The Increase in the Number of Subunits in Eukaryotic RNA Polymerase III Relative to RNA Polymerase II Is due to the Permanent Recruitment of General Transcription Factors

Robert Carter1,2 and Guy Drouin*1,2
1Département de biologie Université d’Ottawa, Ottawa, Ontario, Canada
2Centre de recherche avancée en génomique environnementale, Université d’Ottawa, Ottawa, Ontario, Canada

*Corresponding author: E-mail: gdrouin@science.uottawa.ca.
Associate editor: Claudia Schmidt-Dannert

Abstract

The number of subunits of RNA polymerases (RNAPs) increases during evolution from 5 in eubacteria to 12 in archaea. In eukaryotes, which have at least three RNAPs, the number of subunits has expanded from 12 in RNA polymerase II (RNAPII) to 14 in RNA polymerase I (RNAPI) and to 17 in RNA polymerase III (RNAPIII). It was recently demonstrated that the two additional subunits found in RNAPI relative to RNAPII are homologous to TFIIF, a dimeric general transcription factor of RNAPII. Here, we extend this finding by demonstrating that four of the five RNAPIII-specific subunits are also homologous to transcription factors of RNAPII. We use the available evidence to propose an evolutionary history of the eukaryotic RNAPs and argue that the increases in the number of subunits that occurred in RNAPs I and III are due to the permanent recruitment of preexisting transcription factors.

Key words: RNA polymerase, subunits, transcription factors, eukaryotic.

Introduction

RNA polymerases (RNAPs) are responsible for all cellular transcription. In prokaryotes, a single RNAP handles the full transcription burden of the cell, whereas transcription in eukaryotes is partitioned between at least three polymerases that are common to all eukaryotes. RNA polymerase I (RNAPI) transcribes ribosomal RNA (rRNA); RNA polymerase II (RNAPII) transcribes protein-coding genes and some microRNAs, and RNA polymerase III (RNAPIII) transcribes several small RNAs, including 5S rRNA, transfer RNA, and some microRNAs.

All cellular RNAPs are multisubunit protein complexes that share a core set of homologous subunits across the three domains of life (Best et al. 2004). The eubacterial RNAP is the simplest of the cellular RNAPs. It consists of the active site–forming β′ and β subunits, an α homodimer, and an ζ subunit (Zhang et al. 1999). In archaea, the core bacterial polymerase homologs are retained, but additional subunits have been recruited. It is composed of the following 12 subunits, named according to the scheme used by Korkhin et al. (2009): Rpo1 (homologous to β′); Rpo2 (homologous to β); Rpo3 (homologous to ζ); Rpo4, Rpo5, and Rpo6 (homologous to ζ); Rpo7, Rpo8, Rpo10, and Rpo11 (homologous to ζ); and Rpo12 and Rpo13 (Korkhin et al. 2009). The eukaryotic RNAPs are a further elaboration of the archaeal RNAPs. Eukaryotic RNAPs have homologs of all 12 archaeal RNAP subunits except Rpo13 (Korkhin et al. 2009), and each of the eukaryotic RNAPs additionally contains a homolog of the archaeal general transcription factor (GTF) TFS/M except that these homologs are permanent subunits in the eukaryotic RNAPs. RNAPII is thus composed of 12 subunits, 11 that are homologous to archaeal polymerase subunits (RP1 through RP8 and RP10 through RP12) and 1 (RP9) that is homologous to an archaeal general transcription factor (TFS/M). Each of the 12 subunits of RNAPII is either shared between the three RNAPs or encoded by a paralog in one or both of the other RNAPs. Eukaryotic and archaeal RNAP subunits sharing the same subunit number are homologous, and these subunits are named relative to the RNAPII naming conventions. For example, archaeal and eukaryotic subunits Rpo4, RPA4, RPB4, and RPC4 are all homologous. Eukaryotic polymerase-specific subunits are unique to RNAPI and RNAPIII, which are composed of 14 and 17 subunits, respectively (Jasiak et al. 2006; Proshkina et al. 2006; Kuhn et al. 2007).

