Deciphering the RNA Polymerase II CTD Code

Targeting m^{7}G capping to nascent Pol2 transcripts – The 5' m^{7}GpppN cap is a signature feature of eukaryal mRNA that is required for mRNA stability and efficient translation. Cap synthesis entails three enzymatic reactions: (i) the 5' triphosphate end of the pre-mRNA is hydrolyzed to a diphosphate by RNA triphosphatase; (ii) the diphosphate RNA end is capped with GMP by RNA guanylyltransferase; and (iii) the GpppN cap is methylated by RNA (guanine-N7) methyltransferase. The capping reactions are universal in eukarya and the formation of guanylate caps is essential in all species that have been examined genetically. We have shown how capping is directed to nascent Pol2 transcripts via physical interactions of one or more of the capping enzymes with the carboxyl-terminal domain (CTD) of the Pol2 Rpb1 subunit. The Pol2 CTD consists of tandemly repeated heptapeptides of consensus sequence Y^{1}S^{2}P^{3}T^{4}S^{5}P^{6}S^7. Phosphorylation and dephosphorylation of the Tyr1, Ser2, Thr4, Ser5 and Ser7 residues and *cis–trans* isomerization of Pro3 and Pro6 inscribe a complex "CTD code" read by diverse receptor proteins.

The Ser5-PO$_4$ "letter" of the CTD code plays a vital role in recruiting RNA guanylyltransferase (GTase) to the Pol2 elongation complex. Indeed, the direct binding of GTase to the Ser5-PO$_4$ form of Pol2 CTD is a conserved theme among diverse eukaryal taxa. Crystal structures of *Candida albicans* GTase (Cgt1), mammalian GTase (Mce1), and *Schizosaccharomyces pombe* GTase (Pce1) bound to Ser5-phosphorylated Pol2 CTD ligands illuminated how eukarya take divergent structural routes to achieve GTase•CTD interaction. These three cellular GTases are structurally homologous enzymes composed of two domains: an N-terminal nucleotidyltransferase (NTase) module containing the guanylate binding pocket; and a C-terminal OB fold module. Whereas the CTD docking sites are dominated in each case by interactions of the NTase domain with Ser5-PO$_4$ and Tyr1, the CTD segments bound to the GTases differ in their length and in the phase of the heptad sequence bound. For example: fission yeast Pce1 recognizes an 8-aa CTD segment (S$_{5P}$PSYSPTS$_{5P}$) bracketed by two Ser5-PO$_4$ marks; mammalian Mce1 captures a 6-aa segment (S$_{5P}$PSYSP); and *Candida* Cgt1 has two distinct CTD-docking sites that recognize Ser5-PO$_4$ containing heptads in differently phased registers (TS$_{5P}$PSYSP and PSYSPTS$_{5P}$P). Moreover, the atomic contacts to CTD differ among the three GTases and few of the CTD-interacting side chains are conserved. Thus, capping enzymes from different taxa have evolved unique strategies to read the same Pol2 CTD code. Our analysis of fission yeast Pce1 mutations that disrupt the Pol2 CTD interface shows that at least one of the two Ser5-PO$_4$ binding sites is required for cell viability, and that each site is important for cell growth at 37˚C.

The RNA triphosphatase (TPase) components of the mammalian and budding yeast capping apparatus are recruited passively to the Pol2 CTD, by virtue of their physical association with the GTase: in *cis* as a covalently fused TPase-GTase enzyme Mce1 in mammals or in *trans* as separately encoded subunits of a TPase•GTase complex in budding yeast. The fission yeast *S. pombe* has a distinctive strategy for targeting cap formation to Pol2 transcripts, whereby the TPase (Pct1) and GTase (Pce1) enzymes are not associated physically, but instead bind independently to the Ser5-phosphorylated Pol2 CTD.

