The Pennsylvania State University
The Graduate School
Department of Biochemistry, Microbiology and Molecular Biology

GENOME-WIDE STUDIES OF TRANSCRIPTIONAL RESPONSES IN YEAST

(Saccharomyces cerevisiae)

A Dissertation in
Biochemistry, Microbiology and Molecular Biology

by

Sujana Ghosh

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2013
The dissertation of Sujana Ghosh was reviewed and approved* by the following:

B. Franklin Pugh
Verne M. Willaman Professor of Molecular Biology
Dissertation Advisor
Chair of Committee

David S. Gilmour
Professor of Molecular and Cell Biology

Joseph C. Reese
Professor of Biochemistry and Molecular Biology

Ross C. Hardison
T. Ming Chu Professor of Biochemistry and Molecular Biology

Debashis Ghosh
Professor of Statistics

Scott B. Selleck
Professor and Head of Biochemistry and Molecular Biology

*Signatures are on file in the Graduate School
ABSTRACT

Cells are known to respond to changing environmental conditions by regulating the expression of genes. Regulation of transcription itself is a complicated process. It involves contributions from various cellular activators and repressors towards enhancing the accessibility of an otherwise repressive chromatin to the transcription factors and RNA Pol II for transcription initiation. Till date, a lot of expression studies have been done using *Saccharomyces cerevisiae*, as a model organism to investigate the role of different environmental conditions on cellular expression. Such studies have led to the identification of the environmental stress response (ESR) genes which consists of a common group of genes that are either up-regulated or down-regulated in response to a variety of environmental conditions. In addition to the ESR, the conditions also trigger stress-specific responses.

Expression studies provide valuable information in terms of which genes are expressed or repressed. But attempts to actually dissect how similar or varied the roles of transcription factors (specially assembly of the pre-initiation complex-PIC) in response to diverse conditions are, have not been made. Thus with that objective in mind the question I have tried to address in this thesis is when the genome is re-programmed by two different stress conditions does it use similar or varied PIC assembly pathways. Using moderate resolution tiling arrays (ChIP-chip) I mapped occupancy changes of factors involved in different stages of PIC assembly in
response to DNA damage and heat shock. Surprisingly I found that heat shock inducible genes recruited both TFIID and SAGA co-activator pathways simultaneously. Whereas MMS inducible genes recruit SAGA first, followed by TFIID, suggesting a sequential pathway. Thus although both the conditions recruited both SAGA and TFIID, the timing of the implementation of those complexes varied which could be primarily due to the differences in the kinetics/dynamics of the two stress conditions.

Amongst all the classes of genes in yeast, RP (ribosomal protein) genes are known to be the most highly regulated. They are also the group of genes that get down regulated in unison in response to environmental changes. *S. cerevisiae* has 137 RP genes that encode for 78 ribosomal proteins. 59 out of these 78 are encoded by duplicate genes (paralogs), which are virtually identical. Given the coordinate regulation of RP genes in response to growth conditions the question I addressed in my thesis is how redundant is the transcriptional regulation of the RP paralogs. In particular, I examined the binding of seven known RP gene regulators (Ifh1, Fhl1, Hmo1, Sfp1, Rap1, Gcn4, Abf1) and Crt1 (novel regulator of RP genes) using ChIP-exo. My results suggest that RP paralogs bind the factors differentially and most of these factors track together as a huge complex at RP promoters (localized around -200 to -400 bp upstream of TSS) except for Crt1 which is more localized in the coding region of RP genes. Interestingly, Crt1 was found to play a role in down-regulation of the RP genes. Additionally majority of the factors were found to have
multiple binding sites (>4) in the RP genes that could play a role in cooperative binding or release from the RP genes, for coordinate regulation in response to growth cues.

Taken together, the work in this thesis has made some initial attempts in dissecting the PIC assembly pathways in response to different stress conditions. It also instigates the concept of a large multi-subunit complex working at the RP genes along with the differential binding of factors at its paralogous pairs.
# TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................... ix

LIST OF ABBREVIATIONS........................................................................................................ viii

ACKNOWLEDGEMENTS ........................................................................................................... x

Chapter 1 Introduction ‘Genomic responses and transcriptional regulation’ ............ 1

1.1 A constant struggle for survival ....................................................................................... 1
1.2 The Environmental Stress Response (ESR) ................................................................ 3
1.3 Comparison of heat shock vs DNA damage .................................................................. 4
1.4 Overview of gene expression ........................................................................................... 6
  1.4.1 Orchestration (Role of sequence specific regulators and their cis-regulatory elements) .............................................................................................................. 9
  1.4.1.1 Cis-regulatory elements (CREs) ........................................................................ 9
  1.4.1.2 Trans-factors bind to the cis-regulatory elements (sequence specific regulators) .................................................................................................................. 11
  1.4.2 Access: Chromatin structure, co-repressors and co-activators (chromatin modifiers and chromatin remodelers) ................................................................. 15
  1.4.2.1 Chromatin structure ............................................................................................ 15
  1.4.2.2 Chromatin remodelers and modifiers .................................................................. 18
  1.4.3 General transcription factors (GTFs) and RNA Pol II ............................................ 22
  1.4.4 RNA Polymerase II ................................................................................................ 24
1.5 References ......................................................................................................................... 26

Chapter 2 Sequential recruitment of SAGA and TFIID in a genomic response to DNA damage in *Saccharomyces cerevisiae* ........................................................................... 51

2.1 Summary ........................................................................................................................ 51
2.2 Introduction ...................................................................................................................... 52
2.3 Results ............................................................................................................................. 57
  2.3.1 Differential recruitment of TFIID to MMS-induced vs. heat shock-induced promoter ................................................................. 57
  2.3.2 Unexpected mobilization and retention of factors in response to stress that are not coupled to a transcriptional response ................................................. 67
  2.3.3 Crt1 is not linked to MMS-induced gene ................................................................. 70
  2.3.4 Chromatin remodelers are present constitutively at DNA damage inducible genes .................................................................................................................. 73
  2.3.5 MMS-induced accumulation of Gcn4 at amino acid biosynthetic genes ................................................................. 75
2.4 Discussion ......................................................................................................................... 78
2.4.1 Serial vs parallel implementation of the SAGA and TFIID PIC assembly pathways .................................................................78
2.4.2 Apart from RNR genes, Crt1 does not play a widespread role at DNA damage inducible genes ..................................................................................................................81
2.4.3 Constitutive presence of chromatin remodeling complexes ..........82
2.5 Materials and methods ..............................................................................................83
2.6 References ..................................................................................................................89

