EUKARYOTIC CELL, June 2004, p. 735–740 Vol. 3, No. 3
1535-9778/04/$08.00+0 DOI: 10.1128/EC.3.3.735–740.2004

Functional Unit of the RNA Polymerase II C-Terminal Domain Lies within Heptapeptide Pairs†

John W. Stiller* and Matthew S. Cook

Department of Biology, East Carolina University, Greenville, North Carolina 27858

Received 12 October 2003/Accepted 15 March 2004

Unlike all other RNA polymerases, the largest subunit (RPB1) of eukaryotic DNA-dependent RNA polymerase II (RNAP II) has a C-terminal domain (CTD) comprising tandemly repeated heptapeptides with the consensus sequence Y-S-P-T-S-P-S. The tandem structure, heptad consensus, and most key functions of the CTD are conserved between yeast and mammals. In fact, all metazoans, fungi, and green plants examined to date, as well as the nearest protistan relatives of these multicellular groups, contain a tandemly repeated CTD. In contrast, the RNAP II largest subunits from many other eukaryotic organisms have a highly degenerate C terminus or show no semblance of the CTD whatsoever. The reasons for intense stabilizing selection on CTD structure in certain eukaryotes, and its apparent absence in others, are unknown. Here we demonstrate, through in vivo genetic complementation, that the essential functional unit of the yeast CTD is contained within pairs of heptapeptides. Insertion of a single alanine residue between diheptads has little phenotypic effect, while increasing the distance between diheptads produces a mostly quantitative effect on yeast cell growth. We further explore structural constraints on the CTD within an evolutionary context and propose selective mechanisms that could maintain a global tandem structure across hundreds of millions of years of eukaryotic evolution.

The C-terminal domain (CTD) of the largest subunit (RPB1) of DNA-dependent RNA polymerase II (RNAP II) comprises a set of tandemly repeated heptapeptides ($Y^1S^2P^3T^4S^5P^6S^7$) that are essential for viability in both animals and yeast (1, 2). The CTD functions throughout the RNAP II transcription cycle; it has been likened to a symphony conductor, orchestrating a dizzying array of protein-protein interactions required for proper transcript initiation, elongation, and cotranscriptional mRNA processing (14). Specific binding of additional proteins in vitro also has implicated the CTD in overall genome maintenance and regulation (5). Based on the CTD's central importance to the molecular biology of the cell, it comes as no surprise that its primary structure has been conserved strongly throughout the evolution of animals, plants, fungi, and their nearest protistan relatives (23). In contrast, nearly all eukaryotes outside this group have lost the canonical CTD, if it ever was present in their ancestors (23). Given the degeneration or lack of tandem heptad repeats in many eukaryotes, the specific basis for strong stabilizing selection on CTD structure in animals, plants, and fungi is unclear.

Previous genetic investigations in yeast and animals have shown that residues Y^1, S^2, and S^5 are essential for CTD function and that a severe reduction in the total number of heptads is lethal (2, 19, 27). These results correlate well with observed evolutionary variation in CTD structure (7, 23). Although the yeast CTD itself has relatively few deviations from the consensus YSPTSPS sequence, a wide variety of individual substitutions found in other organisms can be tolerated by yeast cells. This includes complete replacement of the yeast CTD by the longer and more highly substituted sequence from mouse (1), as well as by nonconsensus heptads $(YSPASPA)_{25}$ from the protist *Mastigamoeba invertens* (24). Thus, many of the individual substitutions that have accumulated across broad stretches of eukaryotic evolution are compatible with core CTD function in yeast.

In addition to single amino acid changes, more-severe departures from canonical CTD structure occur in many eukaryotes. Presumably, such deviations are permitted because these organisms lack some of the essential CTD-protein interactions characterized in animals and yeast. If so, yeast cells should not tolerate wholesale disruption of the CTD's overall tandem structure in the same way they accommodate individual substitutions occurring within the confines of a canonical CTD. Here we demonstrate that the essential, conserved functional unit of the CTD lies within paired heptapeptides, using genetic complementation in yeast by both evolutionary and artificial mutants. Our results offer further insights into the preservation of a tandem heptad structure over vast stretches of eukaryotic evolution.

