Structure and Evolution of Prokaryotic and Eukaryotic RNA Polymerases: a Model

DANIELE ARMALÈO

Department of Microbiology, Duke University Medical Center, Durham, North Carolina 27710, U.S.A.

(Received 1 August 1986, and in revised form 2 February 1987)

A comparative overview of the subunit taxonomy and sequences of eukaryotic and prokaryotic RNA polymerases indicates the presence of a core structure conserved between both sets of enzymes. The differentiation between prokaryotic and eukaryotic polymerases is ascribed to domains and subunits peripheral to the largely conserved central structure. Possible subunit and domain functions are outlined. The core's flexible shape is largely determined by the elongated architecture of the two largest subunits, which can be oriented along the DNA axis with their bulkier amino-terminal head regions looking towards the 3' end of the gene to be transcribed and their more slender carboxyl-terminal domains at the tail end of the enzyme. The two largest prokaryotic subunits appear originally derived from a single gene.

1. Introduction

The eukaryotic nuclear RNA polymerases are a family of related proteins with three subgroups, each having distinct structural and functional properties. Class I (or A) enzymes synthesize 18S and 28S ribosomal RNA precursors, class II (or B) messenger RNA precursors, and class III (or C) various small RNAs, transfer RNA and 5S RNA (Roeder, 1976). The purified native proteins (500 to 700 kDa) are composed of 10 to 15 different subunits whose individual sizes range from 10 to 240 kDa. Due to the complexity of these enzymes and to the failure of in vitro reconstitution attempts, very little is known about the roles and topography of their component subunits (Lewis & Burgess, 1982; Paule, 1981). Prokaryotes, on the other hand, transcribe all their genes with a single enzyme, simpler and better understood than its eukaryotic counterparts (Chamberlin, 1982).

In this paper a model is described of the structure and evolution of prokaryotic and eukaryotic RNA polymerases, based on an analysis of subunit molecular weights and sequences. Within the limitations of a simplifying scheme, the model provides a frame for a more targeted dissection of polymerase structure and function.

2. Subunit Size Conservation vs Change in Polymerase Evolution

In Fig. 1(a), eukaryotic RNA polymerases are grouped into their three constituent classes. A prokaryotic and a viral RNA polymerase are also included. Each symbol represents an individual enzyme subunit whose position along the abscissa corresponds to its molecular weight and, along the vertical dimension, to its parent polymerase (listed on the left). Such a comparative representation, warranted by
the relatedness of all eukaryotic RNA polymerases (Armaleo & Gross, 1985b; Ingles et al., 1984; Weeks et al., 1982; Huet et al., 1982b; Paule, 1981), shows that most subunits cluster in two size ranges, one between 10 and 50 kDa, and the other between 120 and 240 kDa, regardless of enzyme type or source. Subclusters can also be recognized within the two major ranges (Fig. 1(b)). This overall similarity in the architecture of eukaryotic RNA polymerases is still discernible after thousands of millions of years of divergence (Woese & Fox, 1977) over a background of substantial molecular weight variability.

Assessing in some detail the relative significance of subunit size conservation vs change in polymerase evolution is at the basis of the model discussed below. The assumption is that, in such a multi-subunit system, the degree of molecular weight drift among comparable polypeptides across many species and polymerase classes reflects the degree of divergence in sequence and function of these subunits. It is on this broad basis that the polypeptides listed in Fig. 1(a) will be grouped into "conserved" or "variable" clusters (Fig. 1(b)). The boundaries between size clusters, however, are not sharp. Therefore, whereas the majority of subunits characterizing a cluster will be considered functionally related, some may in fact be "spill-overs" from neighboring groups with different functions (overlaps in Fig. 1(b)).

Proceeding from the smallest to the largest subunits, one finds a first major 10–30 kDa cluster (A), which includes most highly conserved small subunits common to at least two polymerase classes in any given organism (Lewis & Burgess, 1982). Within the 10–30 kDa group, the histogram in Fig. 1(b) hints at several subdivisions. A second clustering of conserved subunits (●) is seen around 35–45 kDa, and is distinct from the first (Fig. 1(b)). (For the size-conserved subunits in both clusters, sequence conservation has been proven immunologically, by SDS polyacrylamide gel electrophoresis, peptide mapping, and two-dimension electrophoresis; see Armaleo & Gross (1985b) and Lewis & Burgess (1982, and references therein). Proceeding through the region between 50 and 120 kDa, one finds a few widely scattered polymerase I and III subunits (O), but no polymerase II subunits. Finally, the second largest subunits (■) show limited variability (120–150 kDa), whereas the largest subunits (□) cover a wide range of sizes (150–240 kDa).