Each eukaryotic RNAP has a large number of GTFs. The GTFs are a group of proteins and multiprotein complexes that are required for directed transcription in vivo. With two exceptions, each polymerase has its own set of unrelated GTFs. The first exception is TATA-binding protein (TBP), which is present as a GTF in archaeal RNAP and all three eukaryotic RNAPs. The second exception is transcription factor B (TFB), which is present in archaea and exits as GTF paralogs TFIIIB in RNAPII and TFIIIB-BRF in RNAPIII. Although the GTF repertoires mostly differ between the RNAPs, the functions they perform are analogous. In essence, they direct the polymerases to the promoter, help form the transcription bubble, and allow the polymerases to initiate transcription. Unlike the majority of the polymerase subunits, most GTFs are not homologous across the three eukaryotic RNAPs, implying that they arose early in eukaryotic evolution following the formation...
of their associated polymerases. Based on the extent of protein homology, RNAPII appears to be the ortholog of the archaean transcriptional machinery because it is the only eukaryotic RNAP that has homologous counterparts to all three archaeal GTFs, namely TBP, TFB, and transcription factor E (TFE), which are homologous to TBP, TFIIB, and TFIIE in RNAPII, respectively.

Although functions have been attributed to the polymerase-specific subunits of eukaryotic RNAPI and RNAPIII, their origins in early eukaryotes remain a mystery and there exist little structural data describing these proteins. Recently, Cramer and coworkers used HHpred, a hidden Markov model (HMM)–based remote homolog detection method (Söding et al. 2005), to find homologs of the RNAPI-specific subunits RPA34.5 and RPA49 in sequence profile databases (Kuhn et al. 2007). They identified regions of RPA49 and RPA34.5 that were homologous to the N-termini of human TFIIFα of RPA49 and RPA34.5 (Kuhn et al. 2007). They suggested that the subunits specific to RNAPI and RNAPIII originate from the permanent recruitment of GTFs.

**Materials and Methods**

We retrieved the *S. cerevisiae* N-termini of human TFIIFα of RPA49 and RPA34.5 that were homologous to the profile databases (Kuhn et al. 2007). They identified regions RNAPI-specific subunits RPA34.5 and RPA49 in sequence method (Söding et al. 2005), to find homologs of the Markov model (HMM)–based remote homolog detection method (Söding et al. 2005), to find homologs of the RNAPI-specific subunits RPA34.5 and RPA49 in sequence profile databases (Kuhn et al. 2007). They identified regions of RPA49 and RPA34.5 that were homologous to the N-termini of human TFIIFα and β, respectively, which interact to form a novel triple-barrel structure (Gaiser et al. 2000). The authors also proposed that RPC37 and RPC53 might be the RNAPIII counterparts of TFIIF and RPA49–RPA34.5 (Kuhn et al. 2007).

Here, we extend the analysis of Kuhn et al. (2007) by using HHpred to identify homologs of the five RNAPIII-specific subunits of *Saccharomyces cerevisiae* in an effort to determine whether these subunits are also homologous to eukaryotic GTFs. Although no homologs of RPC31 were found, we identified regions in RPC82 and RPC34 that are homologous to portions of RNAPI's GTF TFIIEα and β subunits, respectively. We also identified regions in RPC37 and RPC53 that are homologous to regions in TFIIFα and β subunits, respectively. The regions of homology between RPC37 and TFIIFα and RPC53 and TFIIFβ correspond to the heterodimer binding interface between the TFIIFα and β subunits (Gaiser et al. 2000), which are the same regions that are homologous between RPA34.5 and RPA49 and the subunits TFIIFα and β, respectively (Kuhn et al. 2007). Finally, we propose a scenario for the evolution of the subunits found in the three eukaryotic RNAPs and argue that the subunits specific to RNAPI and RNAPIII originated from the permanent recruitment of GTFs.

