The EMBO Journal vol.12 no.7 pp.2601 – 2607, 1993

Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals

Patrick Schultz, Hervé Célia, Michel Riva[1], André Sentenac[1] and Pierre Oudet

Laboratoire de Génétique Moléculaire des Eucaryotes, 11 rue Humann, F-67085 Strasbourg cedex and [1]Service de Biochimie et Génétique Moléculaire, Centre d'Etudes de Saclay, F-91191 Gif sur Yvette cedex, France

Communicated by A.Sentenac

Two-dimensional crystals of yeast RNA polymerase I dimers were obtained upon interaction with positively charged lipid layers. A three-dimensional surface model of the enzyme was determined by analyzing tilted crystalline areas and by taking advantage of the non-crystallographic internal symmetry of the dimer to correct for the missing viewing directions. The structure shows, at ~3 nm resolution, an irregularly shaped molecule 11 nm × 11 nm × 15 nm in size characterized by a 3 nm wide and 10 nm long groove which constitutes a putative DNA binding site. The overall structure is similar to the *Escherichia coli* holo enzyme and the yeast RNA polymerase II Δ4/7 structures. The most remarkable structural feature is a finger-shaped stalk which partially occludes the entrance of the groove and forms a 2.5 nm wide channel. We discuss the possible location of the catalytic centre and of the carboxy-terminal region of the β-like subunit in the channel. The interference of different DNA fragments with RNA polymerase dimerization and crystallization indicates the orientation of the template in the putative DNA binding groove.

Key words: electron microscopy/image processing/RNA polymerase I (A)/two-dimensional crystallization/three-dimensional structure

Introduction

In eukaryotes, the DNA-dependent RNA polymerases recognize the transcription start site (by interacting with an upstream preinitiation complex), initiate transcription, elongate the nascent RNA chain and finally terminate and release their product. Basal transcription levels and the essential up and down regulations are thus generated by the concerted interaction of a large number of macromolecules. The synthesis of eukaryotic ribosomal RNA precursors is catalysed by class I RNA polymerase, a complex molecular assembly which in the yeast *Saccharomyces cerevisiae* is composed of 14 distinct polypeptide chains with a total mass of ~650 kDa (Sentenac, 1985). Most of the catalytic functions are carried out by the two large polypeptide chains of 186 kDa (A_{190}) (Mémet *et al.*, 1988a) and 136 kDa (A_{135}) (Yano and Nomura, 1991). (The subunit nomenclature identifies the yeast enzyme class to which it belongs A (I), B (II) and C (III) and the apparent size of the polypeptide in kDa × 10^{-3}.) The sequence of A_{190} revealed eight major regions conserved within subunits B_{220} and C_{160} and with the bacterial β′ subunit, colinearly distributed over the entire sequence (Allison *et al.*, 1985; Ahearn *et al.*, 1987; Mémet *et al.*, 1988b; Jokerst *et al.*, 1989). Like β′, the A_{190} subunit or its homologue in other Eukaria and Archea enzymes is probably involved in binding the DNA template (Buhler *et al.*, 1974; Bréant *et al.*, 1983; Sentenac, 1985) and the nascent RNA chain (Gundelfinger, 1983; Riva *et al.*, 1987). Similarly, the sequence of A_{135} shows nine major homology regions with B_{150} and C_{128} subunits (Falkenburg *et al.*, 1987; Sweetzer *et al.*, 1987; James *et al.*, 1991; Yano and Nomura, 1991) and corresponds to the *Escherichia coli* β subunit. Like β (Grachev *et al.*, 1987, 1989), it interacts with the initiator nucleoside triphosphate and the nascent RNA chain (Riva *et al.*, 1987, 1990) and also probably participates in DNA binding (Huet *et al.*, 1982; Gundelfinger, 1983). Suppressor genetic experiments indicate that the carboxy-terminus of A_{135} and the amino-terminus of A_{190}, both containing a zinc-binding consensus sequence, are in close proximity (Yano and Nomura, 1991).

Subunit AC_{40} is required for enzyme assembly, as determined by *in vitro* mutagenesis, and shows sequence homology with a motif present in the prokaryote α′subunit (Mann *et al.*, 1987). Such a motif is also present in subunit AC_{19} (Dequard-Chablat *et al.*, 1991) and in the B_{44} dimer (Woychik and Young, 1990). Therefore the two largest subunits and the α-like homo- or heterodimer constitute a widespread core of subunits present in all higher eukaryotes RNA polymerases and in the bacterial core enzyme.

A set of five polypeptides (ABC_{27}, ABC_{23}, $ABC_{14.5}$, $ABC_{10\alpha}$ and $ABC_{10\beta}$) probably satisfies the specific requirement of eukaryotic cells since they are absent in the bacterial enzyme and are shared with class II and III enzymes. Five additional polypeptides, A_{49}, A_{43}, $A_{34.5}$, A_{14} and $A_{12.2}$, are associated specifically with the yeast form I enzyme. Subunits A_{49} and $A_{34.5}$ were found to be dispensable for RNA chain elongation both *in vitro* (Huet *et al.*, 1975) and *in vivo* (Liljelund *et al.*, 1992; M.Riva, unpublished results).

Recent electron microscopic examination of two-dimensional crystals of *E.coli* RNA polymerase (Darst *et al.*, 1988) and yeast RNA polymerases I (Schultz *et al.*, 1990a) and II (Edwards *et al.*, 1990) grown on charged lipid layers shed some light on their complex molecular arrangement. To interpret the projection map of the crystals formed by the eukaryotic class A enzyme, we have determined the three-dimensional structure of negatively stained crystals, tilted at various observation angles, up to a resolution of ~3 nm. The irregularly shaped structure, 11 nm × 11 nm × 15 nm in size, shows an overall structural homology with the *E.coli* enzyme model (Darst *et al.*, 1989) and a striking similarity with the yeast II Δ4/7 enzyme (Darst *et al.*, 1991). In particular, the presence of a putative DNA binding groove and a finger-shaped domain forming an almost closed

channel are observed in all multisubunit RNA polymerases studied, which probably reflects their sequence and functional homologies. A structure compatible in size and shape with the channel region is observed in the high resolution model of monomeric nucleotide polymerases (Ollis *et al.*, 1985; Chung *et al.*, 1990; Sousa, 1991; Kohlstaedt *et al.*, 1992). We discuss the existence of a common polymerase domain which contains the C-terminal half of the β-like subunit and is located in the channel region. The ability of RNA polymerases complexed to DNA to crystallize and to form dimers was examined. Our results indicate that the 5′ coding strand end interferes with dimerization.