Chapter 3 ChIP-exo mapping of Ribosomal Protein (RP) gene factors suggests differential binding at paralogous RP genes and potential role of Crt1 in down-regulation of RP genes ..............................................................105

3.1 Summary ..................................................................................................................105
3.2 Introduction .................................................................................................................106
3.3 Results ......................................................................................................................110
  3.3.1 Crt1 binds to 1535 Pol II genes genome-wide including 88 ribosomal protein genes suggesting a potential role in global gene regulation ........................................................................................................110
  3.3.2 Crt1 is linked to down-regulation of RP genes ......................................................114
  3.3.3 RP gene regulators Abf1, Fhl1, Ifh1, Sfp1, Hmo1, Rap1, Gcn4 bind at the promoters of RP genes as opposed to Crt1, which binds primarily at the coding region .................................................................116
  3.3.4 Differential binding of RP gene regulators observed at paralogous RP genes with no robust correlation with RP gene promoter activity or TFIIB occupancy ........................................................................................................119
  3.3.5 RP factors co-cluster at promoters of RP genes within a spatial distance of <20 bp of each other .................................................................................................................................122
  3.3.2 Selective motif usage by RP factors at RP genes ....................................................126
3.4 Discussion ..................................................................................................................129
  3.4.1 Crt1 negatively regulates ribosomal protein genes, not DNA-damage inducible genes.................................................................................................................................129
  3.4.2 Non-redundancy of ribosomal protein genes ........................................................130
3.5 Materials and methods ..............................................................................................133
3.6 References ..................................................................................................................138

Chapter 4 Discussion .....................................................................................................148

4.1 Summary of this study ..............................................................................................148
  4.1.1 Stress-specific transcriptional responses ..............................................................148
  4.1.2 Non-redundant regulation of ribosomal protein genes......................................149
4.2 Future directions ........................................................................................................151
4.3 References ..................................................................................................................154
LIST OF ABBREVIATIONS

ChIP  Chromatin ImmunoPrecipitation
CSR   Common Stress Response
DNA   DeoxyriboNucleic Acid
ESR   Environmental Stress Response
GTF   General Transcription Factor
KEGG  Kyoto Encyclopedia of Genes and Genomes
MIPS  Munich Information center for Protein Sequences
Pol II RNA Polymerase II
PIC   Pre Initiation Complex
SGD   Saccharomyces Genome Database
TSS   Transcription Start Site
LIST OF FIGURES

Figure 1.1 Comparison of heat shock vs DNA damage (Methylmethane sulfonate) .........................................................................................................................................................................................5

Figure 1.2 Overview of transcription initiation.............................................................................................................................8

Figure 2.1 Changes in factor occupancy in response to 30 min MMS treatment in comparison to a heat shock response..................................................................................................................58

Figure 2.2 Factor occupancy levels at RNR genes and MMS-induced genes in comparison to heat shock induced genes...................................................................................................................60

Figure 2.3 Delayed acquisition of TFIID at MMS-induced genes.................................................................63

Figure 2.4 Simultaneous acquisition of TFIID and SAGA at heat shock induced genes as opposed to MMS-induced genes................................................................................................................66

Figure 2.5 Occupancy levels at genes that are repressed or unaffected by MMS or heat shock..........................................................................................................................68

Figure 2.6 Crt1 is not linked to MMS induced genes..................................................................................71

Figure 2.7 Constitutive presence of many chromatin remodelers at DNA damage inducible genes..........................................................................................................................74

Figure 2.8 Increased occupancy of Gcn4 and SAGA at MMS-induced amino acid biosynthetic genes..........................................................................................................................76

Figure 3.1 Crt1 binds to 1535 Pol II encoded genes including 88 ribosomal protein genes suggesting a potential role in global gene regulation.................................112

Figure 3.2 Crt1 attenuates expression of Ribosomal Protein genes ..........................................................115

Figure 3.3 Crt1 binds to coding regions of RP genes as opposed to the bonafide RP factors..................................................................................................................................................118

Figure 3.4 Differential binding of RP factors at RP paralogous pairs..........................................................120

Figure 3.5 RP factors co-cluster at promoters of RP genes within a spatial distance of <20 bp of each other ..................................................................................................................124-125

Figure 3.6 Selective motif usage by RP factors at RP genes.................................................................127
ACKNOWLEDGEMENTS

As I think about my PhD journey and its culmination I am reminded of the help and support of many people that made this dream a reality. It is not possible to mention everybody but here are a few that played a very significant role.

**Microarray experiments and data analysis:** I would like to thank Dr Craig Praul and the Penn State Microarray facility for the help provided during generation of the microarray data. My sincere thanks to Bryan Venters for teaching me how to collect and analyze both expression microarrays as well as ChIP-chip assays. I am also grateful to Tom Koerber for helping me learn the nuances of data analysis using Excel. I am also thankful for the bioinformatics support provided by Zhenhai Zhang, Cizhong Jiang and Shin Wachi.

**ChIP-exo data collection and analysis:** I would like to take this opportunity to specially thank HoSung Rhee for helping me with the ChIP-exo procedure/downstream data analysis and sharing his expertise. I am also really indebted to the help provided by Kuangyu Yen for data analysis. She was always the go to person when I had problems in the usage of scripts used for bioinformatics analysis.

**Thesis Committee:** I thank the committee members for their time, support and helpful suggestions and discussions related to the work presented in the thesis.
Pugh Lab: Work presented here would not have been possible without the help and support of my advisor Frank Pugh. I am really grateful for the opportunity he gave me to do some cool science. There is a lot I have learned from him. Lab meetings were wonderful but equally stressful, as one always want to do better each time. It honed my presentation skills and also gave me the confidence to speak publicly. The comments and criticisms have helped shape the scientist in me. I also thank the Pugh Lab for creating such a wonderful working atmosphere. The lab truly became my second home and I have some very fond memories of the time spent with the lab members. I also want to specially thank Christine Walsh for managing the lab so well. Whatever I needed Christine made sure I had it.

