The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II*

Daniel P. Morris and Arno L. Greenleaf‡

From the Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710

We showed previously that the WW domain of the prolyl isomerase, Ess1, can bind the phosphorylated carboxyl-terminal domain (phospho-CTD) of the largest subunit of RNA Polymerase II. Analysis of phospho-CTD binding by four other WW domain-containing Saccharomyces cerevisiae proteins indicates the splicing factor, Prp40, and the RNA polymerase II ubiquitin ligase, Rsp5, can also bind the phospho-CTD. The identification of Prp40 as a phospho-CTD binding protein represents the first demonstration of direct interaction between a documented splicing factor and the phospho-CTD. Domain dissection studies reveal that phospho-CTD binding occurs at multiple locations in Prp40, including sites in both the WW and FF domain regions. Because the conserved repeats of the CTD make it an ideal ligand for multi-site binding events, the implications of multi-site binding are discussed. Our data suggest a mechanism by which the phospho-CTD of elongating RNA polymerase II facilitates commitment complex formation by juxtaposing the 5’ and 3’ splice sites.

The carboxyl-terminal domain (CTD) of the largest subunit of RNA polymerase II (1) plays a central role in mRNA synthesis. The CTD is composed of 26–52 heptad repeats with the consensus sequence, YSPTSPS. These repeats are extensively phosphorylated in RNA polymerase II actively engaged in transcript elongation (2–5).

In the past several years evidence has accumulated indicating that the phosphorylated form of the CTD acts to coordinate pre-mRNA processing events (for review see Refs. 6–10). The 5’ capping complex is not only localized near the pre-mRNA by direct association with the phospho-CTD (11, 12) but is also allosterically activated by phospho-CTD binding (13, 14). A variety of evidence strongly suggests that the phospho-CTD is involved in splicing. The hyperphosphorylated form of RNA polymerase II colocalizes with splicing factors when transcriptionally active (15, 16) and in the nuclear matrix (17–19). Several phospho-CTD binding proteins have been identified that contain SR and RR domains as found in many splicing factors (20, 21). Fusion proteins with a hyperphosphorylated CTD can inhibit splicing in vivo (22). In addition, transcriptionally unengaged but phosphorylated RNA polymerase II is able to stimulate splicing in vitro (23).

Recently, we demonstrated that the prolyl isomerase, Ess1, can bind the phosphorylated form of the CTD and that this binding is mediated by its WW domain (24). Because the prolyl isomerase activity of Ess1 preferentially acts on phospho-Ser-Pro peptide bonds, as are found in abundance in the phospho-CTD, our findings provided a plausible explanation for earlier results implicating Ess1 in pre-mRNA 3’ end formation (25, 26).

WW domains, named for two highly conserved tryptophan residues, are small independently folding protein domains consisting of slightly more than 30 amino acids arranged in three anti-parallel β sheets (27, 28). These domains have been shown to bind proline-rich sequences containing several different motifs (29). Of five proteins in yeast carrying the most characteristic WW domain sequences (Ess1, Rsp5, Prp40, YFL010p, and YPR152p), two are already known to bind the proline-rich CTD. As mentioned, Ess1 binds the phospho-CTD (24), whereas Rsp5, a ubiquitin ligase with specificity to RNA Pol II (30, 31), binds the unphosphorylated CTD via one or two of its WW domains (32). Because many WW domain-containing proteins are involved in RNA transcription or processing, it seemed likely that other WW domain proteins might have CTD or phospho-CTD binding capacity. To test this likelihood, we evaluated the CTD and phospho-CTD binding capabilities of the five yeast proteins mentioned above. In addition to Ess1, we found that Rsp5 and Prp40 can associate with the phospho-CTD.

The identification of Prp40 as a phospho-CTD binding protein represents the first example of direct interaction between a documented splicing factor and the phospho-CTD. The ability of Prp40 to associate with the U1 snRNP (33, 34) and to bind the branch point binding protein, Ms15, is the basis for a proposed bridging between the U1 snRNP-associated 5’ splice site and the Ms15-associated branch point and 3’ splice site (35). Thus the phospho-CTD binding ability of Prp40 suggests a role for the phospho-CTD in commitment complex formation.