Results

Two-dimensional crystallization

The RNA polymerase I preparation was incubated with positively charged lipid layers essentially as previously described (Schultz *et al.*, 1990a). The conditions were optimized to promote the formation of dimers, which are the building blocks of the crystals, by increasing the glycerol concentration to 20%. Moreover the nature of the lipids was chosen to produce planar layers in the absence of vesicular structures. In these conditions mainly RNA polymerase I dimers were observed and, by analysing images of individual dimers, we demonstrated that they are preferentially oriented in contact with the lipid layer (data not shown).

The large but often sheared arrangements of RNA polymerase I were formed by a single protein layer (Figure 1A) as evidenced by the same contrast of the crystallized and of free molecules relative to the background. Three-dimensional growth was not observed under these conditions. The best organized arrays were analysed to produce a translational undistorted average image. The diffraction patterns of the average images were indexed on a lattice of unit cell a = 27.5 nm, b = 27.5 nm and γ = 115.7° (Figure 1B). A two-fold symmetry axis perpendicular to the crystal plane was identified upon refinement of the diffraction peak phase origin. This symmetry was respected for 97 reflections down to a resolution of 1/2.5 nm^{-1} showing a mean phase residual of 11.5°. A symmetrized image was calculated using the averaged terms of seven independent crystals (Figure 1C) which showed the delineated unit cell composed of two RNA polymerase I dimers. The projection of the dimer was the same as previously described (Schultz *et al.*, 1990a) except for the absence of mirror symmetry. In these conditions the orientation of the RNA polymerases to the lipid layer was unique. The dimer is characterized by its elongated shape of 20 nm × 11 nm, which is divided into two asymmetric monomer projections, a and b (dotted contour in Figure 1C). The monomer projection a consists of a large globular domain 10 nm × 9 nm in size with an irregular contour from which a 3 nm thick spike protrudes 5 nm outwards. The monomer projection b exhibits two elongated domains 10 nm × 4 nm and 8 nm × 4 nm in size lying almost parallel, with their long axes perpendicular to the longest dimer dimension.

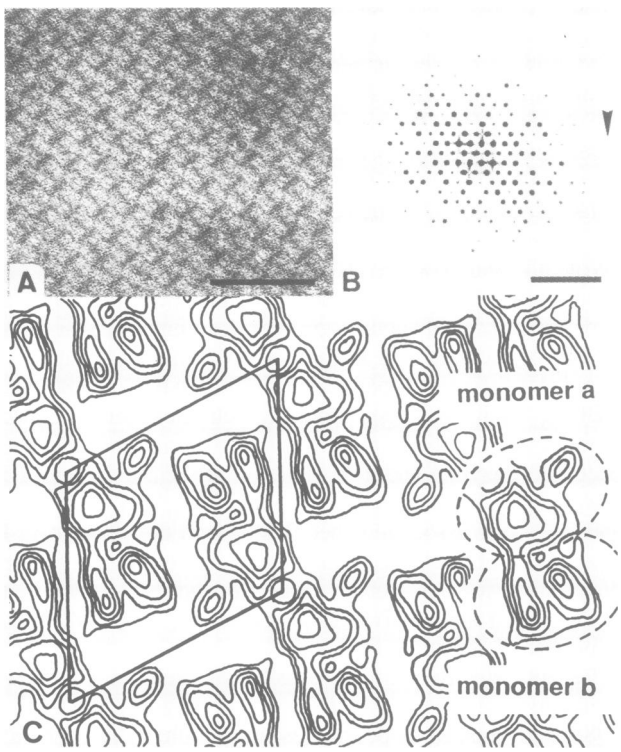

Fig. 1. (**A**) Partial field of a micrograph showing a crystalline area of RNA polymerase I molecules formed on cetyl trimethyl ammonium/ oleyl alcohol layers and negatively stained with uranyl acetate. The flat crystals are 0.2–1.0 μm in size and consist of a single molecular layer. The bar represents 0.1 μm. (**B**) Computed power spectrum of the Fourier transform of a correlation averaged crystal. Peaks up to 1/2.5 nm^{-1} are resolved in this pattern (reflection 12,1 is marked by an arrowhead). The bar represents 0.2 nm^{-1}. (**C**) Two-dimensional symmetrized and contoured projection map synthesized using 97 terms up to 1/2.5 nm^{-1}. The monomers a and b described in the text are contoured with a dotted line on the projection of one RNA polymerase I dimer. The parameters of the unit cell are: $a = b$ = 27.5 nm, γ = 115.7°. The unit cell is delineated by solid lines.

Three-dimensional model

Fourteen independent tilt series ranging from 0 to 58° yielded 61 different projections of the unit cell. Forty average

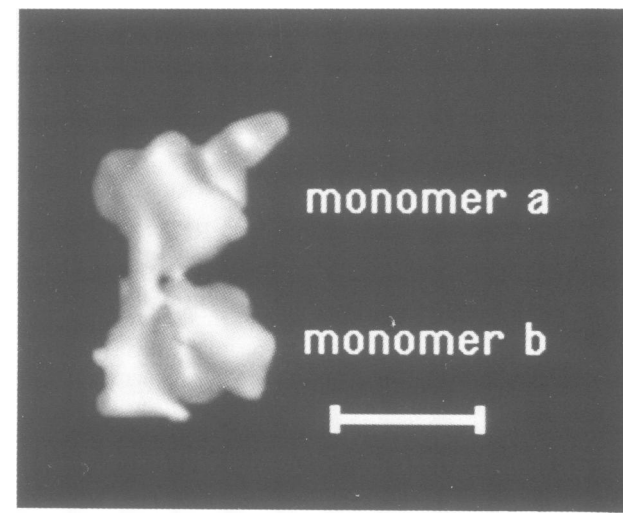

Fig. 2. Three-dimensional surface model of the RNA polymerase I dimer as reconstructed from 40 different crystal projections. The monomers a and b defined in Figure 1C are indicated. The contours represent the stain-excluding volume and were chosen to separate adjacent dimers leaving some contacts between the monomers a and b. The model is viewed from the top of the crystal, the lipid plane lying in the back. Impression of relief in Figures 2, 3 and 5 is provided by shading programs (Saxton, 1985). The bar represents 10 nm.

projections had a resolution better than 3.0 nm and were combined to back-project the stain-excluding volume. Figure 2 represents the shaded surface of the RNA polymerase I dimer. The density threshold was set to separate two adjacent dimers leaving connections between monomers. The volume filled by a dimer is 1160 nm^3, ~73% of that expected for two RNA polymerase I molecules, given a protein density of 1.3 g/cm^3.

