Article No. mb981557 J. Mol. Biol. (1998) 280, 317–322

JMB

COMMUNICATION

Repeated Tertiary Fold of RNA Polymerase II and Implications for DNA Binding

Jianhua Fu[1], Mark Gerstein[1], Peter R. David[1], Averell L. Gnatt[1] David A. Bushnell[1], Aled M. Edwards[2] and Roger D. Kornberg[1]*

[1]*Department of Structural Biology, Stanford University School of Medicine, Fairchild Center, Stanford, CA 94305 USA*

[2]*Banting and Best Department of Medical Research, C. H. Best Institute, 112 College St. University of Toronto, Toronto Ontario, Canada, M5G 1L6*

**Corresponding author*

X-ray diffraction data from two forms of yeast RNA polymerase II crystals indicate that the two largest subunits of the polymerase, Rpb1 and Rpb2, may have similar folds, as is suggested by secondary structure predictions. DNA may bind between the two subunits with its 2-fold axis aligned to a pseudo 2-fold axis of the protein.

© 1998 Academic Press

Keywords: DNA-binding; folding; RNA polymerase II; transcription; X-ray diffraction

Multisubunit RNA polymerases have been extensively conserved through evolution from bacteria to man. The two largest subunits, termed β and β′ in *Escherichia coli* and Rpb1 and Rpb2 in yeast, account for two-thirds of the mass of the polymerases and exhibit many regions of homology across species (Allison *et al*., 1985; Sweetser *et al*., 1987). Biochemical studies have demonstrated polymerase binding to as many as 50 base-pairs of DNA and have implicated the two largest subunits in this interaction (Young, 1991). Structural studies have led to the proposal of a straight path for promoter DNA extending at least 100 Å across the surface of the enzyme (Leuther *et al*., 1996). Here we report an analysis of X-ray diffraction from yeast RNA polymerase II (pol II) crystals and polymerase-DNA cocrystals, as well as results of secondary structure prediction, which suggest a structural basis for the polymerase-DNA interaction.

Yeast RNA polymerase II (587 kDa) lacking two subunits, Rpb4 and Rpb7, which together make up less than 9% of the mass of the protein (Woychik & Young, 1989, 1990) was crystallized in space group *I*222 (Table 1). DNA cocrystals were grown in space group *C*2 of the polymerase with a ''tailed'' template (Kadesch & Chamberlin, 1982), comprising 18 base-pairs of duplex DNA and a 15 base single-stranded extension at one 3′-end (see the legend to Table 1). Such a tailed template may be thought of as half of the unwound DNA ''bubble'' known to occur at the active site of RNA polymerases during transcription. Diffraction extending beyond 4 Å resolution was obtained from both crystal forms. Self-rotation functions (Rossmann & Blow, 1962) revealed peaks corresponding to non-crystallographic 2-fold axes in both crystal forms (Figure 1). The occurrence of these peaks was not dependent on the size of the search sphere nor on the resolution limits of the X-ray data.

Two observations indicate that the non-crystallographic 2-fold symmetry lies within the polymerase molecule and attest to its significance. First, the predicted solvent contents for one polymerase molecule in an asymmetric unit were typical for protein crystals: 64% with V_M† = 3.1 Å^3/Da, assuming one molecule per asymmetric unit of the C2 DNA cocrystal; 52% with $V_M = 2.6$ Å^3/Da, assuming one molecule per asymmetric unit of the *I*222 crystal (Table 1). The possibility that, in both crystal forms, the self-rotation peak arises from overlaps between multiple polymerase molecules within an asymmetric unit is therefore excluded. Second, the

Present address: M. Gerstein, Department of Molecular Biophysics and Biochemistry, Bass Center, 266 Whitney Avenue, Yale University, P.O. Box 208114, New Haven, CT 06520, USA.

Abbreviations used: PDB, Protein Data Base; pol II, yeast RNA polymerase II.

† Volume per unit protein mass in crystal (Matthews, 1968).