State College friends: In the last six years spent in state college I had the privilege of knowing some wonderful people who have now become friends for life. I am truly grateful of the love, help and support I received (specially during my pregnancy) from Nitin Kumar, Meenakshi Singh, Janani Iyer, Megha Wal, Kiran Batta, Saikat Chowdhury and Kris Baker.

Family: What I am and have achieved so far would not have been possible without the unconditional love and support of my parents. They are truly the reason behind all my accomplishments. I dedicate this thesis to them. I am also grateful for the support I received from my in-laws during the completion of my PhD. Last but not least I am grateful of Daipayan, my loving husband who had to bear the brunt of the
rigors of grad school along with me. Thank you for your patience, support and for bearing with all my tantrums. I love you and you are responsible for giving me the biggest accomplishment of my life, our son Devdutt (Gutu). You two make my life complete.

Finally, thank you Lord for everything.
Chapter 1

1 Introduction: Genomic responses and transcriptional regulation

1.1 A constant struggle for survival

‘A constant struggle, a ceaseless battle to bring success from inhospitable surroundings, is the price of all great achievement’- Orison Swett Marden

In the environment, unicellular organisms face a myriad of conditions forcing them to constantly adapt to those changing conditions. Multi-cellular organisms have a slight edge over unicellular organisms whereby they can maintain the homeostasis of their internal system with the help of their tissues or organs and overcome drastic changes in the environment. They have evolved to be able to absorb the shock efficiently. On the other hand for unicellular organisms every moment is a constant struggle to ensure that they maintain their internal milieu for optimal growth. It is a do or die situation. Microorganisms like yeast have been known to be able to survive constant fluctuations in availability of both the kind and amounts of nutrients, changes in temperature, osmolarity, pH of their environment, oxidative exposure, presence of toxic chemicals, ionic /non ionic radiations, competition with other microbes. The question that has been investigated for a long time is how does cell respond to such challenges? Regulation of gene expression is one of the key players in this process. In response to such sudden change in conditions, yeast cells
invoke a multilayered response whereby they immediately shut down (or down-regulate) the most energy consuming processes such as ribosome biogenesis along with concomitant activation of various signal transduction pathways that in turn control global and some gene specific regulatory pathways thereby mounting an effective response. Yeast contains about 6000 genes and it mounts a response by controlling expression of those genes. Although other layers also play an equal role such as changes in cell surface proteins, post-translational modifications of proteins, activation of signaling molecules, etc. It therefore involves a coordinate effort of all the different forms of cellular regulation of which regulation of gene expression will be the primary focus in this thesis.

In Chapter 1 of this thesis, I introduce the process of environmental stress response and describe the stages of transcription initiation, which relies heavily on the environment for cues to either activate or repress genes. Chapter 2 is a comparison of specific transcriptional responses arising from two different stress conditions, heat shock and DNA damage. Chapter 3 puts emphasis on ribosomal protein genes, which are the most highly stress-regulated group of genes. In this chapter, attempts have been made to study the co-ordinate role of the major sequence specific regulators of RP genes. It also includes the study of Crt1, the known repressor of the model DNA damage genes, which was found to play a role in global gene regulation in *S. cerevisiae*, in addition to its role at RP genes.
1.2 The Environmental Stress Response (ESR):

In the late 1990s with the advent of DNA microarrays (Ferea et al., 1999; Lashkari et al., 1997; Schena et al., 1995), many studies investigated the effects of changes in a variety of environmental conditions at the expression level (Causton et al., 2001; Gasch et al., 2001; Gasch et al., 2000; Jelinsky and Samson, 1999) in *Saccharomyces cerevisiae*. These studies monitored the level of transcript abundance, which is the net result of a combination of rate of transcript synthesis and rate of its degradation, known as steady state transcript levels. The outcome of such study involving the reorganization of gene expression led to the identification of the Environmental Stress Response (ESR) in yeast, known to affect ~ 14% of the yeast genome which comprises of roughly 900 genes (Causton et al., 2001; Gasch et al., 2000). The 900 ESR genes in turn can be categorized into two groups based on their expression profiles: group one consisting of around 600 genes whose expression decreased in response to the stress conditions (known as the repressed or down-regulated genes) and group two consisting of around 300 genes whose expression increased (known as induced or up-regulated genes) in response to the stress conditions (Gasch and Werner-Washburne, 2002). These two groups of ESR genes display identical responses but of the opposite magnitude suggesting that they are part of the same cellular response. One important characteristic of the ESR is that cells respond to changes in the environment based on the degree or magnitude of the stress. This was identified through a series of dose-dependent experiments whereby
cells exposed to harsher changes showed a larger ESR response compared to cells exposed to milder changes (Gasch et al., 2000; Jelinsky et al., 2000). Also once cells are brought back to their normal growth conditions after exposure to stressful conditions they respond by showing reciprocal changes in the ESR genes (Gasch et al., 2000).

1.3 Comparison of heat shock vs DNA damage:

Why the need to study the transcriptional mechanisms in response to different stresses when they all invoke a common ESR response? For example, one needs to go from point A to point B. They can use different routes and different modes of transportation to reach the destination (point B). Although the final goal here is to reach point B but the means to the end can vary. Similarly, different stresses utilizes its own specific factor, like Yap1 for oxidative stress (Kuge and Jones, 1994), Hsf1 for heat shock, (Wu, 1995) Hot1/Sko1 for osmotic shock (Rep et al., 1999) to sense the stress. Thus different transcription factors sense the changes in the environment brought about by different conditions, which in turn could orchestrate the transcriptional regulation. In addition to specific factors, there are general factors like Msn2 and Msn4 that produce a more global effect under a variety of stress conditions although they do show stress specific effects on different subsets of genes as well (Martinez-Pastor et al., 1996; Schmitt and McEntee, 1996). After the initial sensing of stress, cells then utilize specific signaling pathways like
Mec1 pathway for DNA damage (Elledge, 1996; Sanchez et al., 1996; Weinert et al., 1994), Hog1 pathway for osmotic stress (Schuller et al., 1994) for relaying the information to downstream regulators. This again could lead to stress specific

![Graph showing gene expression changes for heat shock and DNA damage](image)