That the two monomers constituting the asymmetric unit had different orientations in respect to the lipid plane was a favourable feature which allowed us to complete the missing data due to the limited tilt angles. An iterative search for the highest real space three-dimensional density correlation of the monomeric envelopes showed that the dimer contains a two-fold symmetry axis perpendicular to their long axes and tilted by 30° to the normal of the lipid plane. This additional symmetry element was imposed to calculate an average monomer model (Figure 3) in which the two projection cones overlapped by 28° since the monomer orientations differed by 60° and the highest tilt angle was of 58°.

The three-dimensional structure of yeast RNA polymerase I fits in an elongated irregular envelope of 11 nm × 11 nm × 15 nm and looks like a right hand fist with a prominent back. A thumb-like structure extends 3 nm out of the globular core then curves in and forms an almost closed 2.5 nm wide channel. The channel opens on a flat surface formed by the thumb-like structure and the extremity of an elongated 12–13 nm domain. This domain forms the wall of a bent groove which extends the channel and ends near the dimer contact region. The channel and groove have a total length of ~10 nm. A large and featureless apical region is located behind the groove and contacts the lipids in the monomer b orientation which was shown to be the preferential orientation of isolated monomers in contact with the lipid film (Schultz *et al.*, 1990a).

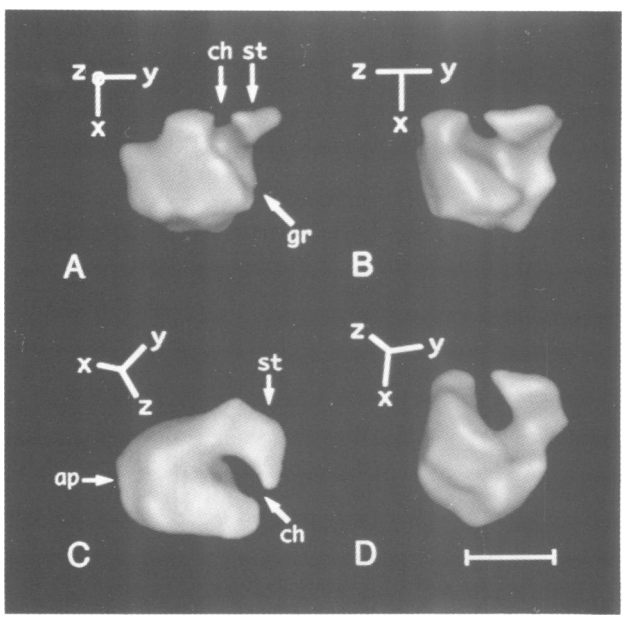

Fig. 3. Surface representation of the yeast RNA polymerase I monomer obtained after imposing the non-crystallographic two-fold symmetry present in the dimer shown in Figure 2. The major structural features described in the text are indicated by ch (channel), st (finger-like stalk), gr (groove) and ap (apical region). The orientations of the model are indicated by orthonormal axes where the orientation in panel A is taken as a reference. Panels A and B represent the orientation of monomers a and b respectively in the crystal; the lipid film is located at the back. The bar represents 6.5 nm.

Interaction with DNA

RNA polymerase was incubated with the different synthetic DNA templates listed in Table I. The ability of the complexes to bind to the lipid layers, to form dimers and to crystallize was then checked. Templates 1 and 2 were respectively 20 and 41 nucleotides long, double-stranded DNA with an internal gap of one nucleotide at positions 9 and 18 respectively. The gaps were shown to be recognized as an initiation site by calf thymus RNA polymerase II (Dreyer and Hausen, 1976). The coding 3′ end of DNAs 3 and 4 present a 9 nucleotide long, single-stranded poly(C) extension and a double-stranded portion respectively 11 and 29 nucleotides long. It has been shown that purified calf thymus RNA polymerase II initiates preferentially at the junctions between single- and double-stranded regions (Kadesch and Chamberlin, 1982).

DNA was not found to interfere significantly with the binding of RNA polymerases to the lipid film up to a DNA:protein molar ratio of 30:1. For higher ratios the amount of bound protein was, however, reduced. The RNA polymerase I dimer to monomer ratio decreased with increasing amounts of DNAs 2 or 4 whereas it remained essentially unchanged when incubated with DNAs 1 or 3 in the same conditions. Moreover crystallization was observed with DNAs 1 and 3 at DNA:protein molar ratios between 1:1 and 30:1 whereas no ordered areas were found using DNAs 2 or 4. The differential behaviour of RNA polymerases incubated with the various DNA templates indicates that the enzyme binds to both types of template [gapped or poly(C) extended]. Crystallization of the nucleoprotein complexes is impaired probably through dimer dissociation when the distance from the 5′ coding strand end to the initiation site increases from 12 to 23 bp.

Discussion

Two-dimensional crystallization

Two-dimensional crystallization of biological macromolecules using either charged (Darst *et al.*, 1988, 1989, 1991; Schultz *et al.*, 1990a) or ligand exposing (Uzgiris and Kornberg, 1983; Ribi *et al.*, 1987; Lebeau

Table I. The synthetic oligonucleotides used to study the interference of their interaction with RNA polymerase with the ability of the enzyme to form dimers and to crystallize

DNA1 20-mer	3'CGAATATATCTTATAATGCC 5'GCTTATAT GAATATTACGG
DNA2 41-mer	3'GCAGCGGTATATATATATATCTTATAATGAATATTGGCGAC 5'CGTCGCCATATATATAT TAGAATATTACTTATAACCGCTG
DNA3 20-mer	3'CCCCCCCCCCTTATAATGCC 5' GAATATTACGG
DNA4 38-mer	3'CCCCCCCCCGGGATGCTTGCATGGTCGGCGCCATTAAA 5' CCCTACGAACGTACCAGCCGCGGTAATTT

The subunit nomenclature identifies the yeast enzyme class to which it belongs—either A (I), B (II) or C (III)—and the apparent size of the polypeptide in kDa × 10^{-3}.

et al., 1990; Mosser *et al.*, 1991) lipid layers has grown to become a powerful tool for electron microscopic studies of soluble proteins. These studies have shown that proteins are concentrated at the lipid–solvent interface upon interaction with the lipids and, in some cases, it has been shown that they are preferentially oriented (Schultz *et al.*, 1990a). The macromolecules can then diffuse in plane and protein–protein contacts can be stabilized leading to ordered areas. Both specific orientations and crystallization are important properties which facilitate image averaging by respectively single molecular image alignment or crystallographic methods.

