPB7 and repress transcription. *Endocrine*, **10**, 281–289.

32. 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.

33. Kabsch,W. and Sander,C. (1983) Dictionary of protein secondary structures. *Biopolymers*, **22**, 2577–2637.