Interestingly, Prp40 appears to be part of an orthologous group which includes a single protein from Schizosaccharomyces pombe, Caenorhabditis elegans, and Drosophila melanogaster and two proteins from mammals (FBP11 and HYPC). Each of these proteins has a pair of WW domains amino-terminal to a series of six FF repeats, a recently identified motif often associated with WW domains (36). One of these apparent Prp40 orthologs, FBP11, has been experimentally implicated in splicing: paralleling the Prp40 interaction with the yeast branch point binding protein, Ms15, FBP11 has been shown to bind a mammalian branch point binding protein, SF1 (37).

Concurrent with the present work, a human protein with both WW and FF domains, CA150, was independently identified as a phospho-CTD binding protein (38). In that study it was shown that sites in the FF region of CA150 bind the...
ess1 10  GLPTP WTVRY SKSKK REXFF NPETK HSQWE EPE
YFL010p 10  QVPSG WKEAVF DDEYQ TYYWY DLSTN SSLQWE PPR
Rsp5 (WW1) 230  RLPPG WERRT DN-FG RTYYV DHNTR TTWTW RPT
Rsp5 (WW2) 332  ELPSG WEQRF TP-EG RAYFV DHNTR TTTWW DPR
Rsp5 (WW3) 388  FLPSG WCMLR TN-TA RVFV DHNTR TTTWD DPR

FIG. 1. Five yeast proteins contain characteristic WW domains. The WW domains of Ess1, Rsp5, Prp40, YFL010p, and YPR152p contain both conserved Trp residues that give the domain its name, a Pro residue three amino acids after the carboxyl-most Trp, and a central set of three hydrophobic residues.

phospho-CTD. In parallel with those results, we show here that although the Prp40 WW domains can bind the phospho-CTD, at least two other phospho-CTD binding sites are present in the FF region of Prp40.

Three of the five yeast proteins with characteristic WW domains have now been shown to bind the CTD or phospho-CTD, at least in part through interactions mediated by WW domains. Because similar WW domains are found in dozens of homologs in higher eukaryotes, it seems very likely that many of these proteins will also interact directly with the CTD.

MATERIALS AND METHODS

Construction and Expression of WW Domain-containing Proteins—GST-Ess1 was constructed and expressed as previously reported (24). Similarly, other GST fusion proteins were expressed using plasmids created by ligating appropriately digested polymerase chain reaction products into the BamHI/EcoRI sites of the polylinker of pGEX2TK (Amersham Pharmacia Biotech). Expression and purification were as described previously using modified elution buffer (50 mM Tris, 10 mM reduced glutathione, 100 mM NaCl brought to pH 8.5–9 with NaOH) for all proteins (24). The amount of full size fusion protein present was crudely estimated by Coomasie staining of gels and comparison with broad range molecular mass markers (Bio-Rad).

Hyperphosphorylating the CTD-containing Fusion Proteins—Yeast CTD kinase I was used as described previously to exhaustively phosphorylate β-Gal-yCTD or GST-yCTD fusion protein (29). Each reaction was monitored to ensure quantitative shift of the fusion protein to slower mobility characteristic of hyperphosphorylated CTD.

Far Western Analysis of WW Domain-containing Proteins Using CTD and Phospho-CTD Probes—Proteins were run on 4–15% precast SDS-PAGE gels (Bio-Rad) until the dye front was 5 mm from the gel bottom. Gels were equilibrated in transfer buffer (Tris/Glycine, 20% Methanol) for 15 min at room temperature prior to transfer onto nitrocellulose (Hybond-C Extra; Amersham Pharmacia Biotech) for 2 h at 0.5 A (5 °C in a 6 liter tank). Blots were incubated overnight in clarified blocking/renaturation buffer (cBRB) with rocking (this and subsequent steps at 5 °C). cBRB was prepared by dissolving 3% nonfat dry milk in 1× PBS, 0.2% Tween-20, 5 mM NaF, 1 mM PMSF for 1 h followed by centrifugation at 10,000 × g for 5 min to remove insoluble material. Blots were probed with β-Gal-yCTD or with fully shifted phospho-β-Gal-yCTD or phospho-GST-yCTD in 10 ml of cBRB with 1× dithiothreitol for 2.5 h.