Capping enzymes can also access nascent Pol2 transcripts via physical interactions with transcription elongation factor Spt5. Spt5 is a large polypeptide, composed of multiple domain modules, that associates with the Pol2 transcription complex shortly after initiation and can exert negative and positive effects on transcription elongation. Fission yeast Spt5 has a distinctive C-terminal repeat domain (the "Spt5 CTD"), composed of 18 repeats of a nonapeptide motif (T^{1}P^{2}A^{3}W^{4}N^{5}S^{6}G^{7}S^{8}K^9), that: (i) binds the RNA capping enzymes Pct1 and Pce1 and (ii) is targeted for threonine phosphorylation by the Cdk9 kinase. Key insights to the interaction of the capping apparatus with Spt5 were gained via the crystal structure of a fission yeast Pce1•Spt5-CTD complex (a collaboration with Chris Lima), which revealed a docking site in the OB domain of the GTase enzyme that captures the Trp4 residue of the Spt5 nonamer repeat. We found that a disruptive mutation in the Spt5-CTD binding site of Pce1 is synthetically lethal with mutations in the Pol2-CTD binding site, signifying that the Spt5 and Pol2 CTDs cooperate to recruit GTase to the transcription elongation complex *in vivo*. We showed that CTD phosphorylation has opposite effects on the interaction of Pce1 with Pol2 (Ser5-PO$_4$ is required for binding) *versus* Spt5 (Thr1-PO$_4$ inhibits binding). We proposed that the state of Thr1 phosphorylation comprises a binary "Spt5 CTD code" that is read by capping enzyme, independent of, and parallel to, its response to the state of the Pol2 CTD. We extended our analysis to the fission yeast RNA triphosphatase Pct1 and its interactions with the Spt5 CTD. We reported crystal structures of the Pct1 apoenzyme and a Pct1•Spt5-CTD complex and showed that Pct1 binding to Spt5 CTD is antagonized by threonine phosphorylation. We established by structure-guided mutagenesis the relevance of the Spt5 CTD interface to Pct1 function *in vivo*. These

results fortify our proposal of an Spt5 CTD code.

Deciphering the Pol2 CTD code in fission yeast – The informational rules that govern the CTD code on a cellular and organismal level can be probed genetically by manipulating the composition and structure of the Rpb1 CTD. *S. pombe* is an attractive model system for CTD structure-function analysis because the native heptad repeat array is relatively homogeneous *vis à vis* other taxa. The *S. pombe* CTD consists of 25 consensus heptad repeats linked to the body of Rpb1 by a "rump" segment of 4 near-repeats that deviate in size and/or sequence from the consensus heptad. Using a fully functional *S. pombe* Rpb1 with a CTD composed of the rump plus 14 consensus heptads, we gauged the importance of all coding "letters" by introducing alanines in lieu of Tyr1, Ser2, Pro3, Thr4, Ser5, Pro6, and Ser7 of every heptad of the Rpb1 CTD array. Conservative (or phosphomimetic) substitutions were tested as well. Salient findings were that: (i) Tyr1, Pro3, Ser5, and Pro6 are essential for viability, by the criterion that Ala substitution is lethal, whereas Ser2, Thr4, and Ser7 are not; (ii) Phe is functional in lieu of Tyr1. We also made double-mutants that subtracted two phosphoacceptors in each heptad and found that *S2A+S7A* and *T4A+S7A* mutants are viable. Thus, Ser5 is the only strictly essential CTD phosphorylation site in fission yeast.