MATERIALS AND METHODS

Construction of CTD evolutionary mutants. Clones encompassing the region 3′ of the G domain of *RPB1*, isolated previously from two red algae (23), were used as PCR templates with primers designed to amplify sequences encoding different portions of their RPB1 C termini (Fig. 1). These primers were constructed with different terminal AvaI restriction sites to promote directional cloning. Fragments were ligated overnight at 15°C at a 1:1 molar ratio with an AvaI-digested and dephosphorylated CTD-less subclone (pSBO) and transformed into *Escherichia coli* cells. Colonies were screened via PCR, and clones containing an appropriate-sized insert were sequenced completely in both directions, through ligation sites, to establish that no incidental mutations had been introduced. pSBO subclones were digested with KpnI and SnaBI, the insert was purified on low-melting-point agarose gels, and then it was ligated into a yeast shuttle vector (pY1) from which the wild-type (WT) CTD had been excised. The resulting pY1

* Corresponding author. Mailing address: Department of Biology, Howell Science Complex, N108, East Carolina University, Greenville, NC 27858. Phone: (252) 328-2738. Fax: 252-328-4178. E-mail: stillerj@mail.ecu.edu.

† Supplemental material for this article may be found at http://ec.asm.org.

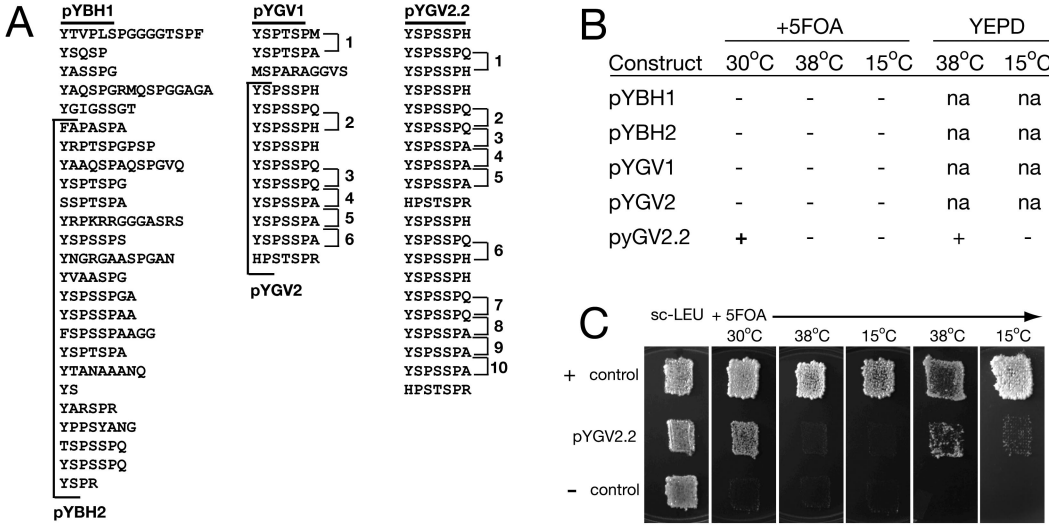


FIG. 1. Sequences used in genetic complementation for the CTD in yeast. (A) Red algal sequences used to replace the yeast CTD. pYBH1 and pYBH2 were recovered from *B. hamifera* using PCR linkers engineered with AvaI sites for directional cloning. pYBH1 includes the entire sequence present inserted in place of the 26 yeast WT CTD repeats, whereas pYBH2 includes only the sequence enclosed in the left-hand bracket. pYGV1 and -2 were isolated in the same manner from *G. vacuolata*. None of these sequences complemented CTD function. pYGV2.2 was made by concatenating two copies of the pYGV2 insert. Right-hand brackets on pYGV sequences represent the presumably functional diheptad units present in these constructs. (B and C) Results of the plasmid shuffle assay in yeast using red algal CTD mutants. A positive control transformed with the pY1 shuttle vector, without substitution for the WT CTD sequence, is at the top of each panel, and a negative control representing yeast cells transformed with a *LEU2* plasmid containing no copy of *RPB*1 is at the bottom. The left-hand panel shows original transformants containing both WT and mutant CTD; subsequent panels show replica plating on 5-FOA, which forces yeast cells to lose the *URA3*-linked copy of *RPB*1 with WT CTD. The two right-hand panels show cells first acclimated to 5-FOA selection at 30°C and then replica plated on YEPD (without 5-FOA) at high and low temperatures. No residual negative control is visible, due to its loss during the first round of 5-FOA selection at 30°C. Although cells from the replica transfer were visible at 15°C, subsequent plating on YEPD at 30°C demonstrated that they were inviable.

subclones were used to transform yeast cells as described below (for a complete description of vectors and subcloning procedures, see references 23 and 27).