Fig. 1. RNA polymerase subunit size distribution. (a) The class and the organism of origin of each enzyme are listed on the left. The abscissa measures molecular weight. Subunits belonging to the same enzyme fall on the same ideal line (parallel to the abscissa) and are represented by different symbols according to the classification described in the text: (△) 10–30 kDa; (●) 35–45 kDa; (O) 50–100 kDa; (■) second largest subunit; (□) largest subunit (proteolyzed subunits are in parenthesis). The data were compiled from the following references: E. coli (Chamberlain, 1982); yeast (Lewis & Burgess, 1982; Buhler et al., 1976, 1980; Dezelee et al., 1976; Valenzuela et al., 1976); Neurospora (Armaleo & Gross, 1985a); Aspergillus (Stunnenberg et al., 1979); Physarum (Smith & Braun, 1978); Agaricus (Vaisius & Horgen, 1979); Acanthamoeba (D'Alessio et al., 1979a, b; Spindler et al., 1978a, b); soybean (Guilfoyle & Jendrisak, 1978); wheat (Jendrisak, 1981); cauliflower (Guilfoyle, 1980); Drosophila (Weeks et al., 1982); Artemia (Huet et al., 1982b); Xenopus (Sklar et al., 1975); rat (Matsui et al., 1976); mouse (Sklar, et al., 1975); calf (Dahmus, 1981a, b; Hodo & Blatti, 1977); human (Jaehning et al., 1977); pox (Baroudy & Moss, 1980). (b) Histogram representing the cumulative number of eukaryotic RNA polymerase subunits displayed in (a) as a function of molecular weight. The horizontal lines and the associated numbers mark the molecular weight ranges of the various subunit populations.
The apparently high variability of the largest subunits can be broken down into three discontinuous ranges related to polymerase class, clustering around 155 kDa in type III polymerases, 195 kDa in type I polymerases, and 220 kDa in type II polymerases. A clue to interpret this pattern may be provided by the largest subunits of type II enzymes. These polypeptides form on the enzyme surface a structurally and functionally independent 40–50 kDa domain (Armaleo & Gross, 1985a), as suggested by the frequent proteolytic loss of a 40 kDa segment from that subunit during enzyme isolation (Guilfoyle et al., 1984) without major alteration in the in vitro catalytic properties (Dezelee et al., 1976). The size of the proteolyzed subunit is reduced to about 180 kDa, a value closer to that of the largest subunits of polymerases I and III. This structural peculiarity of type II polymerases suggests that most of the size divergence among the largest subunits of all three enzyme classes may be due to different domains present at one end of an otherwise “constant” core section of about 160 kDa. In type I polymerases, the extra domains would be less sensitive to proteolysis (their largest subunits are usually recovered intact), whereas type III polymerases would have little or no extension (the size of their largest subunit is close to 160 kDa). Based on indirect evidence, it was suggested that the proteolysis-sensitive extension in polymerase II is carboxy-terminal (Armaleo & Gross, 1985a; Armaleo, 1984), an orientation established by recent sequencing data (see section 3). A similar organization might be anticipated also for the largest subunits of type I polymerases.

In summary, conservation is mostly seen among the 10–30 kDa, the 35–45 kDa, the second largest 140 kDa subunits, and the presumably invariant portions (about 160 kDa) of the largest subunits. Divergence, on the other hand, appears primarily confined to the 50–100 kDa polypeptides and to the variable carboxy-terminal extensions of the largest subunits.