The RNA polymerase I dimer–dimer contacts leading to crystallization are probably unstable since up to four different crystal forms can be obtained (Schultz *et al.*, 1990a; P.Schultz, unpublished observations). Slight modifications in the lipid layer composition or in the aqueous environment can favour peculiar protein—protein interactions. The dimeric organization reflects a more stable structure since it is observed in all crystal forms and appears to be specific for yeast RNA polymerase I since it was not detected for any other RNA polymerase studied so far (Darst *et al.*, 1989, 1990, 1991). The interaction is salt sensitive and mainly monomers are observed at ionic strength higher than 200 mM NaCl (Schultz *et al.*, 1990b).

The precise orientation of dimers to the lipid plane is preserved in all crystal forms indicating a specific protein–lipid interaction. The preferential projection of RNA polymerase I monomers was found to be similar to that of monomer b in the crystal (Schultz *et al.*, 1990a). The orientation of our three-dimensional model with respect to the lipid film indicates that the apical region of the enzyme interacts with the lipid layer (Figure 3) orienting the putative DNA-binding groove (see below) towards the solvent. The question whether such an oriented adsorption has any biological significance for enzyme presentation or segregation in the nucleolus remains open.

Structural similarity with eukaryotic and prokaryotic RNA polymerases

Negatively stained two-dimensional crystals were used to determine a surface model of the RNA polymerase I molecule at a resolution of ~3 nm. The projection of the actual protein density, in the absence of stain, was investigated by analysing similar crystals included in vitreous water. The resulting projection map has similar contours to the stained projection map up to a resolution to 2.5 nm (result not shown) indicating that the surface revealed by the negative stain is a reasonable approximation. The three-dimensional model is consistent with the projection maps of isolated, glutaraldehyde fixed monomers or dimers adsorbed on to carbon film (Schultz *et al.*, 1990a,b) suggesting that the interaction with lipids does not alter the RNA polymerase structure at least at 3 nm resolution. All the described structural features were observed in both monomers forming the dimer. Slight differences were, however, observed in the depth of the groove, the shape of the finger-like stalk and in the monomer contact regions. These differences are probably due to variations in stain penetration, flexibility of some parts of the molecule and missing tilt information.

The overall structural organization of the yeast enzyme I is comparable to that of *E.coli* and yeast type II Δ4/7 enzymes determined by electron crystallography (Darst *et al.*, 1989, 1991). The thumb-like feature, the channel and groove are present in all multimeric RNA polymerases studied. This morphological similitude reflects the sequence homologies between the large polypeptides of the prokaryotic and eukaryotic enzymes. Curiously the *E.coli* holo enzyme appears longer (10 × 10 × 16 nm) than the eukaryotic models despite its smaller molecular weight. This observation can be related to the presence of the σ factor which may be located in the apical region. The yeast RNA polymerase II Δ4/7 envelope shows an additional protein density protruding near the thumb-like domain (density D in Darst *et al.*, 1991). This domain was recently shown to be different from the carboxy-terminal heptapeptide repeat (S.A.Darst, personal communication).

Identification of a common polymerase domain

The most prominent features in the three-dimensional models of multimeric RNA polymerases are the deep cleft and the thumb-like protrusion forming the channel region (Figure 3). It has been shown that the envelope of the Klenow fragment can be superimposed almost exactly on the channel region of the prokaryotic enzyme (Darst *et al.*, 1989). Similar features have been described not only in the atomic structure of the Klenow fragment of DNA polymerase I (Ollis *et al.*, 1985) but also in the T7 RNA polymerase (Chung *et al.*, 1990; Sousa, 1991) and in the HIV reverse transcriptase (Kohlstaedt *et al.*, 1992). This structural similarity within the monomeric nucleotide polymerases is founded on the conservation of the secondary structural elements in number, length and position and on the identification, in most sequences, of three conserved structural motifs (A, B and C) brought into close spatial proximity by the tertiary structure. With regard to this structural homology, a common polymerase fold has been proposed for monomeric nucleotide polymerases (Delarue *et al.*, 1990) which probably reflects the common functional properties of nucleotide polymerases, such as polynucleotide template binding, nucleotide triphosphate binding, nucleotide polymerization, processivity and polynucleotide product binding. The observed morphological similarity with multimeric DNA-dependent RNA polymerases suggests that the polymerase fold hypothesis can be extended to this family of enzymes.

This hypothesis implies that in fact only a small part of the multimeric RNA polymerase structure is sufficient to support the polymerization activity. Presently, experimental data are lacking to demonstrate that fragments of the β or β'-like subunits carry such an activity and it cannot be excluded that both subunits contribute to the correct folding of the polymerization domain. Whatever the case, the C-terminal half of the β-like subunit probably contributes to the active site since chemical cross-linking studies have localized the nucleotide triphosphate binding site in this region in different organisms (Grachev *et al.*, 1987, 1989; Riva *et al.*, 1990). Precise mapping of the amino acids labelled by the nucleotide derivatives identified five lysyl residues clustered in the conserved region H which maps between Asn946 and Met999 in the yeast enzyme II. Site-directed mutagenesis indicated that Lys979 (which is perfectly conserved in all multimeric RNA polymerases) and Lys987 are both indispensable for cell viability and constitute the most probable targets of the nucleotide derivatives (Treich *et al.*, 1992). Similar experiments performed on