For direct probing blots were reacted with 32P-labeled phospho-CTD. These blots were then washed for 8 min with 100 ml of PBS-Tw (1× PBS, 0.2% Tween-20, 5 mM NaF, 1 mM PMSE) and three times for 8 min with 50 ml of PBS-Tw. The blots were dried and analyzed with a PhosphorImager (Molecular Dynamics).

For linked Far Western assays using either phosphorylated or unphosphorylated β-Gal-yCTD as probe, mouse anti-β-galactosidase (Sigma G 4644) was used as a 1st antibody and goat anti-mouse IgG (Fab-specific) horseradish peroxidase conjugate (Sigma A9917) was used as 2nd antibody. Blots were incubated 15 min with 1st and 2nd antibodies, whereas washes were 4 × 50 ml for 3 min. Blots were developed using SuperSignal West Pico Chemiluminescent Substrate (Pierce 34080) and Biomax film (Kodak 870 1302).

CTD and Phospho-CTD Precipitation Experiments—GST fusion proteins were exchanged into 1× PBS with 1 mM PMSE by repeated concentration and dilution in a microcon 10 unit (Amicon). Proteins were adsorbed to glutathione beads to a concentration of about 1 mg of

full sized fusion protein per ml of beads. Beads were equilibrated with prebead buffer (1× PBS, 1 mM PMSE, 0.2% Tween-20, and 5 mM NaF) and resuspended to a 50% slurry. Prebead solution (80 ml) containing both β-Gal-yCTD and 32P-labeled β-Gal-yCTD in prebead buffer was combined with 20 μl of 50% slurry of beads in 250-μl tubes. Sample tubes were incubated at 25 °C for 30 min with mixing by inversion then spun at 500 × g for 30 s to pellet the matrix. Supernatant was removed from each tube (80 μl), mixed with 2× SDS-PAGE sample buffer with 20 mM dithiothreitol (2× SB) and heated for 3 min at 95 °C. The beads and solution remaining (20 μl) were combined with 2× SB and heated for 3 min at 95 °C. All sample were vortexed and spun at 500 × g prior to loading on 4–15% precast SDS-PAGE gels (Bio-Rad). Equal amounts of supernatant and bead solutions (15 μl) were loaded; however, to account for protein dilution of these samples by the beads only 12 μl of the more concentrated prebead solution was used. Gels were stained with Coomassie, dried on filter paper, and analyzed by PhosphorImager.

RESULTS

As part of ongoing efforts to elucidate the role of CTD kinases and CTD phosphorylation in transcription-coupled events, we are identifying proteins capable of binding the CTD and phospho-CTD. In part because two of the five yeast proteins containing characteristic WW domains were CTD or phospho-CTD binding proteins (Ess1 and Rsp5), and a third (Prp40) was involved in pre-mRNA processing, we decided to test the CTD and phospho-CTD binding capability of all five of these proteins (Ess1, Rsp5, Prp40, YFL010p, and YPR152p). The WW domains of these proteins (Fig. 1) contain conserved residues found in many mammalian WW proteins. Other WW proteins are almost certainly present in yeast (e.g. Nop8), but none of these WW domains has all of the characteristic conserved residues.

To investigate CTD and phospho-CTD binding by the yeast WW proteins, we used β-Gal and GST fusion proteins containing the yCTD. These yCTD-containing fusion proteins can be quantitatively hyperphosphorylated by yeast CTD kinase I, generating a product displaying the gel mobility shift characteristic of the phospho-CTD (39). Previous quantitations lead to estimates of the number of phosphates per CTD at from 20 to 45 (40). Thus, although most repeats contain one or more phosphates, it is possible that a small number of repeats lack a phosphate.