Construction of artificial CTD sequences. Complementary 5′-phosphorylated oligonucleotides were designed to encode the consensus CTD heptad in yeast, with additional Ala residues inserted in various positions (Fig. 2). Codon choices matched the most commonly used triplets in the yeast WT CTD. When annealed, the resulting double-stranded fragments were left with overhangs matching the two different AvaI recognition sites to facilitate directional cloning of concatenated fragments. Complementary oligonucleotides were annealed together and ligated into the pSBO vector as described above, but at a 20:1 (insert/vector) molar ratio. Because CTD truncation mutants with less than 13 repeats show at least conditional phenotypes, we screened artificial CTD subclones for inserts containing at least 13 WT heptapeptide motifs.

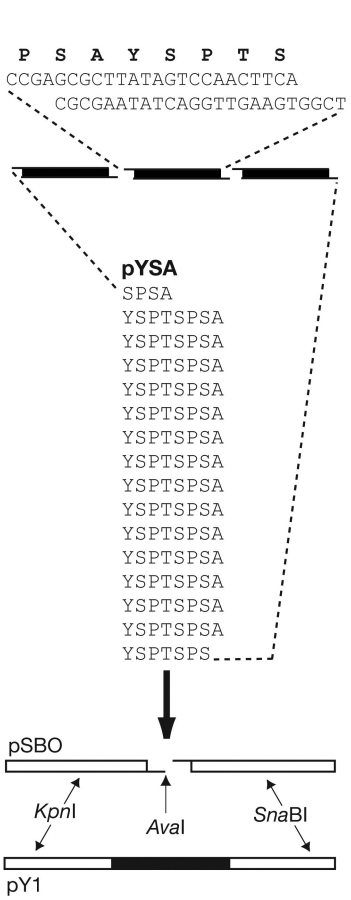
Yeast transformations. Yeast WT CTD was replaced by mutated constructs via the plasmid shuffle. The yeast strain Z26 (20) was transformed by lithium-acetate treatment (16) and selected on synthetic complete (SC)-Leu-Ura medium to retain both the *URA3*-linked WT CTD and *LEU2*-linked mutant genes. Transformed colonies were replica plated onto SC-Leu medium containing 5-fluoroorotic acid (5-FOA) (3) to select cells without the *URA3*-linked RPB1$^+$ gene. Replica plates were incubated at 30, 15, and 38°C. For cells with conditionally lethal phenotypes under direct Leu plus 5-FOA selection, cold and temperature sensitivity also were tested on complete (YEPD) medium after cells acclimated to 5-FOA at permissive temperature. In cases where no growth was observed after 3 weeks, colonies on SC plus 5-FOA were further replica plated onto complete medium at 30°C to verify that the CTD construct was lethal.

To measure growth rates of transformants relative to each other and to a positive control transformed with pY1 (contains *RPB1* with WT CTD), cells were grown in 100 ml of YEPD to an optical density at 600 nm (OD_{600}) of 0.1 and then measured periodically during log phase until they reached an OD of 1.0. Relative growth rates were calculated as the ratio of the generation time of each mutant with that of the positive control. With the exception of pYDA5, replicate cultures were assayed in each case. Finally, to assure that lethality was not due to mutations that had occurred elsewhere in the gene during the cloning and transformation procedures, mutated inserts were removed and the WT CTD was reinserted into several lethal constructs chosen haphazardly (see Fig. S1 in the supplemental material).

RESULTS AND DISCUSSION

Transformation with evolutionary constructs. Phylogenetic analysis of *RPB1* sequences strongly supports a unique group of evolutionarily related organisms (for identification referred to as the CTD clade) in which the canonical sequence and tandem structure of the RNAP II CTD are invariably conserved by strong stabilizing selection (23) (Fig. 3). As a preliminary test of this evolutionary hypothesis, we replaced the yeast CTD with evolutionary mutants derived from two red algae. Red algae appear to be the most recent ancestor of CTD-clade organisms (Fig. 3), and their distal RPB1 sequences contain CTD-like motifs (22). *Bonnemaisonia hamifera* RPB1 has more than the minimum of eight heptad motifs required for viability in yeast (Fig. 1); however, these heptads all have some substitutions deviating from the CTD consensus and, perhaps more importantly, they are not arrayed in tandem (22). Yeast cells were transformed with two different *Bonnemaisonia* constructs; one comprising the entire *Bonnemaisonia* RPB1 C-terminal sequence enriched in common CTD amino acids and the other containing only the portion of sequence most similar to canonical CTD heptads. Both of these constructs proved lethal (Fig. 1).