**Results**

**RNAPIII-Specific Subunits Present in All Eukaryotes Are Homologous to the RNAPII GTFs TFIIE and TFIIF**

We used the HHpred server (Söding et al. 2005) to search for regions of homology between RNAPIII-specific subunits and RNAP GTFs found in sequence profile databases. Because the RNAPIII-specific subunits evolve fast compared with the remainder of the RNAPII subunits (Proshkina et al. 2006), we used this sensitive remote homolog detection method to be able to detect potential regions of homology between these subunits and GTFs that are not detectable by traditional homology search methods due to a lack of sensitivity. The RPC37 and RPC53 subunits of the RPC37–RPC53 heterodimer of *S. cerevisiae* showed regions of homology with the human TFIIFα (RAP74) and TFIIFβ (RAP30) subunits that form the TFIIF heterodimer in RNAPII (fig. 1A and B, respectively). Likewise, subunits RPC82 and RPC34 of the yeast RPC82–RPC34–RPC31 heterotrimer share regions of homology with the TFIIFα and β subunits of the TFIIE GTF of RNAPII (fig. 1C and D, respectively). Consequently, RPC82 and RPC34 likely form a heterodimer in organisms that lack RPC31 as do their paralogous subunits in TFIIE. The final RNAPIII-specific subunit, RPC31, did not contain regions of homology with any HMMs of GTF proteins (results not shown).

The regions of homology between RPC37 and TFIIFα and RPC53 and TFIIFβ correspond to the binding interface of TFIIFα (fig. 2; alignments are shown in the supplementary file, Supplementary Material online). The binding interfaces of the α and β subunits (RAP74 and RAP30 in humans, respectively) of the human TFIIF orthologs form a novel “triple-barrel” dimerization interface that is encoded by the N-terminal regions of the TFIIFα and β monomers (Gaiser et al. 2000). The regions of homology correspond to several of the β sheets that form the central barrel. Trypsin digestion suggests that TFIIFα has a flexible linker immediately downstream from the N-terminal binding interface that connects to a C-terminal globular domain (Yong et al. 1998). The C-terminal globular domain of TFIIFα has been shown to interact with TFIIB (Fang and Burton 1996) and RNAPII (Wang and Burton 1995; Fang and Burton 1996),...
although its structure remains unsolved. Because the PFAM profile of TFIIFα (PF05793) is constructed from the full protein and because the RPC37 protein has no region that is detectably homologous to the TFIIFα C-terminal region, RPC37 apparently lacks this region that is present in its TFIIFα homolog. The C-terminus of RPC53 is homologous to the TFIIFβ-interacting N-terminal dimerization interface of TFIIFβ (fig. 1B; Fang and Burton 1996). Because this region is found at the C-terminus of RPC53, RPC53 lacks a downstream region that is homologous to the C-terminal domain of TFIIFβ (fig. 18).

Notably, the regions of both RPC37 and RPC53 that are homologous to the N-termini of TFIIFα and β are also homologous to those of the RPA49 and RPA34.5 subunits of RNAPI (Kuhn et al. 2007). Thus, all three dimers contain this triple-β-barrel interface. This clearly indicates that the one-polymerase transcription system in early eukaryotes contained a dimer that contained this interface. The location of the RPC37–RPC53 binding interface along the RPC37 and RPC53 primary sequences suggests that this heterodimer has a very different structure than its TFIIF and RPA49–RPA34.5 homologs. Secondary structure alignment of both subunits of TFIIF and RPA49–RPA34.5 shows that the β-barrel binding interfaces are situated at the N-terminus in all four proteins (Kuhn et al. 2007). In contrast, the portion of the β-barrel binding interface of RPC53 is situated at its C-terminus (fig. 18).

RPC53 also contains two regions that are homologous to other non-GTF proteins. Part of the N-terminus of RPC53 is homologous to the C-terminal portion of KOG0921 proteins with a probability of 95.2% (fig. 18). A KOG is an eukaryotic Orthologous Group and is composed of a set of orthologous proteins from at least three of the seven eukaryotic genomes represented in the KOG database.