When equal amounts of full sized GST fusion proteins for all five WW proteins were loaded on a gel, transferred to nitrocellulose, and analyzed by Far Western using 32P-labeled β-Gal-yCTD (Fig. 2) or 32P-labeled GST-yCTD (data not shown), both Rsp5 and Prp40 clearly bound the phospho-CTD, although the signal was somewhat weaker than for Ess1 (Fig. 2). This result was replicated with a linked Far Western assay (data not shown) using hyperphosphorylated β-Gal-yCTD as probe (see

2 D. Morris, unpublished results.
FIG. 2. Analysis of phospho-CTD binding by five WW domain-containing yeast proteins. Duplicate 4–15% SDS gels were run for Coomassie staining and Far Western analysis. About 500 and 50 ng of each full-length GST fusion protein (top band in each lane) was loaded as indicated. The expected masses of GST-Ess1 (45 kDa), GST-YFL010p (49 kDa), GST-Prp40 (95 kDa), and GST-Rsp5 (118 kDa) are consistent with the apparent masses observed by SDS-PAGE; however, the expected mass of GST-YPR152 (80 kDa) is slightly below its apparent mass of 95 kDa. Proteins of one gel were electroblotted to nitrocellulose and probed with hyperphosphorylated 32P-labeled β-Gal-γCTD, after which the bound label was detected by PhosphorImager. In addition to Ess1, both Prp40 and Rsp5 bound the phospho-CTD. Lane 1 contains Bio-Rad broad range markers of the masses indicated (~1 μg/band). Lane 12 contains E. coli proteins from 25 μl of saturated culture (BL21). Lane 13 contains prestained markers with the following apparent masses: 205, 113, 75, 54, 35, 28.8, 21, and 7.2 kDa.
A Splicing Factor Binds the Phosphorylated CTD

FIG. 3. Linked Far Western analysis shows Rsp5 and a fragment of YFLO10p can bind the unphosphorylated CTD. A SDS-PAGE gel loaded with 500 or 50 ng of each WW domain-containing GST fusion protein was run, and the proteins were transferred to nitrocellulose (see Fig. 1 for staining pattern). The blot was probed with β-Gal and fragments thereof present in the markers. Lane 1 contains Bio-Rad broad range markers (10 μg/marker). The signal in lane 1 is from the β-Gal and fragments thereof present in the markers. Lane 2 contains Bio-Rad prestained markers, the actual positions of which are indicated by the bars and whose apparent masses are shown at the far right. Lane 13 contains protein from 25 μl of E. coli cells (BL21).

To analyze the ability of the WW proteins to bind the unphosphorylated CTD (Fig. 3), we employed a linked Far Western system using unphosphorylated β-Gal-yCTD as probe. In agreement with published data (32), the RNA Pol II-directed ubiquitin ligase, Rsp5, bound strongly to the unphosphorylated CTD (Fig. 3, lanes 11 and 12). We note that some of the fragments of GST-Rsp5 bound by the phospho-CTD (Fig. 2, lanes 10 and 11) and the CTD (Fig. 3, lanes 11 and 12) are different. This observation argues that association with the phospho-CTD probe is not simply a consequence of binding rare unphosphorylated repeats. Perhaps significantly, full-length GST-YFLO10p did not react with the CTD, whereas a proteolytic fragment bound the unphosphorylated CTD with sufficient affinity to give a clear signal (Fig. 3, lane 5). This signal occurs under exposure conditions in which all of the E. coli proteins gave no signal (Fig. 3, lane 13). The strong signal in the unstained markers (Fig. 3, lane 1) is a useful positive control that results from the anti-β-Gal antibodies reacting with β-Gal and fragments thereof present in the markers. Under the same conditions used for Fig. 3, a control blot probed using β-Gal showed no binding to any WW domain protein (data not shown).