Because the *Bonnemaisonia* sequences are characterized both by individual amino acid differences and by the absence of a tandemly repeated organization, additional constructs were developed to explore the reasons for their lethality in yeast. First, we determined whether yeast are affected detrimentally by unusual residues found in red algal sequences, even within the context of



pYDA5
SPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPSA
YSPTSPS
YSPTSPSA
YCPTSPS

pYYA
SPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS
YASPTSPS

pYDA9
SPS
YSPTSPSA
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YSPTSPS
YSPTSPSA
YCPTSPS

pYTA
SPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS
YSPTASPS

pYD2A
SPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS
YSPTSPSAA
YSPTSPS

pYAS
SPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSPRS
YSPTSPAS
YSPTSPAS
YSPTSPAS
YSHFTRS
YSPTSPRS
YSPTSPRS
YSPTSPS

pYD5A
SPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS
YSPTSPSAAAAA
YSPTSPS

pYAT
SPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATTPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS
YSPATSPS

FIG. 2. Full sequences of artificial CTD sequences used in yeast transformations, along with subcloning strategy for constructing RPB1 genes with mutated CTDs. Octadapeptide constructs with shorter artificial CTDs were screened, but all proved inviable. Therefore, only the longest sequences are included here and in our documented transformation results (see Fig. 4).

well-ordered tandem repeats. *Glaucosphaera vacuolata* is a unicellular red alga; morphologically and developmentally it is the simplest of the red algae from which *RPB*1 has been sequenced. Curiously, its "CTD" consists of 19 tandem heptads, most of which do not conform to the YSPTSPS consensus. Two constructs of slightly different lengths were developed from *Glaucosphaera*; both were lethal in yeast (Fig. 1).

Yeast cells can survive with as few as 8 CTD repeats, but truncation to less than 13 begins to produce conditional phenotypes (27); these deleterious effects worsen with decreasing numbers of repeats. To determine whether the lethality of *Glaucosphaera* substitutions can be mitigated by increasing the number of heptads present, two pYGV2 inserts were concatenated to form a 19-repeat construct (pYGV2.2) (Fig. 1). In this case, transformed cells were viable at 30°C but had a slow-growth phenotype (Fig. 1). Cells plated directly on selective medium (5-FOA) were temperature sensitive (ts) and cold sensitive (cs) (Fig. 1); when stabilized first at permissive temperature for several days they were capable of very slow growth at 38°C but remained cs (Fig. 1). This construct encodes considerably more than the minimum 13 heptads required for apparent WT growth, and the nonconsensus substitutions present occur mostly at nonessential positions 4 and 7; nevertheless, some of these substitutions clearly are deleterious to CTD function.

Transformation with artificial CTDs. The results of these experiments suggest that, although many noncanonical substitutions present in evolutionary variants are at least minimally compatible with essential CTD functions, disruption of the overall tandem heptapeptide register is not. To investigate the effects of such disruptions further, we constructed a set of artificial sequences in which the consensus YSPTSPS was conserved but alanine residues were introduced between adjacent heptad motifs. Alanines were inserted in all positions except between serine-proline pairs, which serve as essential phosphorylation substrates for CTD kinases (12, 21, 27). Alanine insertions resulted in tandemly repeated octads, each of which contained the consensus heptapeptide broken up at different positions (Fig. 2 and 4). Not surprisingly, given the strongly conserved tandem structure of the yeast CTD, all cells transformed with each of these octad constructs were inviable (Fig. 4).