**Figure 1.1 Comparison of heat shock vs DNA damage (Methylmethane sulfonate)** (modified from Gasch review 2000). Average gene expression changes of induced and repressed ESR genes (Gasch et. Al 2000 and 2001)

mechanisms of transcriptional regulation. Just like there are general stress responsive factors (Msn2, Msn4), there are general signaling pathways like TOR and protein kinase A pathway which play a role in a variety of conditions (Schmelzle and Hall, 2000; Thevelein and de Winde, 1999). Thus even though a degree of commonality exists with different conditions, the cell manages to respond specifically. This specificity comes from how the cell senses, responds and passes the signals in response to different conditions.
With this premise in mind I decided to investigate transcriptional responses triggered by DNA damage & heat shock. Although both heat shock and DNA damage trigger the ESR, there are some fundamental differences in the two conditions, and thus were chosen for comparing transcriptional regulation (Gasch and Werner-Washburne, 2002). As seen in Fig1.1, kinetics of the two stresses differ considerably, where heat shock is transient (response peaks at 15 min and then drops precipitously), DNA damage is more sustained (response continues > 2hrs). Also there are differences in the dynamic range of the response. Heat shock has a higher dynamic range compared to DNA damage. Importantly, heat shock is a more general stress condition given the fact that it up-regulates many metabolic enzymes and increases the growth rate of the cell. MMS treatment (for DNA damage) on the other hand triggers cell cycle arrest, DNA damage repair and cell death (involving the Mec1 pathway). It also leads to methylation of numerous cellular targets that is Mec1 independent (Gasch et al., 2001).

1.4 Overview of gene expression

Gene expression involves synthesis of RNA from DNA by the protein RNA Polymerase. One of the most crucial strategies used to regulate this process is by controlling the rate of transcription of a gene. Thus the synthesis of the optimum amount of the right transcript, at the right time, at the right place needs to be tightly
coordinated. There are other stages involved in the control of gene expression as well, such as processing of the RNA transcripts, splicing events, export to the cytoplasm, control of RNA stability and degradation, translation of the RNA. Although these events are equally important, they will not be discussed here as it is not relevant to the work presented in this thesis.

In eukaryotes, there are three different DNA-dependent RNA polymerases (RNA Polymerase I, II and III) each involved in the synthesis of different RNAs (Sentenac, 1985). RNA Pol II is primarily involved in synthesis of mRNAs (protein coding genes) and small nucleolar RNAs (sno RNAs). Pol I is responsible for synthesis of rRNA and Pol III is responsible for tRNAs, 5S rRNA and some small non-coding RNAs. Plants on the other hand have an additional RNA Polymerase IV, involved in synthesis of small interfering RNAs (si RNAs). In this chapter, the regulation of mRNA synthesis by RNA Pol II will be discussed.

Gene expression is a complex, highly regulated, multi-step process. During this process RNA Pol II not only associates with the DNA it transcribes but also with a number of other cellular proteins that play a role in its regulation. The whole process of transcription initiation (Fig 1.2) can be broadly classified into three stages: (a) orchestration: involves the binding of sequence specific factors at their cognate sequences in response to cellular signals (b) access : involves the recruitment of coactivators or corepressors which also include the chromatin
modifying enzymes, chromatin remodeling enzymes that can covalently modify and mobilize the nucleosomes and alter the structure of the chromatin (c) assembly of the PIC (Pre-Initiation Complex) consisting of RNA Pol II and the general transcription factors (GTFs).

Figure 1.2 Overview of transcription initiation (modified from Venters et.al, 2011)

In this chapter, I have discussed in detail the regulation of the above-mentioned steps involved in transcription initiation with more emphasis on stages relevant to the work in this thesis. Since the model organism I used for the studies described in this thesis is *Saccharomyces cerevisiae*, the focus is primarily on proteins present in yeast.
1.4.1 Orchestration: role of sequence specific regulators and their cis-regulatory elements

1.4.1.1 Cis-regulatory elements (CREs)

A cis-regulatory element (CRE) is a short consensus sequence in the DNA that binds corresponding trans-acting factors (sequence specific regulators) thereby playing a role in regulation of transcription. Based on the number and the kinds of CRE’s genes respond to different signals or cues from the environment primarily mediated by the different trans-factors binding at those sites. One of the common CRE’s is the TATA-box or TATA-like element present at the promoter of most of the genes (Rhee and Pugh, 2012). A promoter of a gene can be defined as a minimal stretch of DNA sequence that can direct the initiation of transcription by RNA Pol II machinery. Accurate initiation of transcription from the core promoter involving RNA Pol II and its associated factors comprising the basal transcription factors is an essential regulatory step in the process of transcription (Butler and Kadonaga, 2002; Orphanides and Reinberg, 2002).

In yeast, the promoter is much simpler consisting of the core promoter and promoter proximal elements. The core promoter spans 60 bp around the TSS (transcription start site, denoted as +1) and includes the TATA box (or TATA-like element). The TATA box is the region that binds TBP (TATA-box binding
protein)(Burley and Roeder, 1996; Roeder, 1996). Only a small percentage of the yeast genome has a highly conserved TATA box consensus whereas majority of the genes have a TATA box with 1 or 2 mismatches and are known as the TATA-like elements(Rhee and Pugh, 2012).

In addition to the core promoter, there are promoter proximal elements in yeast which are regulatory sequences (Smale and Kadonaga, 2003) and they are known as the UAS (upstream activating sequence) or the URS (upstream repressing sequences) depending on whether they activate or repress the genes. The upstream regulatory sequences usually provide binding sites for a number of sequence specific regulators and are usually located 150-200 bp upstream of the TSS (Harbison et al., 2004; Lee et al., 2002) of genes. It is these sequence specific regulators that take the various cues from signaling pathways and orchestrate condition specific or cell-type specific expression patterns at the genes.

In addition to core promoter elements like the TATA-box, there are other elements, namely the Initiator (Inr), the TFIIB recognition element (BRE), downstream promoter element (DPE) and motif ten element (MTE) present in higher eukaryotes but not in yeast(Burke and Kadonaga, 1996; Butler and Kadonaga, 2002; Smale and Kadonaga, 2003). Ribosomal protein genes in Drosophila and mammals contain an additional TCT (polypyrimidine initiator) motif encompassing the TSS, not found in yeast(Parry et al., 2010).
Thus, although it seems that all core promoters function similarly, there is considerable diversity in its structure and function, which directly affects the transcriptional readout of the genes.