"Vocabulary" and output of the CTD code – Elucidating a core CTD vocabulary necessitates answering two key questions: (i) how are essential coding letters organized into readable words? (ii) which essential coding information is read by which essential CTD receptors? We envisioned that genetics could assign an essential CTD coding letter to a specific CTD-receptor pair in the sea of available cellular receptors, if one could bypass the requirement for that letter by delivering a cognate receptor protein to the Pol2 transcription complex via other means. Translating this "thought experiment" into action, we sought to override the requirement for Ser5, and Ser5 phosphorylation, by fusing an essential cellular Ser5-PO_4 receptor – the mRNA capping enzymes RNA triphosphatase and guanylyltransferase – to the carboxyl terminus of the otherwise nonfunctional Rpb1-CTD-S5A protein. To our delight, the experiment worked, i.e., an *rpb1-CTD-S5A-MCE* fusion allele was viable. This result proved that capping enzyme recruitment is a chief function of the Ser5-PO_4 mark *in vivo*. We then asked whether the fusion maneuver could override the essentiality of Pro6 (which is a directing signal for Ser5 phosphorylation by CTD kinases). Indeed it did bypass the lethality of *P6A*. We concluded that Ser5-PO_4–Pro6 comprises an essential two-letter code word that is read by the mRNA capping apparatus. We then tested whether capping enzyme fusion could bypass the lethality of *Y1A* or *P3A* – and found that it did not (*66*). We infer that Tyr1 and Pro3 provide essential "reading material" for additional essential CTD receptors other than the capping enzymes.

Individual letters of the CTD code govern distinct gene expression programs – The fact that four of the five phosphoacceptor coding letters of the CTD heptad are not essential in fission yeast raises important questions as to whether and how these phosphate marks impact gene expression, the extreme situations being that absence of a particular CTD-PO_4 mark has little or no effect, or that loss of a coding cue does exert significant effects albeit on the expression of genes that are not essential under the laboratory conditions surveyed. To explore this issue, we applied high-throughput RNA sequencing methods (RNAseq) to gauge globally the impact of the loss of each of the four inessential CTD phosphoacceptors on gene expression. This analysis (a collaboration with Jürg Bähler's lab) illuminated how individual letters of the Pol2 CTD code affect the expression of limited and distinct sets of genes: to wit, CTD mutations *S2A*, *Y1F*, *S7A*, and *T4A* elicted ≥2-fold dysregulation of only 4.4%, 1.4%, 1.2% and 0.14% of the annotated fission yeast protein-coding RNAs, respectively. The majority of the protein-coding RNAs affected in *Y1F* cells were coordinately affected by *S2A*, suggesting that Tyr1-Ser2 constitutes a two-letter code word. *Y1F* and *S2A* elicited increased expression of genes encoding proteins involved in iron uptake (Frp1, Fip1, Fio1, Str3, Str1, Sib1, etc.), without affecting the expression of the genes that repress the iron regulon, implying that Tyr1-Ser2 transduces a repressive signal in iron homeostasis. *Y1F* and *S2A* cells had increased levels of ferric reductase activity and were hypersensitive to phleomycin, indicative of elevated intracellular iron. We identified CTD letters Thr4 and Ser7 as novel components of the fission yeast phosphate homeostatic response, on which the *T4A* and *S7A* mutations had opposing effects. *T4A* blunted expression of Pho1 acid phosphatase in response to phosphate starvation, while *S7A* de-repressed Pho1 expression under phosphate replete conditions. New experiments exploiting additional CTD mutants implicate Ser5-PO_4, Pro6, and Ser7-PO_4 as letters in a CTD code word that represses the expression of phosphate acquisition genes (*pho1*$^+$ and *pho84*$^+$) under phosphate-replete conditions. This work highlights the fission yeast iron and phosphate regulons as outstanding models to dissect how CTD phospho-sites, interfacing with specific yeast transcription factors, control gene expression in response to environmental changes (iron and PO_4 availability).

McCracken, S., Fong, N., Rosonina, E., Yankulov, K., Brothers, G., Siderovski, D., Hessel, A., Foster, S., Shuman, S., and Bentley, D.L. (1997) 5' Capping enzymes are targeted to pre-mRNA by binding to the phosphorylated C-terminal domain of RNA polymerase II. *Genes Dev*. 11, 3306-3318.

Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998) The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J. Biol. Chem.* 273, 9577-9585.

Ho, C.K., and Shuman, S. (1999) Distinct roles for CTD Ser2 and Ser5 phosphorylation in the recruitment and allosteric activation of mammalian capping enzyme. *Molecular Cell* 3, 405-411.