3. Related Subunits and Cores in Prokaryotic and Eukaryotic Enzymes

A comparison with E. coli RNA polymerase, representative of most prokaryotic RNA polymerases (Chamberlin, 1982), shows that the bacterial enzyme has no subunits in the 10–30 kDa range, that the molecular weight of its α subunit falls in the center of the eukaryotic 35–45 kDa cluster, that the regulatory σ and nusA proteins fall in the 50–100 kDa span, and that β and β' are located at the lower end (150–160 kDa) of the size range of the largest eukaryotic subunits (Fig. 1(a)). Carrying the “molecular weight conservation vs relatedness” argument across the prokaryotic/eukaryotic bridge, one could expect relatedness between α and some of the 35–45 kDa subunits, between β and the second largest eukaryotic subunits, and between β' and the presumed invariant part of the largest eukaryotic subunits.

Direct and indirect experimental evidence for common ancestry does in fact exist.

(1) By way of the cloned DNA sequences, homologies have been found between the E. coli β' subunit and the amino-terminal regions of the largest polymerase II subunits of Drosophila (Biggs et al., 1985), yeast (Allison et al., 1985) and mouse (Corden et al., 1985), the largest polymerase III subunit of yeast (Allison et al., 1985), and the 147 kDa subunit of vaccinia virus polymerase (Broyles & Moss,
The 10-30 kDa polypeptides are thus left as a distinctly eukaryotic subunit population, with no obvious prokaryotic counterpart (except for some archaeabacteria, as discussed below).

The correlations described above suggest (Fig. 2) that today's prokaryotic and eukaryotic cores derive from a common $\beta\beta'\alpha_2$-like ancestor, and that a unique set of prokaryotic and eukaryotic enzymes. For eukaryotic polymerases, the letter e represents the 10-30 kDa, and $\rho$ the various "regulatory" subunits. The carboxy-terminal extensions of the largest eukaryotic subunits are hatched. The 10-30 kDa subunits are considered part of the eukaryotic core as most are common to all three polymerase classes. The contacts, shapes and relative dimensions of the subunits are arbitrary. See section six for a description of the roles of the eukaryotic subunits.

1986); the same data indicate that the carboxy-terminal region of the largest polymerase II subunit is in fact the proteolysis-sensitive domain and that it consists of a unique series of [TyrSerProThrSerProSer] repeats, about 25 in yeast and 50 in mouse, with no known counterpart in *E. coli*; (the establishment of homology between $\beta$ and the second largest eukaryotic subunits awaits the cloning of the eukaryotic sequences).

(2) The $\beta$ and $\beta'$ *E. coli* subunits, like the two largest polymerase II polypeptides, are involved in nucleic acid binding and basic polymerization functions (Carroll & Stollar, 1983; Horikoshi *et al.*, 1983; Huet *et al.*, 1982a; Yura & Ishihama, 1979).

(3) In several cases, a stoichiometry close to two 35-45 kDa subunits per enzyme molecule can be deduced from the SDS gel staining patterns of eukaryotic RNA polymerases (Armaleo & Gross, 1985a; D’Alessio *et al.*, 1979b; Stunnenberg *et al.*, 1979; Vaisius & Horgen, 1979; Spindler *et al.*, 1978a, b; Hager *et al.*, 1977; Jaehning *et al.*, 1977; Valenzuela *et al.*, 1976) and of pox virus RNA polymerase (Barouby & Moss, 1980), in striking correspondence with the known stoichiometry of the $\alpha$ subunit of *E. coli*; (also in this case, the establishment of homology between $\alpha$ and some of the 35-45 kDa subunits awaits the cloning of the eukaryotic sequences).

The 10-30 kDa polypeptides are thus left as a distinctly eukaryotic subunit population, with no obvious prokaryotic counterpart (except for some archaeabacteria, as discussed below).
of novel 10-30 kDa polypeptides (all grouped here under the symbol $e$), as well as carboxy-terminal extensions to the largest subunit of type I and II polymerases, are characteristic of eukaryotic enzymes.