![Fig. 1. Regions of homology between the RNAPIII-specific subunits and the RNAPII GTFs TFIIE and TFIIF. The *Saccharomyces cerevisiae* query subunits are shown with open boxes, and the target HMMs are represented as black bars with thicker regions that represent regions of homology between target and query. The *S. cerevisiae* queries are RPC37 (A), RPC53 (B), RPC82 (C), and RPC34 (D). Protein family targets are labeled as database family ID—description—probability of being a true homolog. Only the target HMMs with a probability greater than 80% of being a true query homolog are displayed; overlapping hits are shown only when more than 50% of its homologous region is nonoverlapping; all hits against GTFs are shown, regardless of significance and overlaps. The scale bar represents length in amino acids.](1037)
KOG0921 is composed of sequences from Drosophila melanogaster, Caenorhabditis elegans, and Homo sapiens, which are the three metazoan genomes represented in the KOG database (Tatusov et al. 2003). Proteins comprising KOG0921 are homologs of the maleless (MLE) protein of dosage compensation complexes. MLE is an RNA helicase that directs the dosage compensation complex to the X chromosome of some male metazoans (Straub and Becker 2007). It upregulates transcription of target genes in the single X chromosome of males in order to match the output of the gene in females where it is present on each of the two copies of the X chromosomes (Straub and Becker 2007).

Part of the C-terminus of RPC53 that overlaps with the TFIIF β homology region is also predicted to be homologous to NTR-like domains with a probability of 92.17% (Tatusov et al. 2003). FIG. 2. Regions of homology between RPC37 and RPC53 and the human TFIIF heterodimer corresponding to the dimer binding interface of TFIIF α and β. (A) The binding interface between the α (green) and β (yellow) subunits of TFIIF. The small figure shows only the two proteins, whereas the larger figure also shows conserved amino acids. Red space-filled amino acids denote high-scoring amino acid columns that are highly similar between TFIIF and RNAIII subunit HMMs (denoted “*” in the alignment shown in part B); yellow space-filled amino acids are medium- to high-scoring amino acid columns (denoted “+” in the alignment shown in part B). (B) HHpred alignment of Saccharomyces cerevisiae query RPC37 HMM (top) and RPC53 HMM (bottom) with target TFIIF α and β HMMs, respectively. Regions in gray in top left figure denote amino acids missing in RPC53 HMM.
NTR-like domains are basic domains that are found at the C-terminus of diverse protein families, including netrins, secreted fizzle-related proteins, complement proteins, type I procollagen C-proteinase enhancer (PCOLCE) proteins, and tissue inhibitors of metalloproteinases (Bányaí and Patthy 1999). Structural studies of the NTR-like domain of human PCOLCE 1 revealed that its NTR-like domain also forms a $\beta$-barrel structure, although it is flanked by $\alpha$ helices and there is no evidence that they are able to form triple-$\beta$-barrels structures (Liepinsh et al. 2003).

RPC82 and RPC34 share regions of homology with TFIIE $\alpha$ and $\beta$ (fig. 1C and D, respectively). The large RPC82 protein has both an N- and a C-terminal region that is homologous to the N-terminal domain of archaeal TFE (fig. 1C). The TFE protein is the archaeal ortholog of TFIIE $\alpha$ (Meinhart et al. 2003). The N-terminal domain of TFE adopts an extended winged helix structure (fig. 3; alignments are shown in the supplementary file, Supplementary Material online). This winged helix domain differs from canonical winged helix domains by having an additional $\alpha$ helix at both the N- and the C-terminal regions of the domain, extended $\beta$. 

**Fig. 3.** Region of homology between the RPC82 C-terminal domain and the Sulfolobus solfataricus N-terminal extended winged helix domain of TFE. (A). The small top right figure shows only the structure of the S. solfataricus extended wing helix domain, whereas the larger figure also shows conserved amino acids. Red space–filled amino acids denote high-scoring amino acid columns that are highly similar between TFIIF and RNAPIII subunit HMMs (denoted “-” in the alignment shown in part B); yellow space–filled amino acids are medium- to high-scoring amino acid columns (denoted “+” in the alignment shown in part B). (B) HHpred alignment of the Saccharomyces cerevisiae RPC82 query HMM C-terminal domain and the HMM of S. solfataricus TFE extended wing helix domain.
strands that make up wing 1, and a negatively charged groove between wing 1 and z3 (Meinhart et al. 2003). RPC82 also has a region that is homologous to part of the winged helix domain of the penicillinase repressor Blal of Bacillus licheniformis (fig. 1C). This region corresponds to helices 1–3 that form the core of the winged helix domain and the first β sheet of wing 1. Whether this region of RPC82 forms another winged helix domain is uncertain as a long z helix is predicted with high confidence to occupy the position in RPC82 where the second β strand of the winged helix domain would otherwise lie (results not shown).