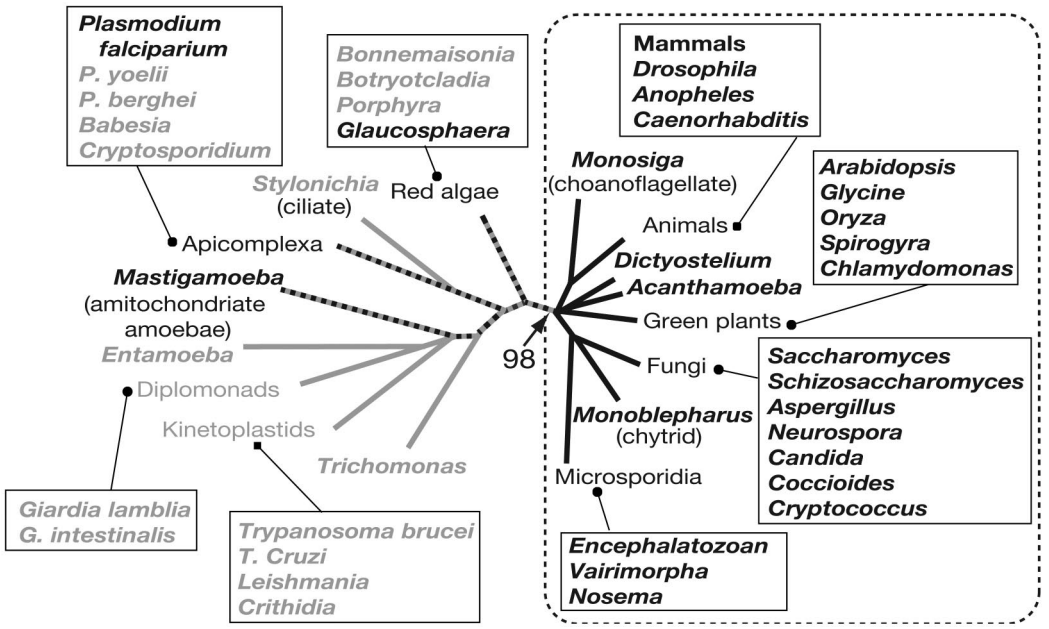


FIG. 3. Phylogenetic tree recovered though Bayesian inference (15) on an alignment of inferred amino acid sequences of RPB1 regions A to H (17) from 31 exemplars, representing 16 lineages of eukaryotes. Maximum-likelihood branch lengths were calculated with TreePuzzle 50 (25). Both analyses applied a discrete estimate of 1 invariable + 4 γ-distributed rates among sites and a Jones, Taylor, and Thornton model for probabilities of changes among amino acids. RPB1 genes from organisms in bold encode a clear set of tandemly repeated C-terminal heptapeptides. The arrow designates an internode on the tree, with a 98% Bayesian confidence level (calculated from 20,000 sampled trees), that divides eukaryotes into two distinct groups. To the right, enclosed in a dashed box, is a CTD clade in which all organisms sampled to date contain a canonical CTD. In contrast, none of the taxa to the left of this divide has a clearly canonical CTD. Although several of these sequences have noncanonical heptad repeats, most have no indication of a tandem heptad structure whatsoever. The *RPB1* sequence alignment is available at the website http://personal.ecu.edu/stillerj/rpb1aln.htm. Also, see reference 23 for a thorough phylogenetic treatment of *RPB1* sequences and the CTD clade.

To this point, the results of our transformation experiments correlated well with the apparently strong stabilizing selection on CTD structure. Our next construct, however, yielded something of a surprise. Although a single residue inserted between every heptad unit always is lethal, introduction of an Ala residue between every other heptad appeared to have little effect on yeast cells (Fig. 4). Transformants containing 9 such diheptad repeats (18 total heptads) were neither cs nor ts and grew at approximately 93% (based on ratio of generation times) the rate of yeast cells carrying a WT CTD. Introducing a second Ala between alternating heptads also was not lethal, but it did result in a further decline (67% of WT CTD) in growth rate (Fig. 4) and extremely slow growth at high and low temperatures. Remarkably, diheptads separated by as many as five Ala residues supported growth in yeast cells, although only at a permissive temperature (under 5-FOA selection) and at a very reduced rate (57% of WT CTD) (Fig. 4).

Zeroing in on the essential unit for CTD function. The unconditional lethality of interrupting all heptad pairs, combined with nearly WT growth when every other heptad is interrupted, indicates that all essential CTD functions are accomplished through interactions of protein factors with a motif, or motifs, that lie within individual diheptads. As distance between these pairs increases, growth rate slows incrementally and conditional phenotypes begin to develop. In other words, although breaking up diheptad units has a clear qualitative effect (a lethal disruption of CTD function), the effects of moving diheptads apart appear to be largely quantitative.

The requirement for a minimum number of diheptads also can help to explain the results from our evolutionary complementation experiments. Based on their need for eight heptads and assuming there is some functional redundancy in overlapping pairs, yeast could require at least seven functional diheptads to be viable. The complete breakdown of a repetitive structure in the *Bonnemaisonia* sequence has resulted in, at most, three potential diheptad units (Fig. 1). It is not surprising, then, that these sequences cannot complement CTD function. Although all of the CTD mutants constructed from the *Glaucosphaera* sequence had enough repeats to provide WT CTD function under the conditions tested, were they consensus heptapeptides, a number of them contained His residues, mostly at position 7. Histidine has a bulky imidazole ring side chain that presumably has a significant impact on three-dimensional structure and binding properties; it is extremely rare in CTD sequences in general and is not found in yeast WT CTD heptads. The presence of a His within the functional unit may effectively block a given diheptad from interacting with one or more of its potential protein partners.