4. Orientation of the Largest Polymerase Subunits on the Transcribed Gene

The data available in the literature allow some detailed proposals to be made on the orientation of the two large polymerase subunits with respect to each other and to the transcribed gene. A primary observation is that the $\beta$ and $\beta'$ subunits of *E. coli* RNA polymerase have a very elongated shape, as indicated by physical studies of the enzyme in solution (Meisenberger *et al.*, 1980; Stoeckel *et al.*, 1980), by experiments measuring the distance traveled by the 5' end of the nascent RNA before leaving the surface of the enzyme (Hanna & Meares, 1983) and, more directly, by mapping the DNA contacts made by the enzyme as it binds to various promoters (Chenchick *et al.*, 1981; Siebenlist *et al.*, 1980; Schmitz and Galas, 1979). The latter studies indicate that $\beta$ and $\beta'$ can each touch the DNA over a distance of more than 200 Å (65 base pairs). From neutron- and small angle X-ray scattering studies (Meisenberger *et al.*, 1980; Stoeckel *et al.*, 1980), $\beta$ and $\beta'$ are each 230 to 240 Å long. These data imply an average width to length ratio of approximately 1:6 for these 150-160 kDa proteins. The main assumption on which the conclusions reached below are based is that, in the elongated polymerase subunits, the overall polarity of the polypeptide chain will tend to follow the longer axis, with the N and C termini near opposite ends. This particular disposition of the termini is observed, for example, in elongated molecules like collagen, myosin, fibrinogen, antibodies, and transcription factor IIIA, and is due to the way many multi-functional proteins appear to be constructed, by sequential linking of individual domains (Rossmann & Argos, 1981).

With these premises, the amino-carboxy orientation of the *E. coli* $\beta'$ and the largest eukaryotic subunit with respect to the direction of transcription can be deduced from the following data. As described, the largest subunits of yeast RNA polymerases II and III and the *E. coli* $\beta'$ subunit are homologous over most of their length. They also share sequence homology, confined to a 150 amino-acid region, with *E. coli* DNA polymerase I and T7 DNA polymerase (Allison *et al.*, 1985). Within the three-dimensional structure of *E. coli* DNA polymerase I (Ollis *et al.*, 1985a), this 150 amino-acid region comprises three $\alpha$-helices involved in binding DNA, and is in close proximity of the 3' end of the primer strand, i.e. the active site. The observed sequence conservation and the probably stricter conservation of structure (Ollis *et al.*, 1985b; Keim *et al.*, 1981) suggest that also in prokaryotic and eukaryotic RNA polymerases the three $\alpha$-helices are located in the general vicinity of the active site. Their position on the linearized amino-acid sequences of the *E. coli* $\beta'$ and the largest polymerase II and III subunits (Allison *et al.*, 1985) places the active site region about 30% away from the amino-terminus; when *E. coli* RNA polymerase binds to a specific promoter (Chenchick, *et al.*, 1981), the transcription start site (which also indicates the position of the active site) lies about 30% away from the "front" end of the $\beta'$ subunit, defined as the end which protrudes into the gene to be transcribed. Thus, the elongated $\beta'$ subunit and its eukaryotic homologues
STRUCTURE AND EVOLUTION OF RNA POLYMERASE

FIG. 3. Alignment of short sequences thought to be involved in nucleotide binding in DNA and RNA polymerases. Seven-residue sequences characteristic of the active site regions of *E. coli* DNA polymerase I and T7 DNA polymerase were used to seek possibly homologous sequences marking the active site regions in the \( \beta \) subunit of *E. coli* RNA polymerase. If present, such sequences are expected in the amino-terminal region of \( \beta \) if it is oriented in parallel to \( \beta' \) (see text for further discussion). Heptapeptide 352-358 of *E. coli* DNA polymerase I, which is in close contact with the bound deoxynucleotide (Ollis et al., 1985a), and heptapeptide 232-238 of the highly homologous T7 DNA polymerase (Ollis et al., 1985b), were matched by computer with the entire sequence of *E. coli* \( \beta \). Scanning residues 1-1200 with either heptapeptide, the best match found was with sequence 186-192; scanning residues 193-1342 with either heptapeptide the best match was with sequence 251-257. It should be noted that the indicated matches were chosen among the approximately 1200 seven-residue stretches which make up the subunit sequence scanned. The ALIGN program (with an upper limit of 1200 residues per run) was used with the unitary matrix (Dayhoff et al., 1983). The \( \beta' \) sequence shown was found by scanning residues 1-1200 of \( \beta' \) with heptapeptide 232-238 of T7 DNA polymerase.

appear oriented with the amino-terminus at the head and the carboxy-terminus at the tail of the transcribing enzyme. In eukaryotic RNA polymerases, this orientation places the carboxy-terminal extension domains at the trailing edge of the enzyme, an area of important regulatory interactions with DNA and with other proteins. In *E. coli*, the tail of the promoter-bound enzyme is thought to interact with a number of regulatory proteins, like CAP or \( \lambda \) repressor (Hochschild et al., 1983; Siebenlist et al., 1980; Schmitz & Galas, 1979).

