THE ROLE OF LEDGF/p75 IN TRANSCRIPTIONAL REGULATION

Jeffrey R. Kugelman

Department of Biological Sciences

APPROVED:

Manuel Llano, M.D. Ph. D., Chair

Max Shpak, Ph.D.

Kristine Garza, Ph.D.

Naijun Sha, Ph.D.

Kyle Johnson, Ph.D. (Advocate)

Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

THE ROLE OF LEDGF/p75 IN TRANSCRIPTIONAL REGULATION

by

JEFFREY R. KUGELMAN, B.S., M.I.T.

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at El Paso
in Partial Fulfillment
of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Department of Biological Sciences
THE UNIVERSITY OF TEXAS AT EL PASO
December 2010

Acknowledgements

I would like to thank my children, Jander and Joelle, for providing inspiration. Your love has kept me going through some very difficult times. Daddy loves you! Para Johanny, gracias por todo. Te adoro mi amor. To my friends and family for providing emotional support, drinking companions and sounding boards. I sincerely hope to repay you all.

To my mentor Dr Llano I would like to say thank you for your guidance both in work and in life. Thank you for pushing me to be better than I was. My thanks to all the members of my committee, and the UTEP faculty at large, for providing the opportunity to pursue this goal.

Thanks to my fellow cadets and ROTC faculty who have provided motivation and to the US Army and HHMI for financial support. To all the members of the Llano lab you have my thanks for your help and company. The following were instrumental in teaching me techniques I needed to prosecute this body of work:

Dr Kyle Johnson and John Roskoff:	RNA isolation and gel electrophoresis.
Dr Carolina Lema:	Microarray analysis.
Dr Armando Varela:	Confocal Microscopy.
Johanny Tonos De Leon:	Densitometry.
Dr Manuel Llano:	Everything else.

This work was supported by grant 1SC2GM082301-01 from the National Institute of General Medical Sciences (NIGMS), National Institutes of Health (NIH) (to ML); a University Research Institute grant from the University of Texas at El Paso (UTEP) (to ML); the Howard Hughes Medical Institute (HHMI) mentorship supported by grant 52005908 (to JRK), and UTEP core facilities funded by the BBRC grant 5G12RR008124.

Abstract

The Lens Epithelial Derived Growth Factor p75 (LEDGF/p75) is a chromatin bound protein whose cellular function is not yet clearly known. A role in transcriptional regulation had been previously proposed based on its interaction with the basal transcriptional machinery and on its effects on the expression of genes involved in the cellular response to environmental stresses. To further elucidate the function of LEDGF/p75, we conducted a global and unbiased evaluation of the role of this protein in gene expression. To that aim, we performed a microarray analysis of cellular gene expression in cells that are severely depleted of LEDGF/p75. To minimize cell type-specific observations, we used three different LEDGF/p75 deficient human cell lines: embryonic kidney epithelial cells (HEK293) and two different lines of CD4+ T cells (SupT1 and Jurkat). By taking the intersection of the three data sets, using a fold-change of greater than two and a student t-test confidence value of 90% we have identified a group of genes whose expression is dysregulated in at least two of the cell types tested.

The potential role of LEDGF/p75 in transcriptional regulation of responsive genes has been described as operating at the level of their promoters as a general co-activator, connecting members of the basal transcriptional apparatus to DNA sequence specific transcriptional activators. However, the observation that LEDGF/p75 is bound to sites within the chromatin across the genome suggests that this protein is also present inside actively transcribed genes. In correlation with this observation, LEDGF/p75 has been demonstrated to promote integration of cDNA copies of the HIV-1 genome within actively transcribed genes. This localization of LEDGF/p75 inside actively transcribed genes suggests a possible role of LEDGF/p75 in transcriptional elongation. To further investigate whether LEDGF/p75 has a role in

transcriptional elongation, we performed a transcriptional profile analysis of genes we previously identified as dysregulated in SupT1 and HEK293 cells. We used chromatin immunoprecipitation and quantitative real time PCR analysis to demonstrate that LEDGF/p75 is located on the intragenic regions of these genes as well as their promoters. LEDGF/p75 occupancy on these genomic locations correlated with the gene transcriptional activity. We also demonstrate that LEDGF/p75 coimmunoprecipitates with CDK9 and the FACT complex, two separate components of the transcriptional elongation complex, and corroborated this association by quantitative confocal colocalization.