1.4.1.2 Trans-factors bind to the cis-regulatory elements (sequence specific regulators)

Sequence specific regulators are known to orchestrate the process of transcription initiation by binding to specific CRE’s and activating downstream regulatory events (Kadonaga, 2004). Some of the earliest work with sequence specific regulators goes back to 1978, when Tjian reported the sequence specific DNA binding of SV40 large T-antigen (Tjian, 1978). Similar studies reported binding of TFIIIA (Engelke et al., 1980), Sp1(Specificity factor 1) (Dynan and Tjian, 1983) using techniques such as DNase I footprinting assays (Galas and Schmitz, 1978). Some of the early yeast regulatory factors which turned out to be sequence specific factors and are still highly worked on were Gal4 (Bram and Kornberg, 1985) and Gcn4 (Hope and Struhl, 1986). One striking characteristic of these sequence specific regulators is that they are mostly modular proteins consisting of a number of domains or modules, first observed in Gal4, (Brent and Ptashne, 1985). Some of the common domains include the DNA binding domain, trans-activation domain and dimerization domains. The DNA binding domain recognizes specific DNA sequences or motifs in the DNA
thereby leading to site-directed and gene specific recruitment. Common signature motifs in the DNA binding domain that assist in their binding to the DNA include helix-turn helix, zinc finger domains, basic leucine zipper and basic helix loop helix domains. The activation domain affects transcription by mediating protein-protein interactions with the transcription machinery directly or with co-activators/co-repressors. Unlike DNA binding domains not much structural information is available for the activation domain. Lastly, the dimerization domain primarily mediates protein-protein interaction leading to formation of hetero or homo-dimers needed for essential functioning of the protein.

In yeast, roughly 120 sequence specific regulators play a role in controlling gene-specific expression patterns (Harbison et al., 2004). These factors can be broadly classified as activators or repressors based on if they increase or decrease the rate of transcription by controlling the interaction of Pol II and GTF machinery with the core promoter. These factors also recruit chromatin remodeling and modifying enzymes (co-activators or co-repressors) that help to alter the chromatin structure thereby enhancing/blocking the accessibility of the template DNA to the transcription machinery.

The sequence specific regulators themselves in turn are regulated at various levels such as; their rate of synthesis (Gcn4 synthesis is repressed under normal growth conditions but gets up-regulated in response to nitrogen or amino acid starvation
(Hinnebusch, 2005; Natarajan et al., 2001)), localization in the cell (Msn2 is a classic example, whereby it is present in the cytoplasm under normal growth conditions and translocated to the nucleus in response to stress conditions,(Gorner et al., 1998)) post-translational modifications (Mig1 rapidly moves in to nucleus in presence of glucose and moves out in its absence mediated by Snf1 dependent phosphorylation (Santangelo, 2006)), all of which affects its role in initiating transcription.

Although a lot of work on individual sequence specific regulators helped us in the better understanding of specific mechanisms involved in transcription, genome-wide global themes of their role in regulation of the transcriptional network came from some pioneering work from the Young lab in early 2000 (Harbison et al., 2004; Lee et al., 2002). Many insightful observations were revealed in those studies, which led to the development of the transcriptional regulatory network or code in the yeast genome. The studies primarily highlighted the different modes of sequence specific factor binding at genes. They showed how some genes can have one binding site for a factor (Gcn4), some have repetitive motifs (Mbp1, Gln3), some genes have multiple regulatory binding sites and some bind co-occurring regulators present in pairs (Mcm1,Ste12 or Ino2/Ino4). Additionally they tested the binding of factors in a variety of growth conditions thereby showing environmental-specific use of these regulatory factors and classified them broadly into four binding patterns: ‘condition invariant’ (eg:Leu3), that binds to similar subset of genes irrespective of the growth
conditions; ‘condition enabled’ (eg:Msn2), that does not bind at detectable levels under normal growth conditions but binds to many genes under different growth conditions (say, nutrient starvation) and is due to its exclusion from the nucleus under normal growth conditions; ‘condition expanded’ (eg:Gcn4), binds to more number of genes under amino acid starvation condition as opposed to normal growth condition; ‘condition altered’ (eg:Ste12): binds to a completely different subset of genes in one growth condition as opposed to another.

Thus it seems that sequence specific regulators really set the stage for the process of transcription and play one of the most important roles in fine tuning condition or cell-type specific transcription of genes. It is therefore not surprising that they are sometimes called the ‘Master regulator of the cell’.
1.4.2. Access: Chromatin structure, co-repressors and co-activators
(chromatin modifiers and chromatin remodelers)

1.4.2.1 Chromatin structure:

Within the nucleus the DNA is present in a complex with proteins and RNA called
the chromatin (Kornberg, 1974; Kornberg and Lorch, 1999). The basic repeating
unit of chromatin is the nucleosome and when viewed under the microscope it looks
like the classic ‘beads on a string’ (Thoma et al., 1979). Each nucleosome consists of
147bp of DNA wrapped 1.65 times around a histone core consisting of two copies
each of the histone H2A, H2B, H3 and H4 (Luger et al., 1997; Richmond and Davey,
2003). The N-terminal tails of the histone subunits themselves add another layer of
regulation to the maintenance of the chromatin structure by protruding from the
surface of the nucleosomes and serving as docking sites for gene regulatory proteins.

In the early 1980s, the role of chromatin structure in transcription was
underappreciated. Studies of transcription factors seemed more attractive
compared to nucleosomes, which were thought to be required only for packing the
DNA into the nucleus. This started changing when studies showed that actively
transcribed genes (which were devoid of nucleosomes) were more nuclease
sensitive compared to inactive genes. Also, the establishment of in vitro studies of
transcription factors on chromatin templates lead to more interesting views on the
role of chromatin in transcription. Some of the earliest evidence on role of nucleosome in expression changes came from studies on the GAL and PHO genes in yeast (Almer and Horz, 1986; Almer et al., 1986; Fedor and Kornberg, 1989; Lohr, 1984, 1997; Lohr and Lopez, 1995). These studies showed that nucleosomes block transcription factors from binding to the promoters when genes are repressed, and that they are removed upon induction. More studies with model genes such as SUC2 (Hirschhorn et al., 1992; Matallana et al., 1992), HIS3 (Iyer and Struhl, 1995), RNR3 (Li and Reese, 2001; Sharma et al., 2003; Zhang and Reese, 2007; Zhang and Reese, 2004) further increased our understanding of chromatin structure and its role in transcription.