DNA polymerase I identified motif B as interacting with nucleotide triphosphates (Joyce, 1991 and references therein). The aligned sequences of viral (Patel and Pickup, 1989), bacterial (Ovchinnikov *et al.*, 1981), archaebacterial (Berghöfer *et al.*, 1988; Leffers *et al.*, 1989; Puehler *et al.*, 1989), chloroplast (Ohme *et al.*, 1986; Ohyama *et al.*, 1986) and eukaryotic (Sweetzer *et al.*, 1987; James *et al.*, 1991; Seifarth *et al.*, 1991; Yano and Nomura, 1991) β-like subunits show that region H is highly conserved in all multisubunit RNA polymerases (Figure 4A). Moreover, region H shares some common features with motif B found in monomeric nucleotide polymerases (Delarue *et al.*, 1990). Two residues (Lys979 and Gly988 in the yeast β-like subunit) are invariant in all enzyme families and their relative positions are identical (Figure 4B). In addition, the environment of these invariant residues appears to be significantly conserved. The invariant Tyr in motif B is changed to a Lys in region H and the motifs A and C of monomeric nucleotide polymerases cannot be unambiguously related to any conserved regions in the β or β′-like subunits, suggesting that there is substantial sequence divergence between the multimeric polymerases and the consensus polymerase motifs. These data indicate that the conserved region H of the β-like subunit contributes to the formation of the active site, and the projection of the monomeric

A

RPO2_VACCV	875	RRP.IL	GDK	FTT	R	TSQ	KG	TVAYVADETEL	P	YDE
RPOB_ECOLI	1058	RRIQ.P	GDK	MAG	R	HGN	KG	VISKINPIEDM	P	YDE
RPOB_TOBAC	812	REIK.V	GDK	VAG	R	HGN	KG	IISKILPRQDM	P	YLQ
RPOB_MARPO	807	RKIQ.I	GDK	VAG	R	HGN	KG	IISKILPRQDM	P	FLQ
RPOU_HALHA	358	RIPE.L	GDK	FAS	R	HGQ	KG	VVGHLAPQEDM	P	FTQ
RPOU_METTH	355	RQPEFI	GDK	FAS	R	HGQ	KG	VVGLIVSQEDM	P	FTE
RPOB_SULAC	868	RIPE.I	GDK	FAT	R	HGQ	KG	VVGILIDQVDM	P	YTA
RPA2_YEAST	909	RTPQ.I	GDK	FSS	R	HGQ	KG	VCSRKWPTIDM	P	FSE
RPB2_YEAST	972	KIPQ.I	GDK	FAS	R	HGQ	KG	TIGITYRREDM	P	FTR
RPC2_YEAST	904	RRPE.L	GDK	FSS	R	HGQ	KG	VCGIIVKQEDM	P	FND
RPC2_DROME	894	RIPR..	GDK	FSS	R	HGQ	KG	VTGLIVEQEDM	P	FND

B

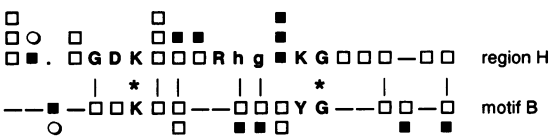

Fig. 4. (**A**) Amino acid sequence alignment of the conserved region H present in the β-like subunit of multimeric RNA polymerases from different organisms. The aligned sequences are vaccinia virus 132 kDa polypeptide (RPO2_VACCV), *E.coli* β subunit (RPOB_ECOLI), the tobacco (RPOB_TOBAC) and liverwort (RPOB_MARPO) chloroplast open reading frames, the *Halobacterium halobium* (RPOU_HALHA) and *Methanobacterium thermoautotrophicum* (RPOU_METTH) B′ subunits, *Sulfolobus acidocaldarius* β-like subunit (RPOB_SULAC), *Saccharomyces cerevisiae* type I (RPA2_YEAST), II (RPB2_YEAST) and III (RPC2_YEAST) second largest subunits and *Drosophila melanogaster* second largest subunit (RPC2_DROME). Conserved residues are framed. (**B**) The conserved region H from the β-like subunit of multimeric RNA polymerases is compared with motif B derived from 20 different monomeric nucleotide polymerase sequences (Delarue *et al.*, 1990). Strictly conserved residues are shown in upper case. Amino acids with >90% conservation are represented in lower case. The properties of non-conserved residues are represented by signs (□ for hydrophobic, ■ for polar, ○ for charged residues and — where there is no conservation of properties). Perfectly conserved properties are represented by three identical signs; two signs or one sign indicate that the properties are shared by >90% or 70–90% of the residues respectively. □■ indicates that all residues at this position are either hydrophobic or polar; ■○ indicates that all residues at this position are either polar or charged.

nucleotide polymerase structures on our model suggests that the nucleotide polymerization site is located at the bottom of the groove in the channel region.

The carboxy-terminal end of the β′-like subunit contains a Cys-rich motif, $CX_2CX_{8-25}CX_2C$, that is important for contacting the amino-terminal region of the β′-like subunit (Yano and Nomura, 1991) and which is therefore likely to be located close to the channel region where the RNA is synthesized. This gives a hint to the explanation for the finding that the largest subunit is involved in contacting the nascent RNA (Gundelfinger, 1983; Riva *et al.*, 1987).

Interaction with DNA

The 2.5 nm wide channel and groove are a suitable size to accommodate a double-stranded DNA helix. An identical location for DNA interaction was proposed for the Klenow fragment of DNA polymerase I (Ollis *et al.*, 1985). The finger-like stalk is likely to be flexible since it was not resolved in the DNA polymerase I crystals. This structure was proposed to act as a clamp preventing the template from dissociating from the enzyme after each polymerization cycle thus allowing the processivity of the reaction.

This DNA-binding groove would hide the probably positively charged residues from the outer surface in keeping with the observation that RNA polymerases do not bind to negatively charged lipids (Darst *et al.*, 1988; Schultz *et al.*, 1990a). The apical region, which interacts with the positively charged lipids, could possibly repel DNA and orient the enzyme so as to allow the DNA to interact with the groove.