In order to understand the mechanism of LEDGF/p75 on transcriptional regulation at the promoter level, we studied the effect of LEDGF/p75 on the shared bi-directional promoter of the LEDGF/p75-regulated genes DTX3L and PARP9. Our data indicate that LEDGF/p75 negatively regulates the expression of this promoter. Two known motifs within LEDGF/p75, the HIV-1 Integrase binding domain and the PWWP domain, were both required for this inhibitory effect. These data suggest that the effect of LEDGF/p75 on the transcription of the DTX3L/PARP9 genes at the chromatin level may involve not only the promoter regions but other genetic elements as well. Alternatively, LEDGF/p75 could require the interaction with other cellular factors that are limiting in cells that are over expressing LEDGF/p75.

LEDGF/p75 has been reported to be involved in the adaptation to cellular stress in cells over expressing this protein; however, our data indicate that under basal endogenous conditions LEDGF/p75 does not significantly affect the regulation of genes involved in stress response. To clarify this potential protective effect we studied the capability of LEDGF/p75-deficient cells to adapt to environmental stresses. Our data indicate a protective effect of LEDGF/p75 to these stresses being necessary for cell viability under these conditions. To further understand this

effect of LEDGF/p75 on cellular response to stress we determined the differential distribution of LEDGF/p75-interactor proteins involved in transcriptional regulation in response to cellular insults. We demonstrate that LEDGF/p75 influences the subcellular distribution of SSRP1 and CDK9 in response to environmental stresses. LEDGF/p75 was required to mediate recruitment of both SSRP1 and CDK9 to transcriptionally active complexes following these cellular insults.

In summary, our data indicate that LEDGF/p75 interacts with components of the basal transcription machinery and of the transcriptional elongation complex, suggesting that this protein regulates the transcription of LEDGF/p75-responsive genes at the initiation and elongation steps of transcription. Additionally, LEDGF/p75 participates in the recruitment of the transcriptional elongation factors SSRP1 and CDK9 to chromatin under basal conditions and in response to cellular stresses.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Gene Expression

The first step of gene expression, wherein RNA copies of the DNA are synthesized, is Transcription. The RNA products are subsequently utilized for the production of proteins in the follow on process of Translation. Transcription is broadly defined in three phases: initiation, elongation (including: promoter clearance, proximal pausing and productive elongation) and termination (Fig. 1) [1].