Schroeder, S., Schwer, B., Shuman, S., and Bentley, D. (2000) Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev*. 14, 2435-2440.

Pei, Y., Hausmann, S., Ho, C.K., Schwer, B., and Shuman, S. (2001) The length, phosphorylation state, and primary structure of the RNA polymerase II carboxyl-terminal domain dictate interactions with mRNA capping enzymes. *J. Biol. Chem*. 276, 28075-28082.

Pei, Y., and Shuman, S. (2002) Interactions between fission yeast mRNA capping enzymes and elongation factor Spt5. *J. Biol. Chem*. 277, 19639-19648.

Hausmann, S., and Shuman, S. (2002) Characterization of the CTD phosphatase Fcp1 from fission yeast: preferential dephosphorylation of serine 2 versus serine 5. *J. Biol. Chem.* 277, 21213-21220.

Hausmann, S., and Shuman, S. (2003) Defining the active site of *Schizosaccharomyces pombe* CTD phosphatase Fcp1. *J. Biol. Chem.* 278, 13627-13632.

Pei, Y., Schwer, B., and Shuman, S. (2003) Interactions between fission yeast Cdk9, its cyclin partner Pch1, and mRNA capping enzyme Pct1 suggest an elongation checkpoint for mRNA quality control. *J. Biol. Chem.* 278, 7180-7188.

Fabrega, C., Shen, V., Shuman, S., and Lima, C.D. (2003) Structure of an mRNA capping enzyme bound to the phosphorylated carboxyl-terminal domain of RNA polymerase II. *Molecular Cell* 11, 1549-1561.

Pei, Y., and Shuman, S. (2003) Characterization of the *Schizosaccharomyces pombe* Cdk9/Pch1 protein kinase: Spt5 phosphorylation, autophosphorylation and mutational analysis. *J. Biol. Chem*. 278, 43346-43356.

Hausmann, S., Erdjument-Bromage, H., and Shuman, S. (2004) *Schizosaccharomyces pombe* carboxyl-terminal domain (CTD) phosphatase Fcp1: distributive mechanism, minimal CTD substrate, and active site mapping. *J. Biol. Chem.* 279, 10892-10900.

Hausmann, S., Schwer, B., and Shuman, S. (2004) An *Encephalitozoon cuniculi* ortholog of the RNA polymerase II carboxyl-terminal domain (CTD) serine phosphatase Fcp1. *Biochemistry* 43, 7111-7120.

Schroeder, S. C., Zorio, D.A.R., Schwer, B., Shuman, S., and Bentley, D. (2004) A function of yeast mRNA cap methyltransferase, Abd1, in transcription by RNA polymerase II. *Molecular Cell* 13, 377-387.

Koiwa, H., Hausmann, S., Bang, W.Y., Ueda, A., Kondo, N., Hiraguri, A., Fukuhara, T., Bahk, J.D., Yun, D.J., Bressan, R.A., Hasegawa, P.M., and Shuman, S. (2004) *Arabidopsis* CPL1 and CPL2 are essential Ser5-specific CTD phosphatases. *Proc. Natl. Acad. Sci. USA* 101, 14539-14544.

Suh, M. H., Ye, P., Zhang, M., Hausmann, S., Shuman, S., and Fu, J. (2005) Fcp1 directly recognizes the CTD and also interacts with a site on RNA polymerase II distinct from the CTD. *Proc. Natl. Acad. Sci. USA* 102, 17314-17319.

Hausmann, S., Koiwa, H., Krishnamurthy, S., Hampsey, M., and Shuman, S. (2005) Different strategies for carboxyl-terminal domain (CTD) recognition by Serine5-specific CTD phosphatases. *J. Biol. Chem.* 280, 37681-37688.

Pei, Y., Du, H., Singer, J., St. Amour, C., Granitto, S., Shuman, S., and Fisher, R.P. (2006) Cdk9 of fission yeast is activated by the CDK-activating kinase Csk1, overlaps functionally with the TFIIH-associated kinase Mcs6, and associates with the cap methyltransferase Pcm1 *in vivo*. *Mol. Cell. Biol.* 26, 777-788.