The \( \beta \) subunit and, by analogy, the second largest eukaryotic subunit, can be oriented on the basis of the following considerations. In *E. coli*, \( \beta \) and \( \beta' \) share a limited amount of sequence homology, which suggests that a single gene was at the origin of both subunits.† Gene duplication and divergence would have led from a dimeric complex in which the two components were very similar or identical (homodimer) to today’s heterodimeric \( \beta \) and \( \beta' \) polypeptides. The orientation of \( \beta \) with respect to \( \beta' \) should reflect the original orientation in the homodimer. For two similar or identical proteins binding to DNA and to each other, two major types of symmetry can be expected. The first relates the two subunits through a rotation around an axis parallel to the DNA and does not change their amino-carboxy polarity (‘parallel’ orientation). The second involves a rotation around an axis perpendicular to the DNA and would invert the amino-carboxy polarity of one

†The sequences of \( \beta \) and \( \beta' \) were compared using the ALIGN program (Dayhoff et al., 1983). The overall alignment score obtained for the 1200 amino-terminal residues was 3.43. This score is the number of standard deviations by which the maximum match score for the two real sequences exceeds the average maximum match score for 100 random permutations of the two sequences. The value 3.43 corresponds to a probability of \( 3 \times 10^{-6} \) for the random generation of an homology as high as, or higher than, that observed.
subunit with respect to the other ("anti-parallel" orientation). The former alternative is supported by two independent lines of evidence. First, small angle X-ray-scattering studies indicate that the *E. coli* polymerase core has an elongated trigonal shape mostly determined by the two attached β and β' subunits, with the center of mass closer to one end of the molecule (Meisenberger et al., 1980). This shape (resembling two adjacent sections of a pie) corresponds best to a parallel orientation of two originally very similar large subunits and is consistent with the disposition of cysteines along the polypeptide chain, as described below. If the orientation were anti-parallel one would expect a more homogeneous overall mass distribution.

Second, a short *E. coli* DNA polymerase I sequence which constitutes part of the binding site for nucleotide monophosphates (Ollis et al., 1985a) resembles two short sequences located 60 residues apart in the amino-terminal region of the β subunit (Fig. 3). A similar sequence is also found in T7 DNA polymerase (Fig. 3), an enzyme highly homologous to *E. coli* DNA polymerase I (Ollis et al., 1985b). Considering that β binds nucleotide triphosphates (Yura & Ishihama, 1979) at two adjacent and catalytically equivalent sites (Panka & Dennis, 1985), it seems plausible that the two neighbouring β sequences may contribute each to one nucleotide triphosphate binding site, thereby localizing the catalytically important area of β in proximity of the amino-terminus and orienting the subunit in parallel to β'. (Figure 3 also shows that a possible related sequence, discussed below, is found in single copy in the β' subunit.)

5. Functional and Structural Map of the Two Large RNA Polymerase Subunits

The β subunit can contact the LacUV-5 promoter between bases −36 and +30 (the transcription start site being +1), and β' between bases −47 and +18 (Chenchick et al., 1981). This staggered disposition of the DNA contacts suggests a staggered disposition of the two subunits, with β being about 12 bases “ahead” of β', i.e. one and a quarter helix turns. In an ancestral homodimer, two parallel subunits could have made equivalent contacts with the phosphate backbone by being 12 base-pairs displaced and rotating one with respect to the other 80–90° around the DNA axis. The “twelve base-pair motif” appears also as another important aspect of RNA polymerase function, namely as the most commonly estimated length of the DNA “bubble” within which RNA synthesis occurs (von Hippel et al., 1984; Hanna & Meares, 1983; Gamper & Hearst, 1982; Reisbig et al., 1979; Siebenlist, 1979). Is it a coincidence that the displacement of the two large subunits along the DNA corresponds to the approximate extension of the RNA–DNA hybrid in the active site of the enzyme? If not coincidental, the twelve base-pair displacement of β and β' might in fact reflect the fundamental symmetry requirements of an ancestral dimer unwinding a twelve base-pair stretch of DNA to allow synthesis of RNA.