Studies on key inducible genes helped to focus the limelight on nucleosomes, thereby leading to the beginning of an era of genome-wide study of the chromatin structure in the 2000s. Using high-resolution genome-wide nucleosome mapping methods, themes in nucleosome organization began to emerge (Jiang and Pugh, 2009a, b; Rando and Chang, 2009; Rando and Winston, 2012). Some of those organization patterns observed in _S. cerevisiae_ were; Firstly, nucleosome occupancy is generally low at promoters, and the transcription termination sites of genes and these regions form the 5’NFR and the 3’NFR (nucleosome free region). Secondly, the 5’NFR in turn is flanked by two highly positioned nucleosomes, the -1 nucleosome (-300 to -150 bp relative to TSS) and the +1 nucleosome covering the TSS, thereby controlling the access of the TSS to the general transcription machinery. The -1
nucleosome is usually a target of a number of regulatory proteins and plays a critical role in the regulation of transcription (Venters and Pugh, 2009; Venters et al., 2011). Thirdly, each nucleosome on average is spaced from the next by 18 bp of linker DNA (Albert et al., 2007; Mavrich et al., 2008). Fourthly, the +1 nucleosome and +2 nucleosome (found downstream of +1) contain the H2A variant, H2AZ (Albert et al., 2007; Guillemette et al., 2005; Raisner et al., 2005) and once into the body of the gene, the nucleosome positions become fuzzier followed by the 3’NFR at the end of the gene which is where Pol II terminates transcription. Given the theme of nucleosomal organization, questions addressing the plasticity of such organization led to studies showing that the primary organization remains constant even when cells undergo a developmental process like sporulation (Zhang et al., 2011). But nevertheless study of the chromatin structure has been important as both nucleosome positioning and occupancy control transcription and is an important step in the regulation of gene transcription (Li et al., 2007).

Now chromatin is mostly considered repressive due to nucleosome organization and is known to block not only PIC assembly (hence transcription) (Narlikar et al., 2002; Struhl, 1999) but also other DNA-template mediated processes like replication, recombination and repair (Ehrenhofer-Murray, 2004). DNA within the nucleosome occludes the CRE’s of the sequence specific regulators and is one of the modes of regulating gene expression.(Almer et al., 1986; Liu et al., 2006; Sekinger et al., 2005; Shivaswamy et al., 2008). To enhance the accessibility of those sequences
for the factors and the transcription machinery the cell relies on the role of the chromatin remodelers and modifiers.

1.4.2.2 Chromatin remodelers and modifiers:

Four families of chromatin remodelers are known in yeast; SWI/SNF (switching defective, sucrose non-fermenting) family, ISWI (imitation switch) family, CHD (chromodomain, helicase, DNA binding) family and INO80 (inositol requiring 80) family (Clapier and Cairns, 2009). All the remodelers are multi-protein complexes (except for Chd1) sharing the ATPase domain and utilize the energy from ATP hydrolysis to reposition, slide, evict or alter nucleosomes (Boeger et al., 2004; Cairns, 2009; Clapier and Cairns, 2009; Fazzio and Tsukiyama, 2003; Hartley and Madhani, 2009; Whitehouse et al., 2007). With the chromatin structure being mostly repressive as discussed previously, remodelers are indispensible in the regulation of transcription. Most of the chromatin remodelers themselves lack DNA sequence specificity and are usually recruited by sequence specific factors and co-regulators (co-activators and co-repressor) to genes following cues provided by the environment (ref).

Although ATP-dependent chromatin remodelers share some common properties such as affinity for the nucleosome, presence of ATPase domain, domains that recognize covalent modifications on histones, interactions with other transcription
factors; they differ in the mechanism of remodeling the chromatin structure and also whether they lead to activation or repression of the target genes.

Of all the remodelers, SWI/SNF was one of the first chromatin remodelers characterized in *Saccharomyces cerevisiae* (Cairns et al., 1994; Cote et al., 1994; Peterson and Herskowitz, 1992; Sudarsanam et al., 2000; Sudarsanam and Winston, 2000; Wu and Winston, 1997). They are recruited to promoters primarily by activator proteins (Neely et al., 2002; Yudkovsky et al., 1999) and its bromodomain stabilizes the binding with the acetylated histones (Hassan et al., 2001). Genome-wide expression studies have shown that SWI/SNF regulates 6% of the yeast genome, Although SWI/SNF complexes are known to be activators, they do act as repressors at some genes (not been shown) (Martens and Winston, 2002; Sudarsanam et al., 2000). Additionally they have been shown to play a role in regulation of ribosomal protein genes and the heat shock response genes (Shivaswamy and Iyer, 2008). The SWI/SNF family in yeast also consists of the RSC (remodel the structure of chromatin) complex. Both the complexes regulate non-overlapping set of genes (Wang, 2003). Unlike SWI/SNF, RSC is essential for cell viability and its concentration is ten times more in the cell compared to SWI/SNF (Cairns, 1998, 2009; Cairns et al., 1998; Flaus et al., 2006; Lorch et al., 1998) and it acts at many Pol III promoters as well (Danelin et al., 2002; Ng et al., 2002). The SWI/SNF family favors ‘histone eviction’ as its main mode of chromatin remodeling. Like SWI/SNF, RSC can both activate and repress transcription (Ng et al., 2002).
Unlike histone eviction preferred by the SWI/SNF family, ISWI family ‘slides’ histones along the DNA thereby re-positioning them (Whitehouse and Tsukiyama, 2006). Another feature of the ISWI is the presence of SANT and SLIDE (SANT-like domains) domains (Grune et al., 2003). ISWI family primarily consists of Isw1 (Vary et al., 2003) and Isw2 (Goldmark et al., 2000), where Isw1 itself is present in two forms. Although SWI/SNF family acts both as activators and repressors of transcription, ISWI family primarily acts as a repressor (Goldmark et al., 2000; Kaplan et al., 2003; Zhang and Reese, 2004). Genome-wide studies involving Isw2 has shown that it binds 20% of Pol II genes and to tRNA genes as well (Whitehouse et al., 2007). Additionally, studies of nucleosome positioning in isw2Δ strains have shown it to represses antisense transcription by positioning nucleosomes on the cryptic TSS’s (Whitehouse et al., 2007). Isw1 on the other hand plays a role in positioning nucleosomes on the coding region of genes, as seen by a shift of nucleosomes from 3’ end to 5’end of genes in isw1Δ mutant strain(Gkikopoulos et al., 2011; Tirosh et al., 2010).