A recognized problem with Far Western analysis is that many protein domains will not renature after SDS denaturation while plastered to nitrocellulose. Because this appears to be the case for some domains of Prp40 (see below), we determined the ability of the CTD and phospho-CTD to bind in solution to bead-bound GST-Ess1, GST-YFLO10p, and GST-YPR152p, or GST-Rsp5 (Fig. 4). The experiment simultaneously partitioned the CTD (β-Gal-yCTD) and phospho-CTD (32P-labeled β-Gal-yCTD) between the solution and beads. In general agreement with the Far Western analysis, GST-YFLO10p and GST-YPR152p bound neither form of the CTD, whereas GST-Ess1 bound only the phospho-CTD and GST-Rsp5 bound the CTD and phospho-CTD.

To further characterize the phospho-CTD binding properties of Prp40, we expressed segments of Prp40 as GST fusion proteins (Fig. 5A). In addition to two amino-terminal WW domains, Prp40 contains a series of six FF motifs. These motifs were identified by sequence homology and named for two conserved phenylalanine residues present in most of the founding members (36). Each of the fusion proteins and full-length GST-Prp40 were analyzed by Far Western using the direct assay with 32P-labeled β-Gal-yCTD (Fig. 5B). Of the Prp40 segments, GST-WW and GST-WW/FF2 clearly bound the phospho-CTD, whereas GST-FF1/FF2 and GST-cFF apparently did not. Note that combining the WW and FF1/FF2 regions (Fig. 5B, lanes 4 and 5) significantly increased binding relative to the WW domains alone (Fig. 5B, lanes 2 and 3). Further, addition of the cFF region in the full sized Prp40 (Fig. 5B, lanes 10 and 11) resulted in another increase in signal intensity. One of the many potential explanations for this behavior is that the FF regions provide some increase in overall binding affinity even though the individual FF regions bound with insufficient affinity to generate a signal under far Western conditions. This hypothesis was made more attractive by the multiple phospho-CTD binding sites present in CA150 (38), a human homolog of Prp40. Like Prp40, CA150 contains WW domains amino-terminal to a series of FF motifs. Unlike Prp40, far Western analysis of the domains of CA150 suggested that the major phospho-CTD binding site resided in the FF region.

To test the hypothesis that the FF motifs of Prp40 were phospho-CTD binding sites by an independent method, each of the segments of Prp40 was bound to beads and in-solution binding to 32P-labeled β-Gal-yCTD measured (Fig. 6). In contrast to results from the Far Western, each of the individual segments of Prp40 bound the phospho-CTD. Most dramatically, beads with GST-cFF (Fig. 6, lane 14) were at least as effective at binding the phospho-CTD as were beads with GST-WW (Fig. 6, lane 5). Binding was specific to the phosphorylated CTD because none of the Prp40 segments bound the unphosphorylated CTD significantly. That each of these fragments has some ability to bind to the phospho-CTD demonstrates Prp40 contains at least three independent phospho-CTD binding sites. As will be discussed below, the repetitive nature of the CTD makes it ideal for multi-site binding. The failure of the FF segments to individually bind the phospho-CTD during Far Western analysis probably resulted from the failure of these motifs to renature sufficiently to generate a signal.

The finding that CA150, a mammalian protein with WW and FF domains similar to those in Prp40, also possesses multi-site phospho-CTD binding capability (38) supports the conclusion that this property is functionally significant, and it argues for some degree of mechanistic similarity between Prp40 and CA150. However, two other mammalian proteins, FBP11 and HYP, are more closely related to Prp40 than is CA150. Indeed, it is quite likely that these two mammalian proteins and Prp40 are part of an orthologous group, which also includes a single protein from S. pombe, C. elegans, and D. melanogaster (Fig. 7). All six orthologs have a distinctive and highly conserved pair of WW domains (Fig. 7B). Alignment within the FF region (Fig. 7C) suggests that all six orthologs contain six FF motifs, and each motif is part of a repeating unit about 70...
amino acids in length. Some of the FF motifs are not readily apparent in Prp40; however, comparison with the more conserved proteins of the multicellular organisms strongly suggests their presence. In addition, all of the orthologs have complete conservation of spacing between the FF motifs. Aside from the members of this group, the next closest homolog to Prp40 is CA150 and its probable orthologs in *C. elegans*, *D. melanogaster*, zebrafish, and mouse.²