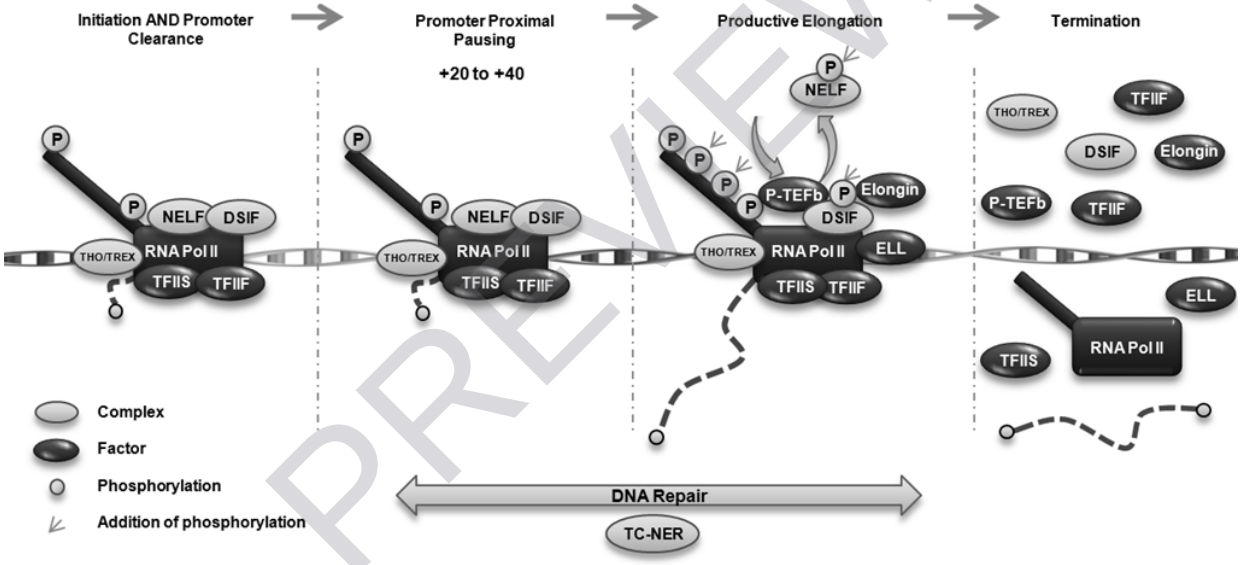

Figure 1. The phases of transcription.

This diagram illustrates the three primary phases of transcription: Initiation, Elongation and Termination. It also clarifies the transition from initiation into productive elongation via promoter clearance and proximal pausing. Factors chosen for this diagram are used to illustrate elongation and are not comprehensive.

Prior to initiation factors are bound to recruit the pre-initiation complex including the RNA polymerase II (PolII). PolII requires ATP hydrolysis for remodeling of the pre-initiation complex by TFIIH. During remodeling of the pre-initiation complex approximately11-15 DNA base pairs at the transcription start site are unwound, allowing single-stranded DNA to enter the active site of PolII and initiate transcription [2, 3]. After initiation the transcriptional complex

disassociates with promoter-sequence elements and some promoter bound factors in promoter clearance [4]. During this step, the association between the transcription complex and the nascent RNA are strengthened and the process is considered complete when it is associated stably [1]. The next step shown in Figure 1 boundaries the transition between initiation and productive elongation is proximal pausing. Proximal pausing is a regulatory step that allows for maturation of the transcriptional elongation complex and rate-limiting control of transcriptional output [5]. This process is thought to be regulated by the presence of negative elongation factor (NELF) and DRB sensitivity-inducing factor (DSIF) [6, 7]. The DSIF complex is made up of the subunits Spt4 and Spt5 [8, 9] and NELF contains four subunits NELF A, B, C/D and E [10].

Release from pausing is initiated by the addition of positive transcription-elongation factor B (P-TEFb) to the transcriptional complex [9]. This complex is made up of two subunits CDK9 and Cyclin T and has kinase activity phosphorylating Rbp1 at the Serine 2 position. Rbp1 is a subunit of PolII containing a heptapeptide repeat that is a defining characteristic differentiating this polymerase from RNA polymerases I and III. Additionally, P-TEFb phosphorylates NELF and DSIF [11, 12]. At this point NELF disassociates from the complex and productive elongation can begin [7]. Of note DSIF, while repressing elongation in the presence of NELF, once phosphorylated and absent of NELF is thought to promote elongation [6, 8, 13]. Factors shown to aid in relief from transcriptional pausing are transcription factors IIF and IIS (TFIIF and TFIIS), Elongin and Eleven-nineteen lysine rich in leukemia (ELL) [14, 15]. Elongation specific factors, ELL and Elongin, can be found when coimmunoprecipitated with phosphorylated PolII and are recruited to heat shock loci [16, 17].

An important pathway involved with the transcriptional complex is the transcriptional coupled repair (TC-NER) pathway [18]. The mechanism involved in this pathway is poorly

understood in eukaryotes yet it is important for correcting DNA lesions and cross-linking found during transcription and associates transiently with the transcriptional complex. Several factors that have been identified in this pathway are: Cockayne syndrome B protein (CSB) a putative transcription repair coupling factor [19], a structure specific DNA endonuclease XPG [20], the PolII subunits Rbp4 and 9 [21, 22], the THO/TREX complex that functions to keep mRNA from binding to DNA to template inappropriately [23, 24], PAF and Ccr4-Not required for efficient TC-NER activity [18].