CHD family is characterized by the presence of the chromodomain motif. Chd1 is one of the less characterized remodelers and in yeast it functions as a monomer (Tran et al., 2000; Woodage et al., 1997). Recent studies suggests the presence of both SANT and SLIDE domains in the DNA binding domain of Chd1(Ryan et al.,
Functionally, Chd1 is known for both assembly and spacing of nucleosomes in vitro (Lusser et al., 2005; Robinson and Schultz, 2003; Stockdale et al., 2006).

INO80 family of remodelers are known to have a characteristic split ATPase domain due to the presence of an insert (Bao and Shen, 2007). INO80 was initially identified as a coactivator of genes involved in inositol metabolism (Ebbert et al., 1999). It is known to be involved in other cellular events like DNA repair as well as replication (Morrison et al., 2004; Papamichos-Chronakis and Peterson, 2008; van Attikum et al., 2004; Vincent et al., 2008). Like ISWI family, INO80 plays a role in maintenance of the spacing between nucleosomes (Udugama et al., 2011). Swr1 also belongs to the same family as INO80 and is part of the bigger complex SWR-C. Swr1 uses energy from ATP hydrolysis for deposition of H2A.Z (H2A histone variant) at the promoter regions (Krogan et al., 2003), thereby altering the properties of the chromatin facilitating its disassembly during gene activation (Krogan et al., 2004). Unlike the other remodelers it alters the nucleosome composition as opposed to the eviction, sliding or spacing of nucleosomes.

Other than the chromatin remodelers, the second major classes of proteins used by the cell to overcome the obstacles provided by the chromatin are the chromatin modifiers. The chromatin modifiers include varied groups of proteins that decorate the surface of the nucleosomes with different kinds of modifications leading to either loosening up or compaction of the chromatin structure thereby regulating
transcription. At least eight different kinds of modifications (acetylation, methylation, phosphorylation, ubiquitylation, Sumoylation, ADP ribosylation, proline isomerization, deimination) are known thus far (Kouzarides, 2007). Although the role of chromatin modifiers is equally important if not more than the other factors regulating transcription, it will not be discussed further as it is not related to the work presented in this thesis.

1.4.3 General transcription factors and RNA Polymerase II:

Once orchestration (binding of sequence specific factors thereby ensuing action of the chromatin modifiers and remodelers) leads to increased accessibility of the underlying DNA at the promoters, the stage is now set for the recruitment of the general transcription factors (GTFs) and RNA Pol II thereby initiating the assembly of the pre-initiation complex (PIC). These general transcription factors were first identified by biochemical fractionation of nuclear extracts. They are broadly known as the GTFs as they play a role in transcription from all Pol II promoters and therefore constitute a part of the basic transcription machinery. The GTFs include TBP, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH (Orphanides and Reinberg, 2002).

PIC assembly at promoters consists of two distinct co-activator pathways involving TFIID and SAGA (Huisinga and Pugh, 2004). SAGA complex was identified in yeast as a multi-subunit complex comprising of about fifteen non-essential and six essential
subunits (Baker and Grant, 2007; Grant et al., 1997). The Spt3 and Spt8 subunit of SAGA help in the recruitment of TBP to the promoters (Bhaumik and Green, 2002). DNA microarray studies have shown that SAGA is preferentially required for transcription of approximately 10% of the yeast genes, which are mostly the TATA-box containing, stress-regulated inducible genes (Huisinga and Pugh, 2004).

Like SAGA, TFIID is also a multi-subunit complex. It is composed of TBP (TATA-binding protein) and TFIID associated factors called TAFs (Dynlacht et al., 1991; Kokubo et al., 1993; Moqtaderi et al., 1996; Reese et al., 1994). In yeast, there are 14 TAFs and some of the TAFs overlap with the SAGA complex (Green, 2000; Tora, 2002). Unlike SAGA, TFIID is known to preferentially regulate ~90% of the yeast genes that are mostly housekeeping genes which lack the canonical TATA box, but recently have been found to have TATA-like elements (Huisinga and Pugh, 2004; Rhee and Pugh, 2012). TFIID also plays a major role in regulation of ribosomal protein (RP) genes, although RP genes are highly occupied by both SAGA and TFIID (Ohtsuki et al., 2010). Thus although the basic classification of SAGA and TFIID dependent genes exist, one must keep in mind that these demarcations are not absolute and there are many genes that rely on both the complexes. TBP, a constituent of both TFIID and SAGA is a saddle shaped protein, which binds to the minor groove of DNA thereby bending it by 90°. This bend in the DNA helps to melt the promoter DNA and facilitates initiation.
The current view of assembly of the PIC involves many ordered steps (Orphanides et al., 1996). The first step is the binding of TFIID including TBP at the promoter (TATA/TATA-like elements) thereby creating a platform for the other GTFs and RNA Pol II. TFIIA is recruited which helps to stabilize the interaction of TBP with DNA. TBP then binds to TFIIB forming the TBP-TFIIA-TFIIB complex at the promoters (Hahn, 2004; Liu et al., 2010; Nikolov et al., 1995; Tan et al., 1996). TFIIB is known to play a role in selection of the transcription start site (Pardee et al., 1998). TFIIF and Pol II are then recruited to the TFIID-TBP-TFIIA-TFIIB complex followed by the binding of TFIIE and TFIIH, that bind downstream of RNA Pol II (Orphanides et al., 1996). Binding of these factors thus completes the formation of the PIC. Once the PIC is assembled, the helicase subunit of TFIIH thereby melts the promoter DNA. TFIIE helps to regulate the helicase activity of TFIIH. Once the DNA is melted, the non-template strand is captured by TFIIF and passed down the active site of RNA Pol II; thereby leading to transcription initiation.