**DISCUSSION**

Yeast offers a relatively simple system in which to examine the behavior of a family of domains and the proteins that contain them. Previously, two of five yeast proteins containing a WW domain similar to those found in many mammalian proteins had been identified as CTD or phospho-CTD binding proteins (32, 24). The present study documents phospho-CTD binding by Prp40 and Rsp5 as well as potential CTD binding by YFL010p. The demonstration that the WW domains of Prp40 can associate with the phospho-CTD reinforces the idea that WW domains are suited to be CTD interacting domains. However, CTD or phospho-CTD binding by a WW domain does not exclude the possibility that the domain also binds other proline-rich targets. Furthermore, in *vitro* CTD or phospho-CTD binding may not always be functionally significant. The ability of a particular WW domain to bind to any protein has to be evaluated in the context of the functional behavior of each protein. In the present context, Ess1 (25, 26), Rsp5 (30, 32, 31), and Prp40 (33–35) have all been linked to processes associated with transcription, where contact with a the CTD or phospho-CTD is likely.

The ability of Rsp5 to bind both the CTD and the phospho-CTD may have important implications for its role in degradation of RNA polymerase II. Although the observed phospho-CTD binding could conceivably be a consequence of the association of Rsp5 with rare unphosphorylated repeats in the phospho-CTD, the ability of some proteolytic fragments to bind only the phospho-CTD argues for a distinct phospho-CTD binding site. The presence of sites for binding both phosphorylated and unphosphorylated CTD repeats suggests that RNA Pol II with a partially dephosphorylated CTD may be the *in vivo* target of Rsp5.

Given the multi-site phospho-CTD binding properties of Prp40, the absence of binding by YPR152p was somewhat unexpected. YPR152p not only has a Prp40-like WW domain but also has at least one FF domain (36). Nevertheless, the presence of both a Prp40-like WW domain and an FF motif suggests this protein may have a role in splicing.

The inability of full-length YFL010p to interact with either form of the CTD is intriguing, because a less abundant fragment clearly bound the unphosphorylated CTD. The possibility that the amino acids missing from the fragment act as a negative regulatory domain is under further study.

The identification of Prp40 as a phospho-CTD binding protein could connect the phosphorylated CTD to the earliest stages of commitment complex formation and may provide a molecular basis for the increasing number of reports linking the phosphorylated CTD and splicing (15–20, 22, 23). Further, the multi-site binding observed for association of Prp40 with the phospho-CTD has far reaching implications not only for...
splicing but probably for the behavior of many phospho-CTD binding proteins. The repeated nature of the CTD makes this domain an ideal target for multi-site binding events. A formal treatment of the consequences of two proteins interacting at two sites was first done by Jencks (41) in the early 1980s and has been termed the A-B site problem. The essence of the analysis is that the first binding event localizes the second set of binding sites in each others presence. Stated alternatively, much of the entropy lost upon binding of protein A and B occurs with the first binding event resulting in less entropy loss upon the second binding event. As a consequence of the decrease in this entropic factor, the observed binding constants for multi-site interactions can be dramatically lower than would be expected from addition of the binding energies of the individual sites.

To examine the consequence of multiple repeats on the po-

---

**FIG. 5.** Far Western analysis of Phospho-CTD binding by segments of Prp40. A, an amino-terminal GST was fused to both WW domains (GST-WW), the WW domains and the first two FF motifs (GST-WW/FF2), the first two FF motifs alone (GST-FF1/FF2), or the region carboxyl-terminal to the first 2 FF motifs (GST-cFF) of Prp40. These GST fusion proteins were analyzed by Far Western using hyperphosphorylated 32P-labeled β-Gal-yCTD probe. For each construct containing a segment of Prp40, about 1000 and 200 ng of full-length fusion protein was loaded in the indicated lanes (top band in Coomassie-stained gel lanes 2–11, except for a contaminating E. coli protein in lanes 4–9 marked with asterisks). Load for the GST-Prp40 fusion protein (500 and 50 ng) was as in Fig. 1, as were loads for the markers and E. coli extract. Expected masses for GST-WW (36 kDa), GST-WW/FF2 (57 kDa), GST-FF1/FF2 (48 kDa), and GST-cFF (64 kDa) agree with the observed masses of the full-length fusion proteins.