If diheptads interrupted by His residues are discounted, the two shorter *Glaucosphaera*-based RPB1 constructs may not contain enough diheptides for yeast viability (Fig. 1). In contrast, there are 8 to 10 His-free diheptads present in pYGV2.2,

A

Construct	Insert	+5FOA 30°C	+5FOA 38°C	+5FOA 15°C	YEPD 38°C	YEPD 15°C
pYSA	$[YSPTSPSA]_{14}$	-	-	-	na	na
pYAS	$[SYSPTSPA]_{18}$	-	-	-	na	na
pYYA	$[SPTSPSYA]_{23}$	-	-	-	na	na
pYAT	$[TSPSYSPA]_{16}$	-	-	-	na	na
pYTA	$[SPSYSPTA]_{15}$	-	-	-	na	na
pYDA9	$[(YSPTSPS)_2A]_9$	++	++	++	++	++
pYDA5	$[(YSPTSPS)_2A]_5$	+	+	+	+	+
pYD2A	$[(YSPTSPS)_22A]_7$	+	-	+	+	+
pYD5A	$[(YSPTSPS)_25A]_8$	+	-	-	+	+

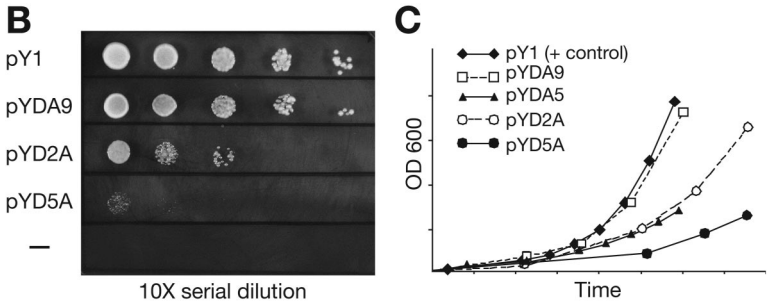


FIG. 4. Growth data for C-terminal constructs comprising diheptad repeats interrupted by Ala residues (see Fig. 2 for specific sequences of each construct). (A) Artificial constructs made by concatenating typical yeast heptads interrupted by Ala residues. Subscripts indicate diheptads as well as the number of repeat units in the construct. Transformants were assayed for the ability to grow directly under 5-FOA selection and on less-stringent media after acclimation to 5-FOA. A pair of bold pluses (++) indicates a growth rate close to that of the WT CTD under the respective plating conditions, a single bold + indicates an intermediate growth phenotype, and a + not in bold indicates an extremely slow growth phenotype. (B) Titration of liquid cultures grown for 48 h in SC plus 5-FOA medium. Cultures all were inoculated with 10^4 cells at time zero and grown in 10 ml of medium at room temperature on an orbital rotator. Cultures were incubated for 2 days to assure that some growth would be observed in pYD5A transformants; consequently, both WT CTD and pYDA9 cultures approached stationary phase when plated and showed a slightly smaller difference in apparent growth rate than was obtained by quantitative measurements. (C) Quantitative comparison of growth rates of WT and various diheptad transformants in YEPD medium. The pYD5A cold temperature slow-growth phenotype is extreme; even on YEPD medium, plated cells did not become visible for several weeks. In general, slow-growth phenotypes were more exaggerated on selective medium (plate B) than on complete (plate C), indicative of a decline in the ability to adapt to some complex transcription requirements, as seen previously in progressively shorter CTD truncation mutants (27).

enough to confer minimal CTD function. In that light, it is interesting that the phenotype of the pYGV2.2 transformant is similar to those exhibited by severe CTD truncation mutants (27).

Flexibility in the spacing of diheptads is consistent with known three-dimensional structures of the CTD complexed with different protein partners. Cocrystallization of the Cgt1 capping enzyme from *Candida albicans* with four YSPTSPS repeats shows two distinct Cgt1 docking sites (CDS1 and CDS2) that interact with a 17-amino-acid segment of the CTD (9). Although this site doesn't fit within a single diheptad, the actual points of contact between Cgt1 and the CTD do; one heptad of the 17-amino-acid segment loops out and is only loosely associated with the Cgt1 surface. Indeed, it appears likely that this loop can include variable numbers of heptads depending on the binding context (9, 11) and, based on our results, the specific length and sequence of the looped region is not of critical importance. A comparison of the Cgt1-CTD complex to that of Pin1 (peptidyl-proline isomerase) (26) suggests a remarkable flexibility in the conformation of bound CTD heptads. This is consistent with the need for the CTD to interact with so many disparate protein structures (5, 9), including a number that appear to bind to multiple CTD locations (19). Given the results presented here, this flexibility extends to the spacing between binding domains as well.