The C-terminal half of RPC34 has a low probability of being homologous to TFIIE β and was only detected when HHpred was initiated with an alignment of hemiascomycete yeast RPC34 sequences instead of the single yeast query sequence (fig. 1D). We have included this weakly predicted homology because RPC34 forms a complex with RPC82, which is homologous to the winged helix domain of TFIIE z. Because TFIIE is a heterodimer of z and β subunits, it is likely that the RPC82 and RPC34 are homologous to the subunits of this GTF. The central portion of RPC34 is also predicted to be homologous to the heat shock σ factor sigma B of B. subtilis with high probability (fig. 1D). Sigma B is a member of the σ20 family of bacterial GTFs, which consists of many groups of similar sigmas that are responsible for coordinating transcription programs in bacterial cells (Gruber and Gross 2003). Although RPC34 was predicted to be homologous to sigma B with high confidence, this relationship was suspect because RPC34 was not predicted to be homologous to any of the other groups of the sigma family. Querying HHpred with sigma B from B. pumilus resulted in highly significant hits against dozens of these different sigma groups, suggesting that RPC34 is not a true homolog of sigma B (results not shown).

Discussion

Polymerase-Specific Subunits Are Paralogs of RNAPII GTFs

Our results suggest that four of the five polymerase-specific subunits of RNAPII are in fact paralogs of subunits of the TFIIE and TFIIF GTFs of RNAPII. Only RPC31 seems to lack homology to other proteins, but this protein is specific to plants and unikonts and is thus not universally present in eukaryotes (Proshkina et al. 2006). RPC82 and RPC34 appear to be paralogs of TFIIE z and β, respectively, although the confidence in the latter paralogy by HHpred itself is weak (fig. 1). As discussed above, we believe it to be a true paralog because of its association with RPC82, which is homologous to TFIIE z subunit that associates with TFIIE β (fig. 3). Finally, we demonstrated that RPC37 and RPC53 are paralogs of TFIIF z and β, respectively, and that the region of homology corresponds to the z–β binding interface of TFIIF (fig. 2). Moreover, this z–β binding interface homology is also present in RNAPI subunits RPA49 and RPA34.5 (Kuhn et al. 2007).

Because archaeal RNAP and RNAPII are highly similar in terms of subunit composition, we argue that the additional subunits of RNAPI and RNAPIII result from the recruitment of existing GTFs that were present in the three ancestral eukaryotic RNAPs. This is opposed to the alternative scenario, in which all three ancestral eukaryotic RNAPs had 17 subunits and where RNAPI and RNAPII experienced several subunit loss events but RNAPIII kept all 17 subunits. If the most recent archaeal and eukaryotic ancestor contained an 11- or a 12-subunit RNAP, three GTF gain/loss events are sufficient to explain the presence of the RNAPI- and RNAPIII-specific subunits. However, many loss events would be required to explain the RNAP subunit compositions if the ancestral RNAP contained 17 subunits. Because RNAPII and the archaeal RNAP have almost identical subunit compositions, the same five subunits of the ancestral RNAP would have to be separately lost along the archaeal and the RNAPII branch. Although this seems unlikely in itself, RNAPI must also have lost its homologs of RPC31, RPC82, and RPC34. We believe that the more likely scenario for subunit expansion of any protein complex is through a transiently interacting protein intermediate. In this scenario, a protein can begin interacting with the complex, become functionally associated, and then eventually become a permanent member of the protein complex. This seems like a much more likely pathway than a completely unassociated protein becoming a permanent member of a protein complex and then dissociating to become a transient interaction partner.