---

**A**

|        | WW Domains | FF Domains |
|--------|------------|------------|
| Prp40  | 1-583      |            |
| WW     | 1-83       |            |
| WW/FF2 | 1-265      |            |
| FF1/FF2| 84-265     |            |
| cFF    | 267-583    |            |

---

**B**

Coomassie Stained Gel

Phosphorimage of Far Western
tential of the CTD for multi-site binding, it is instructive to use some illustrative calculations. Given the 26 repeats in the yeast CTD, the repeat motif in the CTD will achieve very high concentrations in some local environment near RNA Pol II. For instance, the diameter of a sphere containing a yeast CTD in which the seven amino acid repeats are 10 mM is 202 angstroms. Given the relatively unstructured nature of the CTD (42), it is quite possible the yeast CTD can extend to the edge of a 202-angstrom sphere because the radius of this sphere (101 angstroms) is much shorter than the fully extended length of the CTD (about 655 (3.6) angstroms). Thus in the environment around RNA Pol II even relatively weak binding sites (mM $K_d$s) may bind the CTD. Although the present data for Prp40 do not document an entropic contribution from multi-site binding, it seems very likely that constraining several binding sites in an environment with a high concentration of phosphorylated repeats will significantly contribute to phospho-CTD binding behavior of Prp40.

The presence of Prp40 in the U1 snRNP (33, 34) and its ability to interact with the branchpoint binding protein, Msl5, are the basis for the proposal that Prp40 functions as a coactivator linking the U1 snRNP at the 5' site with the Msl5/Mud2 complex at the branch point 3' splice site (35). The ability of Prp40 to bind the phosphorylated CTD may play a role in the initial steps of spliceosome assembly, perhaps by speeding association of Prp40 with the U1 snRNP or by facilitating formation of a bridging complex.

The fact that the WW domains of Prp40 can bind both the phospho-CTD and Msl5 may have important functional significance. This feature of Prp40 raises issues of interdomain binding order and specificity. For example, binding of the phospho-CTD to the WW domains might prevent premature association of Prp40 with free Msl5 prior to Msl5 binding to the 3' splice site. At this point the colocalized Msl5 may achieve a high enough local concentration to displace the phospho-CTD. It should be noted that a detailed examination of WW domain binding specificity revealed considerable promiscuity in the target sequence (43), suggesting that a single WW domain could be capable of binding the phospho-CTD and the proline-rich region of Msl5. On the other hand, it may be that the two WW domains of Prp40 have different specificity, one binding the phospho-CTD and one binding Msl5. Considerably more work will be required to elucidate the details of these interactions.

The existence of a family of proteins orthologous to Prp40 by sequence suggests these proteins and their individual domains may have conserved functions. The fact that the WW domain region of Prp40 can bind the phospho-CTD and is highly conserved in the orthologs suggests the phospho-CTD binding ability will also be conserved. Similarly, conservation of the FF region argues for conservation of multi-site phospho-CTD binding. In addition, the conservation of multi-site phospho-CTD binding in CA150 (38), which contains related WW domains and a six-repeat FF motif region, strongly suggests the Prp40 orthologs will retain phospho-CTD binding ability.

Evidence that the Prp40 orthologs play a conserved role in splicing is also available. The WW domains of Prp40 have been shown to bind directly to the branch point binding protein, Msl5 (35). Similarly the WW domains of FBP11 have been shown to bind directly to a mammalian branch point binding protein, SF1, through a PPLP motif in SF1 (37). Although this exact motif is not present in Msl5, there is a proline-rich amino-terminal region containing two PPL sequences.