Evolutionary conservation of CTD tandem structure. If the required CTD functional unit lies within individual diheptads, why is the domain's overall tandem structure conserved so strongly? One explanation simply may be that, once a tandem structure is established, most insertions or deletions are likely to disrupt rather than fall between functional domains. In addition, although alanine insertions are well-tolerated, other residues (e.g., His, as suggested by our pYGV2.2 mutant) could disrupt CTD function even when inserted between functional units. Thus, the overall tandem structure of the CTD may be somewhat self-policing with regard to stabilizing selection. Nevertheless, given a billion or more years of evolution, the broad diversity of organisms comprising the CTD clade, and the functional redundancy of CTD heptads (20, 27), it appears likely that additional selective forces are responsible for the remarkable conservation of overall tandem CTD structure.

Although not required for essential CTD functions in vivo, maintaining heptads in methodical tandem repeats offers at least two selective advantages. First, for a given number of heptads, tandem repeats yield more individual diheptads. For

example, six discontinuous heptads can provide, at most, three diheptad units, whereas six tandemly repeated heptads form five diheptads. Control over elongation and processing events during the RNAP II transcription cycle is a complex and dynamic process, involving reversible phosphorylation of CTD residues (18, 21) and their interactions with a variety of initiation, elongation, and processing factors (12–14). A maximum number of functional units within a given investment of sequence could increase the efficiency of sequential and/or competing reactions; this is best achieved through continuously tandem repeats.

The continuous decline in growth rates and induction of conditional phenotypes, as comparable numbers of diheptad units are moved farther apart, require an additional explanation. There must be mechanical constraints involving interactions among multiple CTD binding sites. This is not surprising, given that many initiation and processing steps are interdependent and require the cooperative or at least concurrent actions of multiple protein factors that associate with the CTD (4, 6, 8, 21). Evidence that such proteins not only bind the CTD but also concentrate in specific subnuclear regions suggests that the CTD plays the role of an organizational platform for coordinating a variety of transcriptional and processing activities into multifunction transcriptosomes (5, 8, 11, 14). A simple increase in physical separation or change in orientation of such proteins, each bound to different CTD diheptad units, may be sufficient to reduce the efficiency of their interactions and account for the slow-growth phenotypes we observed. It is reasonable to conclude that potentially cooperative interactions among CTD-bound proteins, as well as individual protein factors that bind to multiple CTD locations, have been honed by evolution to function at peak efficiency when diheptads are spaced in an overall tandem structure.

Conclusions. The canonical CTD has proven to be an extremely useful and flexible sequence, which has been adapted as a staging platform for coordinating a variety of steps in the RNAP II transcription cycle. As that cycle has evolved independently in different eukaryotes, some CTD-protein interactions specific to each system have developed; for example, exon definition during pre-mRNA splicing occurs in mammals but not yeast (14, 28). In more complex organisms with longer and more-highly substituted CTD sequences, such as animals and plants, some of these interactions may well involve functional units longer than individual diheptads. In mammals, the distinct functionalities of canonical proximal repeats, as opposed to more-highly substituted distal heptads (10), may very well involve different CTD binding requirements. Thus, it would be surprising to find that maintenance of basic diheptad units is sufficient for viability in a system as complicated as mammalian RNAP II transcription. Nevertheless, the structure and consensus sequence of the CTD has been conserved strongly in a distinct group of eukaryotes and across large evolutionary distances (Fig. 3). This supports the hypothesis that a core set of CTD-protein interactions are present and essential in all of these CTD-clade organisms (23). Based on the results presented here, the functional unit required for those core interactions appears to reside within individual pairs of CTD heptapeptides.

ACKNOWLEDGMENTS

This research is based upon work supported by the National Science Foundation under grant MCB 0133295.

We thank J. Corden for providing the original CTD-less subclone and yeast shuttle vector used in transformation experiments, B. Hall for help with initial implementation of the plasmid shuffle, and A. Greenleaf and H. Phatnani for prereview of the manuscript and helpful feedback. We also thank three anonymous reviewers for their comments and suggestions.