0022–2836/98/280317–06 $30.00/0

© 1998 Academic Press

Table 1. Crystallographic data on yeast pol II and pol II/DNA complex

Molecule	pol II	pol II/DNA complex
Unit cell parameters	$a = 131.3$ Å, $b = 223.5$ Å	$a = 176.9$ Å, $b = 220.4$ Å,
	$c = 369.2$ Å	$c = 196.5$ Å, $\beta = 102.9°$
X-ray source	SSRL BL7-1	CHESS F-1
Space group	*I*222	*C*2
R_{symm} (%)	12 for 50–4.0 Å	10.3 for 40–3.9 Å
Completeness overall (%)	72	65
Z (Z')[a]	8, (1)	4, (1)
V_s (%, v/v)	52	64
V_M (Å^3/Da)	2.6	3.4
Predicted crystal density (g/cm^3)	1.200	1.160
Measured crystal density (g/cm^3)	1.22 ± 0.03	–

Yeast pol II lacking subunits 4 and 7 was purified and crystallized as described (Edwards *et al.*, 1990, 1994), with modifications extending diffraction to 3.5 Å resolution (unpublished results). Cocrystals were prepared with the following synthetic DNA:

```
5′-CGTCTAACTTTTCTTCCCCCCCCCCCCCCCCCC-3′
3′-GCAGATTGAAAAGAAGGG-5′
```

The crystals were flash-frozen for X-ray diffraction experiments. All data were processed with the HKL package (Otwinowski, 1993). To confirm the estimated values of Z, V_s and V_M, crystal density was measured with three different gradient media and averaged. One of the gradient media was organic (bromobenzene/toluene mixture saturated with crystal mother liqor) and the others aqueous (Pertoft *et al.*, 1978; Westbrook, 1985). The measured density of the *I*222 pol II crystal agreed well with the predicted value. The plate-shaped pol II/DNA cocrystal (space group *C*2) was too thin to observe when submerged in gradient madia, so no experimental value for its density was obtained. The unit cell dimensions of the *C*2 DNA cocrystal, however, dictate that $Z' = 1$, because assuming $Z' = 2$, results in $V_s = 28\%$ and $V_M = 1.7$, unreasonable for protein crystals (Matthews, 1968). The space group ambiguity (*I*222 or $I2_12_12_1$) for the orthorhombic crystal of pol II was later resolved by results from a difference Patterson calculation between native and heavy-atom derivatized crystals (unpublished results).

[a] Z is the number of molecule(s) inside a unit cell; Z' is the number of molecule(s) per asymmetric unit.

non-crystallographic self-rotation peaks had large values: 74% of that for the crystallographic 2-fold axis in the *C*2 DNA cocrystal; 50% in the *I*222 crystal of pol II alone. As a basis for comparison, we constructed a hypothetical heterodimer of *Eco*RI and *Bam*HI endonucleases, which exhibit no sequence similarity but which possess a common tertiary fold (Newman *et al.*, 1994), and whose ratio of masses is about the same as for Rpb1 and Rpb2. When this heterodimer was packed in a *C*2 unit cell, a self-rotation function was calculated showing a non-crystallographic 2-fold peak in the direction of the pseudodyad between the *Eco*RI and *Bam*HI monomers with a value 50% of that for the crystallographic 2-fold axis†.

The strength of the non-crystallographic 2-fold peaks in the two different crystal forms and their intramolecular nature indicate that a large fraction of the polymerase molecule must contain a repeated tertiary fold. Three subunits, Rpb3, Rpb5 and Rpb9, occur in two copies in the polymerase (Woychik & Young, 1990), but they make up only 28% of the mass of the enzyme, so a 2-fold relationship between these subunit pairs alone is insufficient to account for the non-crystallographic 2-fold symmetry observed. Rather, a contribution from the two largest subunits Rpb1 (190 kDa) and Rpb2 (140 kDa), which make up 61% of the mass of the molecule, seems necessary. These two subunits, together with pairs of Rpb3, Rpb5 and Rpb9, form a molecular mass about 89% of that of pol II (lacking Rpb4 and Rpb7) that may be self-symmetric at the secondary structure level. Evidence for a common fold in Rpb1 and Rpb2 was unexpected, since the two subunits exhibit no discernible sequence similarity (see the legend to Figure 2). We sought support from a comparison of predicted secondary structures of Rpb1 and Rpb2 (Figure 2). Two separate multiple sequence alignments for Rpb1 and Rpb2 homologs identified regions of sequence conservation only within each group of the homologs. Secondary structures were calculated for the entire sequences of Rpb1 and Rpb2 (except for the C-terminal repeat domain, unique to Rpb1 (Young, 1991)) and were averaged over all homologs at each position of the conserved regions. The secondary structures were then