Functional Similarities between Homologous GTFs

Both TFIIE and its homologs RPC37–RPC53 as well as TFIIF and its homologs RPA34.5–RPA49 and RPC37–RPC53 have maintained functional similarities across the RNAPs. In RNAPII, TFIIE is involved in both transcription initiation and the transition from initiation to elongation (Holstege et al. 1996; Watanabe et al. 2003; Tanaka et al. 2009). Whereas TFIIE z likely interacts with TBP through its N-terminal extended winged helix domain in the RNAPII pre-initiation complex (PIC) (Maxon et al. 1994; Bell et al. 2001), little is known about the function of its RNAPIII homolog RPC82. Because RNAPIII also uses TBP as a GTF, RPC82 and TFIIE z might interact similarly with this protein. TFIIE β has been implicated in both transcription initiation and the transition to elongation in RNAPII (Watanabe et al. 2003; Tanaka et al. 2009). Although its homolog RPC34 is involved in the recruitment of RNAPIII to the PIC, it has also been implicated in transcription initiation, early elongation, or both via mutagenesis studies (Brun et al. 1997). No RNAPI component has yet been discovered that is involved specifically in either transcription initiation or the transition to elongation.

TFIIF and its homologs are implicated in transcription elongation in all three RNAPs. The RPA34.5–RPA49 heterodimer is required in vitro for elongation and aids elongation in vivo (Kuhn et al. 2007). Although not a subunit of RNAPII like its homologs in RNAPI and RNAPIII, TFIIF is the GTF that is most intimately associated with RNAPII. It enters the PIC bound to RNAPII and associates with stalled RNAPII, which is consistent with a role in elongation (Zawel
et al. 1995). Finally, RPC37/RPC53 is also involved in elongation but appears to slow down the transcription rate of RNAPIII to allow recognition of termination signals encoded in its target genes (Landrieux et al. 2006).

**Evolution of the Cellular Transcriptional Machinery**

Based on the known homologies between transcriptional proteins in the three domains of life, a scenario for the evolution of much of the transcriptional machinery emerges (fig. 4). Here, we propose such a scenario for the archaeal and eukaryotic RNAPs due to their highly similar polymerase and GTF protein repertoires. We exclude eubacterial RNAPs due to the distinct differences in their GTFs compared with those of archaea and eukaryotes.

Most of the subunit complexity that is observed in eukaryotic RNAPII had emerged before the archaeal–eukaryotic split as the archaeal RNAP and eukaryotic RNAPII have almost identical subunit composition (fig. 4). The only archaeal-specific subunit is Rpo13, which was recently identified as a subunit in the high-resolution crystal structure of the crenarchaeote *Sulfolobus shibatae* (Korkhin et al. 2009). Because this subunit is specific to crenarchaeota (minus Rpo13). Thus, the ancestral archaea already contained most of the complement of RNAP subunits and GTFs that appear in extant archaeal cells, and elaborations of the RNAPs occurred mostly along the lineage leading to eukaryotes after their divergence from archaea.

Shared subunit composition of the eukaryotic RNAPs and GTFs evokes a scenario for the evolution of the three-RNAP transcriptional system of extant eukaryotes. The earliest eukaryotes likely contained the full complement of RNAPII subunits, which almost all have orthologs in archaeal RNAPs (fig. 4). The only exception is the archaeal protein TFS/M, which is a GTF in archaea but has a homolog in the eukaryotic RNAPs (RP9, RPB9, and RPC9) that are bona fide subunits. Because part of RP9 is homologous to the GTF/elongation factor TFIIS (Langer and Zillig 1993), there was likely a duplication event of the ancestral TFS/M to generate these paralogs in RNAPII. Early eukaryotes contained TBP, TFB, and TFE as their GTFs. The homology between the triple-barrel binding interface of TFIIF α/β, RPA49/RPA34.5 (Kuhn et al. 2007), and RPC37/RPC53 (this work) indicates that a dimeric protein containing this domain became associated with the ancestral RNAP prior to the expansion of the one-RNAP transcriptional system in eukaryotes. This triple-barrel binding interface is known to be involved in transcription start-site selection, early bond formation, and stabilization of an early elongating complex in TFIIF of *S. cerevisiae* RNAPII (Khapersky et al. 2008).