A scheme exemplifying in this light the hypothetical structure–function relationships of β and β' is presented in Fig. 4. The rectangles represent β and β' 20% displaced with respect to one another (the equivalent of about 12 base pairs, since each subunit can cover about 60 base pairs). Indicated by the triangles are the positions of the putative nucleotide binding sequences: the two on β (▲) presumably
Fig. 4. Scheme of the hypothesized relationships between $\beta$, $\beta'$ and the substrate. The $\beta$ and $\beta'$ subunits are represented by rectangles whose lengths were made proportional to the primary sequence lengths. N and C, amino- and carboxy-terminal regions. The location of the putative nucleotide binding sequences is indicated by triangles. (▲) catalytic nucleotide binding; (△) effector nucleotide binding. The regions presumably involved in the binding of double-stranded nucleic acid are marked by the double-headed arrows. The region on $\beta'$ found homologous to the DNA binding domains of *E. coli* DNA polymerase I (Allison *et al.*, 1985) is marked by the bracket. The dots indicate the positions (within ten amino-acid residues) of cysteines. Nucleic acids (——DNA; ——RNA) are drawn so as to emphasize the primary interactions with each subunit. The direction of transcription is from right to left.

contributing to the binding of the catalytic triphosphates; the one of $\beta'$ (△) corresponding to residues 172–180 in Fig. 3 and marking a possible effector binding site (Kingston *et al.*, 1981, have shown, for instance, that *E. coli* RNA polymerase binds ppGpp as a modulator of activity). Upstream of the putative nucleotide binding regions, the double stranded nucleic acid binding domains are also located in corresponding segments of the two subunits, $\beta'$ being involved in DNA binding as the available experimental evidence indicates (Allison *et al.*, 1985; Yura & Ishihama, 1979), and $\beta$ being the major contributor to the catalytic site (Yura & Ishihama, 1979) and probably to the RNA-DNA hybrid binding region. The 5' end of the nascent RNA (— — —) has been shown to follow the boundary between $\beta$ and $\beta'$ (Hanna & Meares, 1983). However, as the nascent end first leaves the RNA-DNA hybrid, it appears closer to $\beta$ (Hanna & Meares, 1983), reinforcing the notion that $\beta$ is in fact the major contributor to the hybrid binding domain. The DNA helix interacts with both large subunits outside the drawn regions (Chenchick *et al.*, 1981). The cysteine residues of both $\beta$ and $\beta'$, also mapped in Fig. 4 (●), cluster in the amino-terminal two thirds of the molecule, leaving a cysteine-free stretch of 400 to 500 amino-acids towards the carboxy terminus of both polypeptides. The sequences of the largest subunits of yeast polymerases II and III (Allison *et al.*, 1985) also show a similar overall arrangement of cysteines (not drawn). No inter-chain disulfide bonds have been observed in *E. coli* or in eukaryotic RNA polymerases, indicating that most cysteines will be involved in intra-chain structural stabilization. The
distribution of the known disulfide bridges among a wide variety of proteins (Thornton, 1981) indicates that most disulfide-linked cysteines are within the same domain and separated by less than 30 residues. They are rarely more than 150 residues apart, a fact probably related to the maximum size of individual domains in large proteins (Freedman & Hillson, 1980). Each of the large polymerase subunits might therefore fold as a succession of domains roughly colinear with the polypeptide chain, compacted by disulfide bonds towards the amino-terminus and more extended and flexible in the cysteine-free carboxy-terminal regions.

This pattern can be related to the overall structural and functional features of the polymerase discussed in the preceding section. It is likely that the bulkier area of the molecule (as defined by small-angle X-ray scattering, see Meisenberger et al., 1980) comprises the amino-terminal halves of the two large subunits and is involved in the central catalytic functions of helix unwinding, rewinding, and of RNA synthesis, which require the largest transfers of free energy. The more slender and flexible carboxy-terminal region (with the additional extension in eukaryotes) would be a tail-like area amenable to regulatory interactions with specific DNA and protein sequences. It may also remain bound to the DNA even if the front end of the enzyme were displaced by the binding of repressor-like molecules, as observed by Schmitz & Galaz (1979). Alternatively, during elongation, the front end may be bound to the template while the tail domains are not, an interpretation of the footprinting data of Carpousis & Gralla (1985). RNA polymerase may thus display some overall flexibility around a central "hinge" region: the whole elongated enzyme can contact the DNA in an open promoter complex, or either the front or the rear domains are displaced from the DNA. A number of regulatory factors, both positive and negative, may mediate such conformational changes. No attempt was made here to define the topography of the \( \alpha_2 \) dimer or of \( \sigma \) in relation to \( \beta, \beta' \) and the DNA, given the limited data available (Chenchick et al., 1981; Meisenberger et al., 1980; Simpson, 1979; Hillel & Wu, 1977).

