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 (YSPTSPS) 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 Y4, S7, and S5 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 (YSASPA)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 of the G domain of RBP1, 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 trans-
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 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 heptadpeptide 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 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 RPBI- 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 8 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 RPBI with WT CTD), cells were grown in 100 ml of YEPD to an optical density at 600 nm (OD600) 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 RPBI 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 RPBI sequences contain CTD-like motifs (22). Bonnemaisonia hamifera RPBI 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 RPBI 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
Glaucosphaera vacuolata is a unicellular red alga; morphologically and developmentally it is the simplest of the red algae from which RPB1 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).
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 appeared to have little effect on yeast cells (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).

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 consenssus 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.
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 mammals, a conserved structure with an essential function. Mol. Cell. Biol. 8:321–329.

Boeke, J. D., J. Trueheart, G. Natoulsis, and G. R. Fink. 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 154:164–175.

Calvo, O., and J. L. Manley. 2003. Strange bedfellows: polyadenylation factors at the promoter. Genes Dev. 17:1321–1327.

Carty, S. M., and A. L. Greenleaf. 2002. Phospho-CTD-association proteins in the nuclear precompe link transcription to chromatin modification and RNA processing. Mol. Cell. Proteomics 1:598–610.

Cho, E.-J., T. Takagi, C. R. Moore, and S. Buratowski. 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.

Greenleaf, A. 2003. Getting a grip on the CTD of Pol II. Structure 11:900–902.

Hart, G. A. 2003. Transcription elongation by RNA polymerase II. Curr. Opin. Genet. Dev. 13:119–126.

Hirose, Y., and J. L. Manley. 2000. RNA polymerase II and the integration of nuclear events. Genes Dev. 14:1415–1429.

Howe, K. J. 2002. RNA polymerase II demonstrates a dependency on pre-mRNA processing activities. Biochim. Biophys. Acta 1577:308–324.

Huiskensteel, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Ito, H., Y. Fukuda, K. Murata, and A. Kimura. 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. 153:163–168.

Jokirinter, R. S., J. L. 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.

Kohor, M. S., and J. Greenblatt. 2002. Regulation of transcription elongation by phosphorylation. Biochim. Biophys. Acta 1577:261–275.

Morris, D. P., and A. L. Greenleaf. 2001. The splicing factor, Prp40, binds to the phosphorylated carboxy-terminal domain of RNA polymerase II. J. Biol. Chem. 276:3935–3943.

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.

Prelich, G. 2002. RNA polymerase II carboxy-terminal domain kinases: emerging clues to their function. Eukaryot. Cell 1:153–162.

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.

Stiller, J. W., and B. D. Hall. 2002. Evolution of the RNA polymerase II C-terminal domain. Proc. Natl. Acad. Sci. USA 99:6061–6066.

Stiller, J. W., B. L. McConaughy, and B. D. Hall. 2000. Evolutionary complementation for polymerase II CTD function. Yeast 16:57–64.

Strimmer, K., and A. von Haeseler. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13:964–969.

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 domain motifs. Nat. Struct. Biol. 7:639–643.

West, M. L., and J. L. Corden. 1995. Construction and analysis of yeast RNA polymerase II CTD deletion and substitution mutations. Genes 14(1):223–233.

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.

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.