We addressed the question of how the DNA would become

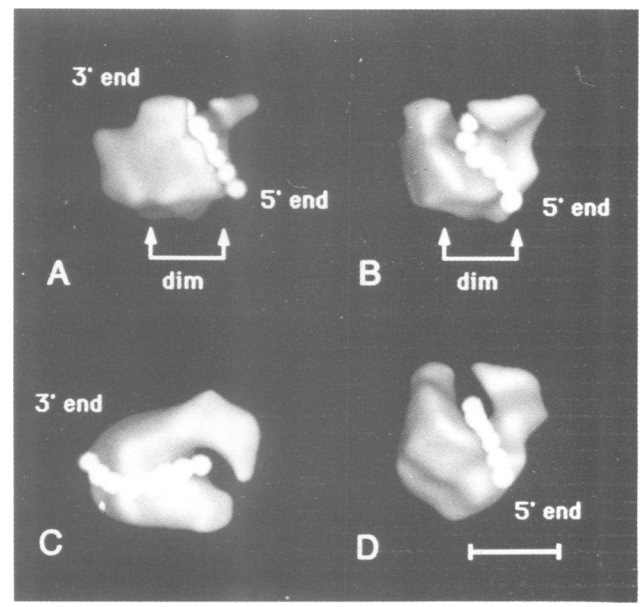

Fig. 5. Schematic representation of the proposed location and orientation of a DNA fragment interacting with RNA polymerase I. The three-dimensional model of the enzyme and its orientation are identical to those in Figure 3. The DNA is symbolized by a beaded filament 1.7 nm in diameter such that one turn of the DNA helix is represented by two beads. The template is placed in the channel and groove whose sizes are suitable to accommodate a double-stranded DNA molecule. The 5′ end of the coding strand is placed close to the dimer interaction surface (represented by arrows) in order to explain the observed interference described in the text. The representation in panel C is meant to illustrate that a 40 bp long upstream DNA fragment can interact with the enzyme surface. The bar represents 6.5 nm.

oriented during transcription. In the proposed orientation of DNA in the Klenow fragment (Ollis *et al.*, 1985) the 5′ end of the template strand is localized in the groove and the 3′ end in the channel. Thus transcription would proceed from groove to channel. The synthetic DNAs used in this study are likely to be specifically bound and oriented in the enzyme since both single-stranded breaks (Dreyer and Hausen, 1976) and single-stranded–double-stranded junctions in 3′ poly(C) extended DNAs (Kadesch and Chamberlin, 1982) have been shown to constitute preferential initiation sites for eukaryotic RNA polymerases. A possible reason for this preferential recognition is that single-stranded–double-stranded junctions mimic the structure of an open complex. Our experiments reveal that the enzyme–DNA interaction interferes with the ability of polymerases to form dimers and subsequently to crystallize. This interference is not due to a non-specific charge effect since DNA fragments of half the length do not perturb crystallization at a 10 times higher concentration. Our experiments further indicate that the dimerization of the enzyme is impaired when the distance from the initiation site to the 5′ end of the template is increased from 11 to 23 bp. This observation suggests that the 5′ end is located near the dimer contact region at the extremity of the groove (Figure 5) and is consistent with the proposed template binding model in DNA polymerase I.

Footprinting experiments performed on *E.coli* RNA polymerase binary complexes and ternary complexes with precise RNA chain lengths are also consistent with this proposed orientation (Metzger *et al.*, 1989). Hydroxyl radical footprints identified two distinct regions in the binary complexes. The recognition domain extends from −14 to −55 bp from the initiation site and shows a partial protection characterized by a 10 bp repeat whereas the melting domain is fully protected over 31 bp from positions −13 to +18. In ternary complexes, the melting domain is almost unchanged until the product reaches 11 nucleotides in length and the protection is reduced to 21 bp when transcription proceeds. The recognition domain disappears immediately after transcription starts. It has been proposed that this differential protection reflects the different accessibility of the DNA molecule when placed in the channel or in the groove and results in an orientation reverse to ours (Darst *et al.*, 1991). However, the authors had to take into account that channel and groove cannot bind 73 bp and that additional DNA has to interact with the enzyme surface.

To be consistent with these footprinting data, our model places the fully protected 31 bp long melting domain in the channel and groove, which are large enough to accommodate ~30 bp (Figure 5). The 42 bp long transient recognition domain needs then to interact with the enzyme surface in order to produce a one-sided protection. Figure 5C illustrates that a 14 nm long DNA stretch might be able to interact with the enzyme. When the upstream DNA exits the channel it has to be bent close to the initiation site. Such a bend, generated only in the presence of σ, has been described at position −3 (Heumann *et al.*, 1988). The *E.coli* enzyme was recently observed to interact preferentially with the apices of supercoiled DNA loops where the radius of curvature is maximal (ten Heggeler-Bordier *et al.*, 1992); this is in keeping with our model. Thus our proposed orientation of the template within RNA polymerase, derived from DNA binding experiments, is consistent with the orientation proposed for DNA polymerase I and with footprinting studies.

Image analysis of RNA polymerase–DNA complexes either crystallized or as single molecules failed to detect any significant difference from the projection map of unbound RNA polymerases. This absence of signal may be due to the poor contrast of DNA which is probably both positively and negatively stained and the poor crystallinity of our arrangements. Observation of frozen hydrated specimens might help to localize more directly the path of DNA in the nucleoprotein complexes.

Materials and methods

RNA polymerase purification

Yeast RNA polymerase I was purified and assayed as previously described (Buhler *et al.*, 1974; Huet *et al.*, 1975). The purified enzyme which showed the characteristic polypeptide content was stored at −80°C at a concentration of 1.4 mg/ml in the presence of 50% glycerol.

Specimen preparation

Lipid–protein interactions and crystallization were performed essentially as described by Schultz *et al.* (1990a). Briefly, 10 μl incubation buffer (30 mM Tris pH 7.6, 20 mM KCl, 5 mM $MgCl_2$, 20% glycerol, 0.15 mM DTT, 0.015 mM EDTA) were placed in a Teflon well and a lipid film was formed at the surface of the droplet by adding 2 μl lipids at a concentration of 0.5 mg/ml in chloroform–hexane (1:1, v/v). The lipids used were oleyl alcohol (Sigma) and cetyl trimethyl ammonium (Aldrich Chemical Co.) at a molar ratio of 4:1. The RNA polymerase preparation was diluted to a concentration of 60 μg/ml in incubation buffer and 5 μl of this suspension was injected into the subphase. After a 1–3 h incubation at 18°C, the lipid layers were transferred on to an electron microscopy grid and negatively stained for 30 s with a 2% uranyl acetate solution.