† For calculating a self-rotation function from a hypothetical, unrelated heterodimer, monomeric coordinates of the endonucleases *Eco*RI (leri) and *Bam*HI (1bam: Newman *et al.*, 1994) were extracted from the PDB. The two monomers were combined into a dimer and transformed into a *C*2 unit cell using the MidasPlus package and a crystal packing program (E. Abola, personal communication). Each of the monomers was then rotated and translated so they could be brought to overlap by a pseudodyad perpendicular to the *b*-axis. During this process, the molecular packing was monitored and unit cell parameters adjusted (final cell: $a = 125.2$ Å, $b = 60.4$ Å, $c = 140.6$ Å and $\beta = 110.5°$), to ensure appropriate intermolecular distances (no steric clash between any neighbors, but with the molecules close enough to each other to mimic intermolecular contacts in real protein crystals.). A self-rotation function was calculated with the program MERLOT (Fitzgerald, 1988), using reflections between 10 and 4 Å and a 20 Å cutoff radius.

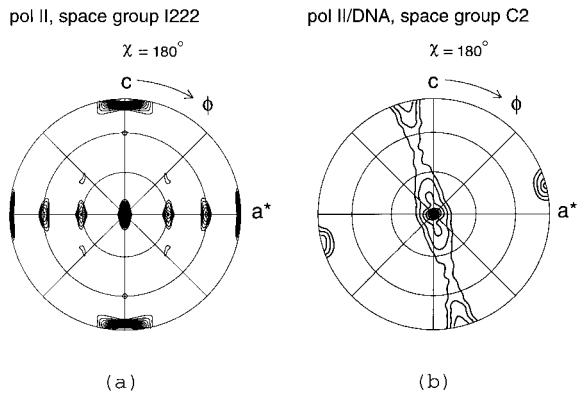

Figure 1. Stereographic projections of self-rotation functions at $\chi = 180°$ for: (a) pol II crystal in space group *I*222; (b) pol II/DNA cocrystal in space group *C*2. Contours starting at 2.5 and 2.0 σ with intervals of 0.5 and 0.2 σ are displyed in (a) and (b), respectively. Crystallographic *b*-axes for both crystal forms are perpendicular to the plane of the page. Directions of *a** and *c*-axes are indicated for both forms. In both (a) and (b), the ϕ angles run from 0° at the *c*-axis designation to 180° on the opposite side (arrows). The ψ angles vary from 0° at the center to 30° at the innermost circle, 60° at the next larger circle, and 90° at the outermost circle. In (a), peaks along the principal cell axes are crystallographic 2-folds. The non-crystallographic peaks of lower contours along $\phi = 90°$ and 270°, with $\psi = 30°$ and 60°, are symmetry-related by the crystallographic 2-folds along the *a**(*a*) and *b*-axes. Since there is one pol II molecule per asymmetric unit (Table 1), the non-crystallographic peak(s) in this crystal form can arise only from an intramolecular dyad. The direction of intramolecular dyad in this crystal form remains undetermined between $\psi = 30°$ and 60°. In (b), the peak along the *b*-axis corresponds to the crystallographic 2-fold. The lower peak along $\phi = 76°$ and $\psi = 90°$ is non-crystallographic and, since $Z' = 1$ (Table 1), can arise only from an intramolecular dyad as well. It is obvious that the non-crystallographic dyads in the two different crystals in (a) and (b) reveal the same molecular pseudosymmetry of pol II. All the self-rotation functions were calculated with the program AMoRe (Navaza, 1994), using reflections from 10 to 4.0 Å and a range of values of Patterson cutoff radius (25 to 50 Å, compared with diameter of the polymerase of about 130 Å (Darst *et al.*, 1991)). Any translational component of the pseudo-symmetry axis remains to be determined.

aligned manually, with considerable success: 45 of 69 helices and 56 of 90 strands could be matched, with especially good agreement in the conserved regions.