![Model for the evolution of the three eukaryotic RNAPs](image-url)
The first RNAP duplication event(s) likely resulted in two RNAPs: RNAPI and the ancestral RNAPII/RNAPIII. This is supported by the existence of three homologous proteins that are found in the RNAPII and RNAPIII transcriptional systems but with no homologs in RNAPI. The first two proteins are TFIIE $\alpha$ and $\beta$ that are found in RNAPII and that are homologous to RPC82 and RPC34 in RNAPIII, respectively. In contrast, RNAPI lacks a homolog of either subunit of TFIIE, indicating a loss of TFE in this RNAP. Thus, the ancestral TFE was likely lost along the RNAPII lineage but maintained in both RNAPII and RNAPIII following duplication of the RNAPII/RNAPIII ancestor. The third such protein is the archaeal TFB, which is common to both RNAPII (where it is named TFIIB) and RNAPIII (where it is named TFIIB-BRF) but absent from RNAPI (fig. 4). These recruitment events resulted in the expansion of the number of RNAP subunits from 12 in the ancestral eukaryotic RNAP to 14 in RNAPI (Kuhn et al. 2007) and 16 in RNAPIII (17 if we include the unikont- and plant-specific RPC31 protein; Proshkina et al. 2006).

Although we interpreted the evolution of the cellular transcription system in the framework of the three-domain tree of life (Woese et al. 1990), it is equally consistent (parsimonious) with the eocyte hypothesis in which the crenarchaea and eukaryotes form a monophyletic clade and are sister taxa of euryarchaea (Rivera and Lake 1992; Cox et al. 2008). The two different topologies are almost equivalent in terms of explaining the origin of RNAP proteins and only differ in terms of the location of the origin of the Rpo8/RP8 (fig. 4). In the three-domain tree, Rpo8 is present in the most recent common ancestor of archaea and eukaryotes and was lost along the euryarchaeal lineage (fig. 4), whereas in the eocyte tree, it does not originate until the advent of the crenarchaea and does not require loss of this protein in any lineage. Further support for the eocyte origin of eukaryotes is the recent identification of a crenarchaeote, Cenarchaeum symbiosum (Hallam et al. 2006; Cox et al. 2008; Kwapisz et al. 2008).

In conclusion, we have identified homologies between four of the RNAPII-specific subunits and GTFs of RNAPII and argued that the increase in the number of subunits in RNAPI and RNAPIII was due to the recruitment of GTFs. Accordingly, subunit expansion of RNAPIII occurred in the same manner as that of RPA by the recruitment of transiently associated GTFs into permanently associated bona fide polymerase subunits. This appears to be a repeating theme throughout RNAP evolution because at least three separate permanent GTF recruitment events occurred: 1) a homolog of the archaeal TFS/M, which is a GTF in archaea, became permanently recruited to the ancestral eukaryotic RNAP to form a bona fide subunit, RP9; 2) the RNAPI and RNAPIII homologs of TFIIF were recruited as permanent subunits in early eukaryotes to become RPA49 and RPA34.5 in RNAPI and RPC37 and RPC53 in RNAPIII; and 3) TFIIE $\alpha$ and $\beta$ were recruited to become RPC82 and RPC34 in RNAPIII.

**Supplementary Material**

**Supplementary file** is available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

We thank the anonymous referees for their useful and constructive comments on a previous version of this manuscript. This work was supported by a Discovery Grant from the Natural Science and Engineering Research Council of Canada to G.D.

**References**


Holstege FC, van der Vliet PC, Timmers HT. 1996. Opening of an RNA polymerase II promoter occurs in two distinct steps and requires the basal transcription factors IIe and IIh. *EMBO J.* 15:1666–1677.