The DNA templates were formed by hybridization of synthesized oligonucleotides. Absence of free single-stranded DNA was checked by ^{32}P end labelling the oligonucleotides and gel electrophoresis. RNA polymerase diluted to a concentration of 60 μg/ml in incubation buffer was mixed with the DNA at a protein:DNA molar ratio ranging from 1:1 to 100:1. After a 20 min incubation at 37°C, the complexes were allowed to interact with the lipid layers as described above.

Electron microscopy

Micrographs were recorded on a Philips CM12 transmission electron microscope operating at 100 kV on Kodak SO163 plates. Crystalline areas were recorded at a magnification of 43 300×, as calibrated with the 2.3 nm reflection of tobacco mosaic virus, and with a dose of 100–1000 electrons/nm^2. Tilt series were recorded from 58° to 0° with 5–15° increments. A maximum of five images was recorded from a single area. An untilted image was recorded after each series to determine the orientation of the crystal and to assess the absence of information loss in the 2.5–3 nm resolution range as an effect of repeated irradiation.

Image processing

The original micrographs were checked visually for proper stain embedding and by optical diffraction for absence of astigmatism and optimal contrast transfer function. The best micrographs were digitized on the microdensitometer equipment developed by the GSTS (Groupement Scientifique de Télédétection Spatiale) and the SERTIT (Service Régional de Traitement d'Image et de Télédétection, Strasbourg) at 25 μm raster size producing a pixel spacing of 0.56 nm. The image processing was performed using the IMAGIC software kindly provided by M.van Heel (Van Heel and Keegstra, 1981) which was adapted to the SUN-UNIX environment by R.Fritz in our laboratory. Images were stored as 3400 × 2800 points, small ordered areas 512 × 512 pixels in size were selected out and an average unit cell was calculated for each area by the correlation averaging method (Saxton *et al.*, 1984). The tilt parameters were determined from the distortion of the unit cell as analysed from the autocorrelation functions and coded as Euler angles. Untilted images were aligned on the *p2* symmetry axis present in the unit cell and each tilted projection was aligned on the tilted image immediately below it. A preliminary three-dimensional model was then calculated by filtered back projection (Harauz and Ottensmeyer, 1984; Harauz and Van Heel, 1986). The three-dimensional phase origin was further refined by aligning translationally the experimental projections against the out-projections of the model. The *p2* symmetry present in the unit cell is imposed during reconstruction by assigning two symmetric observation angles to each projection. The non-crystallographic symmetry present in the dimer was imposed after real space three-dimensional density correlation of the

two monomeric envelopes (Bricogne, 1976). The rotational and translational matrices were iteratively refined to minimize the density difference between the two envelopes. The surface relief was represented by shaded images using the programs originally developed by Saxton (1985).

Acknowledgements

We thank M.Van Heel for helpful suggestions throughout this work. Support from G.Bricogne for real space three-dimensional correlation was highly appreciated. Constant encouragement from R.D.Kornberg, S.A.Darst and A.M.Edwards is gratefully acknowledged. We thank V.Mallouh and R.Fritz for support in image processing. Oligonucleotides were synthesized by F.Ruffenach and A.Staub. P.S. was a recipient of a Fondation pour la Recherche Médicale fellowship. Part of this work was supported by the EEC Science Program grant no. SC1-CT91-0702.

References

Ahearn,J.M., Bartolomei,M.S., West,M.L., Cisek,L.J. and Corden,J.L. (1987) *J. Biol. Chem.*, **262**, 10695–10705.

Allison,L.A., Moyle,M., Shales,M. and Ingles,C.J. (1985) *Cell*, **42**, 599–610.

Berghöfer,B., Kröckel,L., Körtner,C., Truss,M., Schallenberg,J. and Klein,A. (1988) *Nucleic Acids Res.*, **16**, 8113–8128.

Bréant,B., Huet,J., Sentenac,A. and Fromageot,P. (1983) *J. Biol. Chem.*, **258**, 11968–11973.

Bricogne,D. (1976) *Acta Crystallogr.*, **A32**, 832–847.

Buhler,J.-M., Sentenac,A. and Fromageot,P. (1974) *J. Biol. Chem.*, **249**, 5963–5970.

Chung,Y.J., Sousa,R., Rose,J.P., Lafer,E. and Wang,B.C. (1990) In Papas,T.S. (ed.), *Gene Regulation and AIDS: Transcriptional Activation, Retroviruses, and Pathogenesis*. Portfolio Publishing Co., The Woodlands, TX, pp. 3–7.

Darst,S.A., Ribi,H.O., Pierce,D.W. and Kornberg,R.D. (1988) *J. Mol. Biol.*, **203**, 269–273.

Darst,S.A., Kubalec,E.W. and Kornberg,R.D. (1989) *Nature*, **340**, 730–732.

Darst,S.A., Edwards,A.M., Kubalec,E.W. and Kornberg,R.D. (1991) *Cell*, **66**, 1–20.

Delarue,M., Poch,O., Tordo,N., Moras,D. and Argos,P. (1990) *Protein Engng*, **3**, 461–467.

Dequard-Chablat,M., Riva,M., Carles,C. and Sentenac,A. (1991) *J. Biol. Chem.*, **266**, 15300–15307.

Dreyer,C. and Hausen,P. (1976) *Eur. J. Biochem.*, **70**, 63–74.

Edwards,A.M., Darst,S.A., Feaver,W.J., Thompson,N.E., Burgess,R.R. and Kornberg,R.D. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 2122–2126.

Falkenburg,D., Dworniczak,B., Faust,D.M. and Bautz,E.K.F. (1987) *J. Mol. Biol.*, **195**, 929–937.

Grachev,M.A., Kolocheva,T.I., Lukhtanov,E.A. and Mustaev,A.A. (1987) *Eur. J. Biochem.*, **163**, 113–121.

Grachev,M.A., Lukhtanov,E.A., Mustaev,A.A., Zaychikov,E.F., Abdukayumov,M.N., Rabinov,I.V., Richter,V.I., Skoblov,Y.S. and Chistyakov,P.G. (1989) *Eur. J. Biochem.*, **180**, 577–585.

Gundelfinger,E.D. (1983) *FEBS Lett.*, **157**, 133–138.