It occurred to us that a pseudo 2-fold symmetry of Rpb1 and Rpb2 might play a role in recognition of the 2-fold symmetric DNA double helix (Figure 3). This mode of polymerase-DNA interaction would be analogous to that of dimeric DNA-binding proteins with 2-fold symmetric DNA sequences. In the case of RNA polymerase, however, the requirement is for minimal specificity, to accommodate a wide range of promoter DNA sequences. Most sequences form essentially straight DNA, with a well-defined 2-fold axis perpendicular to the helix axis. The length of straight DNA at a promoter must be at least 30 base-pairs, extending from the TATA box to the transcription start site, corresponding to about 100 Å of DNA, comparable to the diameter of the polymerase molecule (Leuther *et al.*, 1996). Rbp1 and Rpb2 would seem well suited for binding such a length of straight DNA, with a 2-fold relationship between the subunits to match the 2-fold symmetry of the DNA and, in view of the large size of these subunits, sufficient for interaction with 100 Å of DNA.

The role of Rpb1 and Rpb2 in DNA binding may extend further. DNase protection mapping of transcription elongation complexes has shown the protection of 20 to 24 base-pairs (80 Å) upstream and 19 to 23 base-pairs downstream of the active (nucleotide addition) site (Rice *et al.*, 1993), including a region of about 15 base-pairs in single-stranded form around the active site. The mode of polymerase-DNA interaction described above for the upstream region could apply to the downstream region as well. The upstream segment would follow a straight path between Rpb1 and Rpb2 as described, bend around the polymerase in the active-center cleft by virtue of the flexibility of single-stranded DNA in this region, and then return on a straight path between Rpb1 and Rpb2 down the back side of the enzyme.

A straight path for DNA downstream of the active site may provide a structural basis for transcription pausing and termination. Sequences up to about 25 base-pairs downstream are important for pausing and termination, and these sequences often cause DNA bending (Kerppola & Kane, 1990; Lee *et al.*, 1990). A straight DNA-binding channel on the surface of the polymerase may be unable to accommodate bent DNA, at least in some orientations, and additional features of the sequence, such as flexibility of the double helix and structure of the RNA transcript, may contribute to pausing and termination as well.

Our conclusions regarding the relationship of Rpb1 and Rpb2 in yeast RNA polymerase II and our proposals for the mode of interaction with DNA are consistent with reports on *E. coli* RNA polymerase. It has been suggested that *E. coli* β and β′ subunits are arranged in an antiparallel manner (Landick & Roberts, 1996; Weilbaecher *et al.*, 1994; Nudler *et al.*, 1996). Scanning force micrographs of elongation complexes have shown the entry and exit of DNA from nearly the same point on the surface of the polymerase (Rees *et al.*, 1993), in keeping with a path of DNA up one side of the enzyme and down the other, as described above. Finally, results of electron crystallography on core *E. coli* RNA polymerase are indicative of a 2-fold symmetric distribution of protein density at low resolution (S. A. Darst, personal communication).