Ghosh, A., Shuman, S., and Lima, C.D. (2008) The structure of Fcp1, an essential RNA polymerase II CTD phosphatase. *Molecular Cell* 32, 478-490.

Schwer, B., Schneider, S., Pei, Y., Aronova, A., and Shuman, S. (2009) Characterization of the *Schizosaccharomyces pombe* Spt5-Spt4 complex. *RNA* 15, 1241-1250.

Schneider, S., Pei, Y., Shuman, S., and Schwer, B. (2010) Separable functions of the fission yeast Spt5 CTD in capping enzyme binding and transcription elongation overlap with those of the RNA polymerase II CTD. *Mol. Cell. Biol*. 30, 2353-2364.

Ghosh, A., Shuman, S., and Lima, C.D. (2011) Structural insights to how mammalian capping enzyme reads the CTD code. *Molecular Cell* 43, 299-310.

Schwer, B., and Shuman, S. (2011) Deciphering the RNA polymerase II CTD code in fission yeast. *Molecular Cell* 43, 311-318.

Schwer, B., Sanchez, A.M., and Shuman, S. (2012) Punctuation and syntax of the RNA polymerase II CTD code in fission yeast. *Proc. Natl. Acad. Sci. USA* 109, 18024-18029.

Doamekpor, S.K., Sanchez, A.M., Schwer, B., Shuman, S., and Lima, C.D. (2014) How an mRNA capping enzyme reads distinct RNA polymerase II and Spt5 CTD phosphorylation codes. *Genes & Development* 28, 1323-1336.

Doamekpor, S.K., Schwer, B., Sanchez, A.M., Shuman, S., and Lima, C.D. (2015) Fission yeast RNA triphosphatase reads an Spt5 CTD

code. *RNA* 21, 113-123.

Schwer, B., Ghosh, A., Sanchez, A.M., Lima, C.D., and Shuman, S. (2015) Genetic and structural analysis of the essential fission yeast RNA polymerase II CTD phosphatase Fcp1. *RNA* 21, 1135-1146.

Schwer, B., Sanchez, A., and Shuman, S. (2015) RNA polymerase II CTD phospho-sites Ser5 and Ser7 govern phosphate homeostasis in fission yeast. *RNA* 21, 1770-1780.

Chatterjee, D., Sanchez, A.M., Goldgur, Y., Shuman, S., ands Schwer, B. (2016) Transcription of lncRNA *prt*, clustered *prt* RNA sites for Mmi1 binding, and RNA polymerase II CTD phospho-sites govern the repression of *pho1* gene expression under phosphate-replete conditions in fission yeast. *RNA* 22, 1011-1025.

Schwer, B., Sanchez, A.M., Garg, A., Chatterjee, D., and Shuman, S. (2017) Defining the DNA binding site recognized by the fission yeast Zn2Cys6 transcription factor Pho7 and its role in phosphate homeostasis. *mBio* 8, e01218-17.

Sanchez, A.M., Shuman, S., and Schwer, B. (2018) Poly(A) site choice and Pol2 CTD Serine-5 status govern lncRNA control of phosphate-responsive *tgp1* gene expression in fission yeast. *RNA* 24, 237-250.

Garg, A., Sanchez, A.M., Shuman, S., and Schwer, B. (2018) A long noncoding (lnc) RNA governs expression of the phosphate transporter Pho84 in fission yeast and has cascading effects on the flanking *prt* lncRNA and *pho1* genes. *J. Biol. Chem*. 293, 4456-4467.

Project Members

Angad Garg

About Us

Overview

Leadership

Administration

History

Contact Us

Research

Overview

Research programs

Research labs

Core facilities & resources

Education & Training

Overview

Postdoctoral training

Gerstner Sloan Kettering Graduate School

Joint graduate programs

Programs for college & high school students

News & Events

Overview

Seminars & events

Open Positions

Overview

Faculty positions

Postdoctoral positions