Harauz,G. and Ottensmeyer,F.P. (1984) *Ultramicroscopy*, **12**, 309–320.

Harauz,G. and Van Heel,M. (1986) *Optik*, **73**, 146–156.

Heumann,H., Ricchetti,M. and Werel,W. (1988) *EMBO J.*, **7**, 4379–4381.

Huet,J., Buhler,J.M., Sentenac,A. and Fromageot,P. (1975) *Proc. Natl Acad. Sci. USA*, **72**, 3034–3038

Huet,J., Phalente,H., Buttin,G., Sentenac,A. and Fromageot,P. (1982) *EMBO J.*, **1**, 1193–1198.

James,P., Whelen,S. and Hall,B.D. (1991) *J. Biol. Chem.*, **266**, 5616–5624.

Jokerst,R.S., Weeks,J.R., Zehring,W.A. and Greenleaf,A.L. (1989) *Mol. Gen. Genet.*, **215**, 266–275.

Joyce,C.M. (1991) *Curr. Opin. Struct. Biol.*, **1**, 123–129.

Kadesch,T.R. and Chamberlin,M.J. (1982) *J. Biol. Chem.*, **257**, 5286–5295.

Kohlstaedt,L.A., Wang,J., Freidman,J.M., Rice,P.A. and Steitz,T.A. (1992) *Science*, **256**, 1783–1790.

Lebeau,L., Regnier,E., Schultz,P., Wang,J.C., Mioskowski,C. and Oudet,P. (1990) *FEBS Lett.*, **267**, 38–42.

Leffers,H., Gropp,F., Lottspeich,F., Zillig,W. and Garret,R.A. (1989) *J. Mol. Biol.*, **206**, 1–17.

Liljelund,P., Mariotte,S., Buhler,J.M. and Sentenac,A. (1992) *Proc. Natl Acad. Sci. USA.*, **89**, 9302–9305.

Mann,C., Buhler,J.-M., Treich,I. and Sentenac,A. (1987) *Cell*, **48**, 627–637.

Mémet,S., Saurin,W. and Sentenac,A. (1988a) *J. Biol. Chem.*, **263**, 10048–10051.

Mémet,S., Gouy,M., Marck,C., Sentenac,A. and Buhler,J.-M. (1988b) *J. Biol. Chem.*, **263**, 2830–2839.

Metzger,W., Schickor,P. and Heumann,H. (1989) *EMBO J.*, **8**, 2745–2754.

Mosser,G., Ravanat,C., Freyssinet,J.-M. and Brisson,A. (1991) *J. Mol. Biol.*, **217**, 241–245.

Ohme,M., Tanaka,M., Chunwongse,J., Shinozaki,K. and Sugiura,M. (1986) *FEBS Lett.*, **200**, 87–90.

Ohyama,K., Fukuzawa,H., Kohchi,T., Shirai,H., Sano,S., Sano,T., Umesono,K., Shiki,Y., Takeuchi,M., Chang,Z., Aota,S., Inokuchi,H. and Ozeki,H. (1986) *Nature*, **322**, 572–574.

Ollis,D.L., Brick,P., Hamlin,R., Xuong,N.G. and Steitz,T.A. (1985) *Nature*, **313**, 762–766.

Ovchinnikov,Y.A., Monastyrskaya,G.S., Gubanov,V.V., Guryev,S.O., Chertov,O.Y., Modyanov,N.N., Grinkevich,V.A., Makarova,I.A., Marchenko,T.V., Polovnikova,I.N., Lipkin,V.M. and Sverdlov,E.D. (1981) *Eur. J. Biochem.*, **116**, 621–629.

Patel,D.D. and Pickup,D.J. (1989) *J. Virol.*, **63**, 1076–1086.

Puehler,G., Lottspeich,F. and Zillig,W. (1989) *Nucleic Acids Res.*, **17**, 4517–4534.

Ribi,H.O., Reichard,P. and Kornberg,R.D. (1987) *Biochemistry*, **26**, 7974–7979.

Riva,M., Schäffner,A.R., Sentenac,A., Hartmann,G.R., Mustaev,A.A., Zaychikov,E.F. and Grachev,M.A. (1987) *J. Biol. Chem.*, **262**, 14377–14380.

Riva,M., Carles,C., Sentenac,A., Grachev,M.A., Mustaev,A.A. and Zaychikov,E.F. (1990) *J. Biol. Chem.*, **265**, 16498–16503.

Saxton,W.O. (1985) *Ultramicroscopy*, **16**, 387–394.

Saxton,W.O., Baumeister,W. and Hahn,M. (1984) *Ultramicroscopy*, **13**, 57–70.

Schultz,P., Celia,H., Riva,M., Darst,S.A., Colin,P., Kornberg,R.D., Sentenac,A. and Oudet,P. (1990a) *J. Mol. Biol.*, **216**, 353–362.

Schultz,P., Nobelis,P., Colin,P., Louys,M., Huet,J., Sentenac,A. and Oudet,P. (1990b) *Chromosoma*, **99**, 196–204.

Seifarth,W., Petersen,G., Kontermann,R., Riva,M., Huet,J. and Bautz,E.K.F. (1991) *Mol. Gen. Genet.*, **228**, 424–432.

Sentenac,A. (1985) *CRC Crit. Rev. Biochem.*, **18**, 31–90.

Sousa,R. (1991) Ph.D. thesis, University of Pittsburgh.

Sweetzer,D., Nonet,M. and Young,R. (1987) *Proc. Natl Acad. Sci. USA*, **84**, 1192–1196.

Ten Heggeler-Bordier,B., Wahli,W., Adrian,M., Stasiak,A. and Dubochet,J. (1992) *EMBO J.*, **11**, 667–672.

Treich,I., Carles,C., Sentenac,A. and Riva,M. (1992) *Nucleic Acids Res.*, **20**, 4721–4725.

Uzgiris,E.E. and Kornberg,R.D. (1983) *Nature*, **301**, 134–136.

Van Heel,M. and Keegstra,W. (1981) *Ultramicroscopy*, **7**, 113–130.

Woychik,N.A. and Young,R.A. (1990) *Trends Biochem. Sci.*, **15**, 347–351.

Yano,R. and Nomura,M. (1991) *Mol. Cell. Biol.*, **11**, 754–764.

Received on December 14, 1992; revised on April 13, 1993