Figure 2. An alignment of predicted secondary structures of the two largest subunits of yeast RNA polymerase II. The deduced amino acid sequences of the second largest subunit (Rpb2) (top) and the largest subunit (Rpb1 or rpo21) (bottom) are shown. The range of numbers for each line of sequence is indicated in the left panel. Large insertions in either sequence are indicated by [...] and are omitted, as are N and C-terminal sequences. Predicted secondary structures, beneath the sequences, are symbolized by e with light shading for strands and h with dark shading for helices. The filled bars above each sequence indicate regions of sequence conservation among homologs of Rpb1 and Rpb2. The alignment was based on the consensus secondary structures predicted seperately for the two subunits. First, the sequences of both subunits were run against the OWL database (Bleasby *et al.*, 1994) using the FASTA program (Pearson & Lipman, 1988) with an *e*-value threshold of 0.001. In these searches it was never possible to get any appreciable fragment of Rpb2 or one of its homologs to match any fragment of Rpb1 or one of its homologs, confirming the absence of any sequence similarity between the subunits. Then, a multiple sequence alignment was constructed for each of the two subunits using the CLUSTALW program (Thompson *et al.*, 1994). (There were 54 sequences in the multiple alignment for Rpb2 and 52, for Rpb1.) There were five conserved ''blocks'' in the multiple alignment for each of the subunits. Third, secondary structure prediction was done on all 106 sequences in the multiple alignments using the GOR program (Garnier *et al.*, 1996). For each group of alignment, the prediction was averaged at each position to give a consensus secondary structure prediction. These procedures were checked by sending the sequences of the two largest subunits to the PHD prediction server at EMBL (Rost, 1996), which automatically does a database search, multiple alignment, and secondary structure prediction using slightly different methods. We obtained consistent results for the secondary structure prediction (data not shown, but available on the web at http://bioinfo.mbb.yale.edu/align). Finally, the secondary structures were aligned by eye, trying to maximize the number of overlapping elements (as opposed to overlapping residues) and to minimize the gaps in the conserved regions.

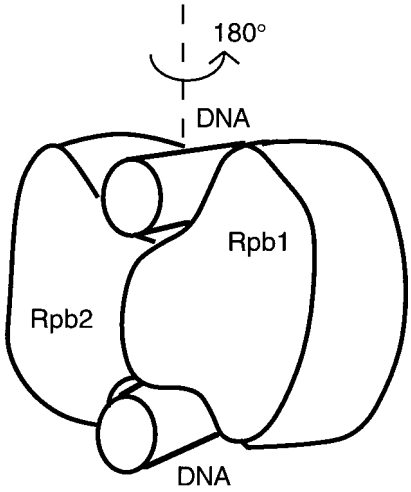

Figure 3. A diagram of RNA polymerase II - DNA interaction. The pseudo 2-fold axis relating the two largest subunits, Rpb1 and Rpb2, is indicated by a broken line, coinciding with the 2-fold axis of the DNA double helix.

Acknowledgments

We thank K. Flaherty, K. Ng, A. Kolatkar and J. Wedekind for asistance with X-ray facilities, and M. Zhang and C. Poglitsch for help with oligonucleotide purification. J.F. is supported by a fellowship from the Universitywide AIDS Research Program, University of California. M.G. acknowledges support from the NSF grant DBI-9723182. This research was funded by NIH grant AI49985 to R.D.K.

References

Allison, L. A., Moyle, M., Shales, M. & Ingles, C. J. (1985). Extensive homology among the largest subunits of eukaryotic and prokaryotic RNA polymerases. *Cell,* **42**, 599–610.

Bleasby, A. J., Akrigg, D. & Attwood, T. K. (1994). OWL: a non-redundant composite protein sequence database. *Nucl. Acids Res.* **22**, 3574–3577.

Darst, S. A., Edwards, A. M., Kubalek, E. W. & Kornberg, R. D. (1991). Three-dimensional structure of yeast RNA polymerase II at 16 Å resolution. *Cell,* **66**, 121–128.

Edwards, A. M., Darst, S. A., Feaver, W. J., Thompson, N. E., Burgess, R. R. & Kornberg, R. D. (1990). Purification and lipid layer crystallization of yeast RNA polymerase II. *Proc. Natl Acad. Sci. USA,* **87**, 2122–2126.

Edwards, A. M., Darst, S. A., Hemming, S. A., Li, Y. & Kornberg, R. D. (1994). Epitaxial growth of protein crystals on lipid layers. *Nature Struct. Biol.* **1**, 195–197.

Fitzgerald, P. M. D. (1988). MERLOT: an integrated package of computer programs for the determination of crystal structures by molecular replacement. *J. Appl. Crystallog.* **21**, 273–278.

Garnier, J., Gibrat, J. F. & Robson, B. (1996). GOR method for predicting protein secondary structure from amino acid sequence. *Methods Enzymol.* **266**, 540–553.

Kadesch, T. R. & Chamberlin, M. J. (1982). Studies of *in vitro* transcription by calf thymus RNA polymerase II using a novel duplex DNA template. *J. Biol. Chem.* **257**, 5286–5295.