The present study combined with previously published data suggests a model for phospho-CTD enhancement of splicing. The association of Prp40 with the phospho-CTD suggests that the U1 snRNP will remain associated with transcribing RNA Pol II. As a consequence, the U1 snRNP and Prp40 may be positioned to efficiently interact with Msl5 after this protein binds the branch point near the 3' splice site. In this manner, the phospho-CTD could facilitate the association of the 3' and

---

**Fig. 6.** Binding of segments of Prp40 to the phosphorylated and unphosphorylated CTD. Glutathione beads containing GST-WW, GST-WW/FF2, GST-FF1/FF2, GST-cFF, GST, or no protein were incubated with a solution containing β-Gal-yCTD and 32P-labeled β-Gal-yCTD and analyzed as for Fig. 4. See “Materials and Methods” for details.
splice sites, an obvious benefit, particularly in mammalian systems where these sites can be separated by many kilobases. A similar model has been suggested based on the ability of RNA polymerase II to enhance commitment complex formation in mammalian cell extracts (23). Clearly the mammalian homologs of Prp40 could play a central role in this process. Because commitment complex formation and splicing can be reconstituted in vitro in the absence RNA Pol II (44, 45), it seems plausible that a principal function of the phospho-CTD vis a vis splicing may be to increase the efficiency of the process rather than to perform a discrete obligatory step in the pathway. Such a facilitating role, consisting mainly of recruiting critical components to the RNA Pol II neighborhood, could certainly contribute to the speed of the overall process; it could also modulate specificity of the process, such as splice site choice. The potential that the phospho-CTD may be the landing platform

Fig. 7. Comparison of Prp40 and its probable orthologs. A, other than the WW domain (WW) and FF motif (FF) regions, no obvious homology is present in proteins from both unicellular and multicellular organisms including Saccharomyces cerevisiae (S.c.), S. pombe (S.p.), C. elegans (C.e.), D. melanogaster (D.m.), or mouse (mHYP and mFBP11). However, the proteins from multicellular organisms share a series of basic patches (+ + +) near the carboxyl terminus and proline enrichment (P) near the amino terminus. GenBank™ accession numbers for uncharacterized proteins are 2330816 (S.p.), 1353114 (C.e.), and 7295860 (D.m.). B, the WW domains (underlined) of all of the proteins show a very high level of conservation. Further, many residues between the domains are conserved, as is the interdomain spacing in five of the six proteins. C, alignment of the FF motif region indicates there are six repeating units composed of about 70 amino acids each containing an FF motif (underlined). Note that the first, second, third, and fifth FF motifs of proteins from multicellular organisms are followed by similar acidic and basic sequence elements.
for regulatory factors controlling commitment complex formation has far reaching implications.

Acknowledgment—We thank Dr. Mariano Garcia-Blanco for excellent scientific advice.

REFERENCES

1. Corden, J. L. (1990) Trends Biochem. Sci. 15, 383–387
2. Cadena, D. L., and Dahmus, M. E. (1987) J. Biol. Chem. 262, 12468–12474
3. Weeks, J. R., Hardin, S. E., Shen, J., Lee, J. M., and Greenleaf, A. L. (1993) Trends Biochem. Sci. 18, 385–387
4. O’Brien, T., Hardin, S., Greenleaf, A. L., and Lis, J. T. (1998) Trends Biochem. Sci. 23, 245–254
5. Greenblatt, J. (1997) Trends Biochem. Sci. 22, 378–382
6. Greenleaf, A. L. (1993) Trends Biochem. Sci. 18, 383–387
7. Steinmetz, E. J. (1997) Proc. Natl. Acad. Sci. U. S. A. 94, 6568–6573
The Splicing Factor, Prp40, Binds the Phosphorylated Carboxyl-terminal Domain of RNA Polymerase II
Daniel P. Morris and Arno L. Greenleaf

J. Biol. Chem. 2000, 275:39935-39943.
doi: 10.1074/jbc.M004118200 originally published online September 7, 2000

Access the most updated version of this article at doi: 10.1074/jbc.M004118200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 44 references, 22 of which can be accessed free at http://www.jbc.org/content/275/51/39935.full.html#ref-list-1