REFERENCES

1. **Allison, L. A., J. K. Wong, V. D. Fitzpatrick, M. Moyle, and C. J. Ingles.** 1988. The C-terminal domain of the largest subunit of RNA polymerase II of *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and mammals: a conserved structure with an essential function. Mol. Cell. Biol. **8:**321–329.
2. **Bartolomei, M. S., N. F. Halden, C. R. Cullen, and J. L. Corden.** 1988. Genetic analysis of the repetitive carboxyl-terminal domain of the largest subunit of mouse RNA polymerase II. Mol. Cell. Biol. **8:**330–339.
3. **Boeke, J. D., J. Trueheart, G. Natsoulis, and G. R. Fink.** 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. **154:**164–175.
4. **Calvo, O., and J. L. Manley.** 2003. Strange bedfellows: polyadenlyation factors at the promoter. Genes Dev. **17:**1321–1327.
5. **Carty, S. M., and A. L. Greenleaf.** 2002. Phospho- CTD-association proteins in the nuclear proteome link transcription to DNA/chromatin modification and RNA procession. Mol. Cell. Proteomics **1:**598–610.
6. **Cho, E.-J., T. Takagi, C. R. Moore, and S. Buratowski.** 1997. mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain. Genes Dev. **11:**3319–3326.
7. **Corden, J. L.** 1990. Tails of polymerase II. Trends Biol. Sci. **15:**383–387.
8. **Corden, J. L., and M. Patturajan.** 1997. A CTD function linking transcription to splicing. Trends Biochem. **22:**413–416.
9. **Fabrega, C., V. Shen, S. Shuman, and C. D. Lima.** 2003. Structure of an mRNA capping enzyme bound to the phosphorylated carboxy-terminal domain of RNA polymerase II. Mol. Cell **11:**1549–1561.
10. **Fong, N., and D. L. Bentley.** 2001. Capping, splicing, and 3′ processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. Genes Dev. **15:**1783–1795.
11. **Greenleaf, A.** 2003. Getting a grip on the CTD of Pol II. Structure **11:**900–902.
12. **Hartzog, G. A.** 2003. Transcription elongation by RNA polymerase II. Curr. Opin. Genet. Dev. **13:**119–126.
13. **Hirose, Y., and J. L. Manley.** 2000. RNA polymerase II and the integration of nuclear events. Genes Dev. **14:**1415–1429.
14. **Howe, K. J.** 2002. RNA polymerase II conducts a symphony of pre-mRNA processing activities. Biochim. Biophys. Acta **1577:**308–324.
15. **Huelsenbeck, J. P., and F. Ronquist.** 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17:**754–755.
16. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153:**163–168.
17. **Jokerst, R. S., J. R. Weeks, W. A. Zehring, and A. L. Greenleaf.** 1989. Analysis of the gene encoding the largest subunit of RNA polymerase II in *Drosophila.* Mol. Gen. Genet. **215:**266–275.
18. **Kobor, M. S., and J. Greenblatt.** 2002. Regulation of transcription elongation by phosphorylation. Biochim. Biophys. Acta **1577:**261–275.
19. **Morris, D. P., and A. L. Greenleaf.** 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. **275:**33935–39943.
20. **Nonet, M., D. Sweetser, and R. A. Young.** 1987. Functional redundancy and structural polymorphism in the large subunit of RNA polymerase II. Cell **50:**909–915.
21. **Prelich, G.** 2002. RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. Eukaryot. Cell **1:**153–162.
22. **Stiller, J. W., and B. D. Hall.** 1998. Sequences of the largest subunit of RNA polymerase II from two red algae and their implications for rhodophyte evolution. J. Phycol. **34:**857–864.
23. **Stiller, J. W., and B. D. Hall.** 2002. Evolution of the RNA polymerase II C-terminal domain. Proc. Natl. Acad. Sci. USA **99:**6091–6096.
24. **Stiller, J. W., B. L. McConaughy, and B. D. Hall.** 2000. Evolutionary complementation for polymerase II CTD function. Yeast **16:**57–64.
25. **Strimmer, K., and A. von Haeseler.** 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13:**964–969.
26. **Verdecia, M. A., M. E. Bowman, K. P. Lu, T. Hunter, and J. P. Noel.** 2000. Structural basis for phosphoserine-proline recognition by group IVWW domains. Nat. Struct. Biol. **7:**639–643.
27. **West, M. L., and J. L. Corden.** 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. Genetics **140:**1223–1233.
28. **Zeng, C., and S. M. Berget.** 2000. Participation of the C-terminal domain of RNA polymerase II in exon definition during pre-mRNA splicing. Mol. Cell. Biol. **20:**8290–8301.