Kerppola, T. K. & Kane, C. M. (1990). Analysis of the signals for transcription termination by purified RNA polymerase II. *Biochemistry,* **29**, 269–278.

Landick, R. & Roberts, J. W. (1996). The shrewd grasp of RNA polymerase. *Science,* **273**, 202–203.

Lee, D. N., Phung, L., Stewart, J. & Landick, R. (1990). Transcription pausing by *Escherichia coli* RNA polymerase is modulated by downstream DNA sequences. *J. Biol. Chem.* **265**, 15145–15153.

Leuther, K. K., Bushnell, D. A. & Kornberg, R. D. (1996). Two-dimensional crystallography of transcription factor IIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell,* **85**, 773–779.

Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A,* **50**, 157–163.

Newman, M., Strzelecka, T., Dorner, L. F., Schildkraut, I. & Aggarwal, A. K. (1994). Structure of restriction endonuclease BamHI and its relationship to EcoRI. *Nature,* **368**, 660–664.

Nudler, E., Avetissova, E., Markovtsov, V. & Goldfarb, A. (1996). Transcription processivity: protein-DNA interactions holding together the elongation complex. *Science,* **273**, 211–217.

Otwinowski, Z. (1993). Oscillation data reduction program. In *Proceedings of the CCP4 Study Weekend: Data Collection and Processing*, pp. 56–62, SERC Daresbury Laboratory, Warrington WA4 4AD, UK.

Pearson, W. R. & Lipman, D. J. (1988). Improved tools for biological sequence analysis. *Proc. Natl Acad. Sci. USA,* **85**, 2444–2448.

Pertoft, H., Laurent, T. C. & Laas, T. (1978). Density gradients prepared from colloidal silica. *Anal. Biochem.* **88**, 271–182.

Rees, W., Keller, R. W., Vesenka, J. P., Yang, G. & Bustamante, C. (1993). Evidence of DNA bending in transcription complexes imaged by scanning force microscopy. *Science,* **260**, 1646–1649.

Rice, G. A., Chamberlin, J. & Kane, C. M. (1993). Contacts between mammalian RNA polymerase II and the template DNA in a ternary elongation complex. *Nucl. Acids Res.* **121**, 113–118.

Rossmann, M. G. & Blow, D. M. (1962). The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallog.* **15**, 24–31.

Rost, B. (1996). PHD: predicting one-dimensional protein secondary structure by profile-based neural networks. *Methods Enzymol.* **266**, 525–539.

Sweetser, D., Nonet, M. & Young, R. A. (1987). Prokaryotic and eukaryotic RNA polymerases have homologous core subunits. *Proc. Natl Acad. Sci. USA,* **84**, 1192–1196.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensibility of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.

Weilbaecher, R., Hebron, C., Feng, G. & Landick, R. (1994). Termination-altering amino acid substitutions in the β′ subunit of *Escherichia coli* RNA

polymerase identify regions involved in RNA chain elongation. *Genes Dev.* **8**, 2913–2927.
Westbrook, E. M. (1985). Crystal density measurements using aqueous ficoll solutions. *Methods Enzymol.* **114**, 187–196.
Woychik, N. A. & 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.
Woychik, N. A. & Young, R. A. (1990). RNA polymerase II: subunit structure and function. *Trends in Biochem. Sci.* **15**, 347–351.
Young, R. A. (1991). RNA polymerase II. *Annu. Rev. Biochem.* **60**, 689–715.

Edited by A. Klug

(Received 18 June 1997; received in revised form 14 November 1997; accepted 28 November 1997)

Note added in proof: After this paper went to press, an NMR structure of Rpb8 was published showing a 2-fold symmetry of the tertiary fold, despite a lack of sequence repetition (Krapp, S., Kelly, G., Reischl, J., Weinzerl, R. O. & Matthews, S. (1998). *Nature Struct. Biol.* **5**, 110–114). Although Rpb8 is too small (17 kDal) to account for the non-crystallographic symmetry peak reported here, the Rpb8 structure may form part of a more extensive internal structural repeat of the RNA polymerase molecule.