In eukaryotes, PolII transcription is terminated when the transcript is cleaved. This cleavage of the new transcript is followed by polyadenylation. Polyadenylation is the addition of multiple adenines to the 3' end of the transcript. The cleaved 3' product is degraded as well as any un-polyadenylated transcripts [1]. This releases the transcript and the transcriptional complex and its constituent factors are then disassociated from each other and from the DNA. It has been demonstrated that RNA processing mechanisms such as: splicing, capping, cleavage, polyadenylation, are linked to the transcription-elongation complex [25].

The process of transcription is being actively studied to identify new factors that have roles in regulation, pausing, and repair. The Lens epithelium Derived Growth Factor (LEDGF/p75) has been described as co-transcriptional activator linking sequence specific transcription factors to the basal transcription machinery [26]. All the sites of interaction described for LEDGF/p75 transcriptional regulatory activities have been derived from promoter interactions yet the mechanism of these interactions are still poorly understood.

LEDGF/p75 Cellular Role

LEDGF/p75 is a member of the hepatoma derived growth factor family. The Hepatoma Derived Growth Factor (HDGF) family of genes is named after its founder member. There are

currently 6 members in this family of genes including: HDGF, LEDGF/p75 and its splice variant p52, and Hepatoma Derived Growth Factor Related Proteins (HRP) 1, 2, 3 and 4. Defining characteristics for membership in this gene family are an N-terminal PWWP domain (pro – trp – trp – pro motif) which is responsible for DNA binding [27-29] and nuclear sub-cellular localization.

HDGF

HDGF has been shown to induce growth factor activity when over-expressed endogenously or added exogenously [30-32]. Involvement in early tissue development has been demonstrated for several organs including: intestine, kidney, liver and the cardio vascular system [30, 33-36]. Following injury HDGF is involved in tissue repair of lung, vascular and colon tissues by promoting proliferation [37, 38]. Another role of HDGF has been described in cell survival as cell lines depleted for HDGF were more prone to induction of apoptosis [39, 40]. This protein is well studied for its involvement in carcinogenesis demonstrating an increased expression profile in multiple tumor types [41, 42] and its involvement in metastasis makes HDGF over-expression a marker for poor prognosis [28, 43-48]. Recently, a study of HDGF over-expression effects on gene regulation was completed in mouse primary aortic vascular smooth muscle cells. They found that it functioned largely as a transcriptional repressor in 66 genes and activated a small group of 9 genes. Their dataset was validated by qPCR for two groups of genes that are involved in cardiovascular development and transcriptional regulation [49] tying this dataset to known protein function. On casual inspection there seems to be no overlap between this dataset and microarray data available for Human LEDGF/p75 depletion.

HRP-1, 3 and 4

Beyond their localization and growth factor activity very little functional data exists for these proteins. HRP-1 and 4 are only expressed in the testis and 3 are limited to the nervous system [50-52]. No gene regulation studies have been conducted for these proteins.

HRP-2

Expressed in many cell types this protein is the only other protein in the known Human proteome to contain an IBD [50]. In Vitro both HRP-2 and LEDGF/p75 have the ability to bind HIV-1 Integrase and potentiate integration [53]. An analysis of function In Vivo, however, demonstrates that something prevents this protein from binding to chromatin sufficiently to support HIV-1 infection despite its structural similarities to LEDGF/p75 and ability to translocate to the nucleus [54]. One theory supported by the discrepancy is that HRP-2 contains a much higher percentage of Serines, often in repeats. This would lend the protein to be much more heavily phosphorylated than LEDGF/p75. Perhaps a kinase important to the endogenous post-translational conditioning of HRP-2 is missing in *E.coli.* Whether the inability of HRP-2 to bind to chromatin is due to targeting, binding of additional factors masking the PWWP, folding conformation or Phosphorylation of binding domains is unknown. No studies have been completed to date on HRP-2's effect on gene regulation but due to its poor chromatin association homologous function to LEDGF/p75 may be difficult to recreate.

LEDGF

Two splice variants, LEDGF/p75 and p52, are generated from the PSIP1 gene [55]. LEDGF proteins are ubiquitously expressed although their relative abundance varies in a tissue specific manner. These proteins are tightly bound to chromatin during all the phases of the cellular growth cycle. Chromatin binding is essential for a role of these proteins as general transcriptional co-activators [56]. LEDGF proteins contain an identical N-terminal region (325