It is assumed that throughout evolution the basic structural and functional features described above have been largely conserved between prokaryotic and eukaryotic RNA polymerases.

6. Possible Roles of the Various Sets of Eukaryotic Subunits

For eukaryotic enzymes, the evidence discussed in the preceding sections can provide the following functional and structural information on the various sets of subunits.

(1) As described, the central nucleic acid binding and polymerization functions can be assigned largely to the \( \beta\beta'\alpha_2 \)-type subunit complex. In a eukaryotic enzyme, this complex would include the invariant part of the largest subunit, the second largest subunit, and two copies of a 35–45 kDa subunit.

(2) The highly conserved 10–30 kDa subunits are a dominant characteristic of eukaryotic polymerases, while nucleosomes and other highly conserved chromatin components are a hallmark of the eukaryotic template. The parallel evolution and conservation of two sets of proteins involved in the same metabolic process (i.e.
transcription) suggests specific interactions between them. Several of the small 10-30 kDa eukaryotic subunits may thus bind to chromatin components and induce the fundamental alterations necessary for the enzyme to interact directly with DNA. The coexistence of the small subunits with a structurally and functionally autonomous ββ'α2-like structure further suggests that most are probably located on the outer surface of the polymerase molecule. This suggestion is reinforced by the likely occurrence of uncontrolled dissociation of 10-30 kDa subunits from the enzyme during isolation, leading all the way from variable yields of some of the small subunits to possible cases of severe depletion (see type I polymerases in rat and calf, Fig. 1(a)).

(3) The main features distinguishing one polymerase class from another would not be evenly distributed throughout the structure of each enzyme, but would rather depend on individual protein domains evolving either as carboxy-terminal extensions of the largest core subunit or as individual 50-100 kDa subunits (grouped under ρ in Fig. 2). Due to their presumed roles in defining the individuality of each polymerase class, the functions of these domains and subunits can be broadly classified as "regulatory". Somewhat paradoxically, the absence of 50-100 kDa subunits in purified type II polymerases might in fact reflect the very large number of regulatory proteins binding reversibly to polymerase II and modulating transcription of a wide variety of genes. Unlike some of their counterparts stably bound to the less versatile polymerases I and III, most or all polymerase II regulatory subunits would be lost upon enzyme purification due to their reversible binding and low concentration, varying with cell type and metabolic state. Some loss of regulatory components is of course expected also with type I and III enzymes, since no isolated eukaryotic RNA polymerase is able to recognize specific initiation sites in vitro without addition of cellular extracts or purified transcription factors. For instance, the amount of a polypeptide of approximately 65 kDa shows considerable variation among type I polymerases isolated from various sources. It was found in Aspergillus (Stunnenberg et al., 1979) and not in yeast (Huet et al., 1975), it was present in sub-stoichiometric amounts in Neurospora (Armaleo & Gross, 1985b) and its presence or absence distinguished two populations of polymerase I in both rat and calf (Matsui et al., 1976; Gissinger & Chambon, 1975).

7. The Case of Sulfur-dependent Bacteria

Prokaryotes have been subdivided into Eubacteria, which include most of the classical prokaryotes, and Archaeabacteria, which comprise the methanogens (Woese & Gupta, 1981; Woese & Fox, 1977). Recently and with some controversy, sulfur-dependent bacteria living in thermal springs at temperatures up to 90°C and pH as low as 2 have been grouped into a new kingdom, that of Eocyta, distinct from Archaeabacteria (Pace et al., 1986; Lake et al., 1984, 1985). These subdivisions stem from the discovery of a number of profound biochemical differences among prokaryotes. Of concern here is a major discrepancy in subunit composition between eubacterial RNA polymerases and those isolated from representatives of several genera of sulfur-dependent bacteria and methanogens (Huet et al., 1983; Zillig et
Most of these non-eubacterial enzymes possess, besides two large (100-130 kDa) and one to three 35-45 kDa subunits which comprise the presumed $\beta\beta'\alpha_2$-like core, also three to six subunits in the 10-30 kDa range (Zillig & Stetter, 1980). In view of the proposal made here that most of the 10-30 kDa subunits in eukaryotes functionally co-evolved with nuclear chromatin, the presence of small subunits in some archaeabacterial polymerases may suggest that their templates are extensively complexed with protein; a stabilizer, perhaps, in the face of the extreme habitats colonized by these prokaryotes. In eukaryotes, evolution and differentiation of chromatin and RNA polymerases would have been influenced by more complex variables than the need to stabilize the chromosome against heat or low pH.

It should be noted that these speculations on the evolution of nucleo-protein complexes do not imply a choice, by themselves, between the various proposals on the relationships of the primordial eukaryotic and prokaryotic lineages (Pace et al., 1986; Lake et al., 1985).

8. Concluding Remarks

A model of RNA polymerase structure and evolution is described which suggests (1) the existence of a basic $\beta\beta'\alpha_2$-like design common to eukaryotic and prokaryotic RNA polymerases, comprising a bulkier front region mostly involved in catalysis flexibly connected to a more slender tail largely involved in regulation, (2) that within this basic design, two originally similar large subunits (both looking with the amino-terminal region towards the 3' end of the transcribed gene) diverged one towards binding the RNA-DNA hybrid and catalytic nucleotides, and the other towards binding DNA and effector nucleotides, and (3) that the evolution differentiating bacterial and eukaryotic enzymes essentially concerned a set of domains and subunits peripheral to the underlying $\beta\beta'\alpha_2$-like structure. These are (a) a group of largely conserved 10-30 kDa "chromatin-unraveling" subunits on the surface of the molecule, found in all three eukaryotic classes, (b) class-specific carboxy-terminal extensions of the largest subunits (at the tail end of the enzyme), and (c) the variable, class-dependent set of 50-100 kDa subunits with "regulatory" functions. Although the $\beta\beta'\alpha_2$-like structure is presumed to have undergone relatively limited change, the real extent of such change can only be revealed by gene-sequencing. In general, more conservation is expected in the front regions involved in catalysis than in the regulatory areas towards the back.

As described in the preceding section, the model suggests that the chromosome of sulfur-dependent bacteria and methanogens is extensively complexed with proteins. Another suggestion is that the unicellular algae Dinoflagellates, as their nucleo-protein is organized very differently from that of the other eukaryotes (Rizzo & Burghardt, 1982; Li Jing-Yan, 1984), should have RNA polymerases differing from other eukaryotic enzymes with regard to the 10-30 kDa subunits. No RNA polymerases have yet been purified from these organisms, although an $\alpha$-amanitin sensitive RNA polymerase activity has been observed in isolated Dinoflagellate nuclei (Rizzo, 1979).

In conclusion, I think that the message conveyed by the more than 300 subunits in Fig. 1 and by the observed sequence homologies, however blurred by the large
evolutionary distances it bridges, allows to unify a growing number of scattered data into a picture of some coherence and amenable to experimental testing.

I would like to thank Arno Greenleaf, David Price, Stephen Johnston and Mike Ostrowski for useful criticisms and David Pickup for help with the ALIGN program.

**Note Added In Proof**

Through their recent cloning of the gene for the second largest subunit of yeast RNA polymerase II, Sweetser et al. (1987) arrive at some of the same conclusions presented here on the homologies between eukaryotic and prokaryotic RNA polymerases. However, the authors’ implication that a GTP/GDP-binding sequence and a DNA-binding “zinc finger” domain, both located near the C-terminus of the subunit, are involved in catalysis, is not consistent with the known roles of such sequences. They are conserved in a wide variety of proteins and in no case have been involved directly in nucleotide polymerization, but are rather associated with regulatory functions (la Cour et al., 1985; Berg, 1986; Chowdhury et al., 1987). The latter notion is consistent with the regulatory roles ascribed here to the C-terminal regions of both large RNA polymerase subunits.

**REFERENCES**