RNA Polymerase II is a multi-subunit complex consisting of 12 subunits. The C-terminal domain (CTD) of yeast RNA Pol II contains 26 hepta-peptide repeats (Corden, 1990). During the transcription cycle CTD undergoes phosphorylation at the serine residue 2 and 5 in the hepta-peptide repeat. Variation in the phosphorylation status of the serine residues catalyzed by TFIIH (Kin28) and Ctk1 help in the transition from initiation to elongation phase of transcription (Feaver et al., 1994; Lee and Greenleaf, 1989).
In addition to Pol II many other factors and complexes play a role in initiation as well as elongation phase of transcription. Factors such as the Mediator, FACT, Paf, TFIIS, Spt4/5 all play equally critical roles in transcription (Biddick and Young, 2005; Mason and Struhl, 2003; Pokholok et al., 2002; Wery et al., 2004), but will not be discussed here further given the scope of this thesis.
1.5 References:


Msn4p are required for transcriptional induction through the stress response element (STRE). EMBO J 15, 2227-2235.


mutation on Taf2 occupancy and provides indirect evidence for different TFIID conformations at different promoters. Nucleic Acids Res 38, 1805-1820.


Chapter 2

Sequential recruitment of SAGA and TFIID in a genomic response to DNA damage in *Saccharomyces cerevisiae*

Work presented in this chapter was previously published as “Sequential recruitment of SAGA and TFIID in a genomic response to DNA damage in *Saccharomyces cerevisiae*. Sujana Ghosh and B.F.Pugh, Mol. Cell Biol. 2011 Jan; 31 (1): 190-202 and is reprinted here with permission.

2.1 Summary:

Eukaryotic genes respond to their environment by changing the expression of selected genes. The question I address here is whether distinct transcriptional responses to different environmental signals elicit distinct modes of assembly of the transcription machinery. In particular, I examined transcription complex assembly by the stress-directed SAGA complex vs the housekeeping assembly factor TFIID. I focussed on genomic responses to the DNA damaging agent methyl methanesulfonate (MMS) in comparison to responses to acute heat shock, looking at changes in genome-wide factor occupancy as measured by ChIP-chip. My data suggests that MMS-induced genes undergo transcription complex assembly sequentially, first involving SAGA then TFIID, whereas heat shock genes utilize the SAGA and TFIID pathways in parallel. Also Crt1, the repressor of model MMS-inducible ribonucleotide reductase genes does not play a wide spread role in
repression of DNA damage inducible genes. Taken together my findings reveal distinct involvement of gene and chromatin regulatory factors in response to DNA damage versus heat shock, and suggests different implementations of the SAGA and TFIID assembly pathways that may depend upon whether a sustained or transient change in gene expression ensues.

2.2 Introduction:

One way that eukaryotes respond to environmental signals is to change the expression of their genes (Causton et al., 2001; Gasch et al., 2000). A plethora of proteins are involved in regulating gene expression, and include those that recognize specific DNA sequences, modify or remodel chromatin, assemble the pre-initiation complex, and regulate transcription elongation (Orphanides and Reinberg, 2002; Struhl et al., 1998; Venters and Pugh, 2009a, b). Both the sequence-specific factors and the general transcription machinery might vary in composition depending upon the signaling events that are involved (D'Alessio et al., 2009; Harbison et al., 2004; Holstege et al., 1998; Lee et al., 2002; Ren et al., 2000).

However, beyond the role of different sequence-specific factors, it remains unclear as to what extent the downstream events in the transcription cycle are uniform or specialized when the organism is exposed to different environmental conditions.

The heat shock and DNA damage response pathways have been two well-studied model systems for understanding the steps in the transcription cycle
(Workman et al., 2006; Zanton and Pugh, 2004, 2006). In *Saccharomyces*, both heat shock, DNA damage, and many other environmental stresses/signals induce a common set of environmental stress response (ESR/CER) genes (Gasch et al., 2001; Gasch et al., 2000). However, the response to DNA damaging agents creates an additional genomic response profile that is not found with other stresses. In particular, cell cycle arrest occurs and the DNA damage repair pathway is induced. Unlike the transient heat shock response, the response to DNA damage is sustained for many hours (Elledge et al., 1993; Gasch, 2007; Gasch et al., 2001; Huang et al., 1998; Weinert, 1998). Induction of ESR/CER genes by DNA damaging agents is attenuated in strains in which the DNA damage repair signaling pathway (i.e. *MEC1* and *DUN1*) is disrupted, but the same genes are properly induced in response to heat shock (Gasch et al., 2001). This suggests that distinct signaling pathways converge to activate a common set of genes. Thus, these two systems represent suitable models to ascertain the extent to which the mechanics of a transcriptional response can differ or is similar when cells are exposed to conditions, which invoke common as well as stress-specific responses.

The study of the heat shock response in addition to previous work has corroborated the finding of two mutually compatible pathways for pre-initiation complex (PIC) assembly (Huisinga and Pugh, 2004; Kuras et al., 2000; Lee et al., 2000; Mencia et al., 2002; Struhl, 1986). One involves SAGA-directed PIC assembly through genes that contain a TATA box and binds a TAF-free form of TBP. Such genes represent only about 10-20% of all yeast genes, and tend to be stress-induced
and subjected to widespread inhibition by repressors of chromatin and PIC assembly. In contrast, the TFIID-directed pathway for PIC assembly tends to dominate at TATA-less genes, which comprise 80-90% of the yeast genome. These genes tend to be housekeeping genes that are expressed at constitutively low levels, except in the case of the ribosomal protein (RP) genes. These pathways are compatible in that loss of one can be partially compensated by the other. Additionally, many genes are occupied by varying levels of both SAGA and TFIID (although one pathway tends to dominate). For instance in the TFIID dominated RP genes SAGA co-localizes simultaneously along with TFIID (Ohtsuki et al.). Also both SAGA and TFIID are recruited to heat shock-induced genes (Zanton and Pugh, 2006).

Repression and induction of DNA damage response genes has been studied primarily at the RNR2, 3,4 and HUG1 genes (Huang et al., 1998; Zhou and Elledge, 1992). At these genes, Crt1 binds to its recognition motif and recruits the chromatin repression complex SSN6-TUP1, as well as histone deacetylases to maintain genes in a repressed state (Huang et al., 1998; Li and Reese, 2001; Sharma et al., 2007). The DNA damaging agent methyl methanesulfonate (MMS) has been used to study induction (Jelinsky et al., 2000). Damage sensing and induction occurs through a Mec1p-Rad53p-Dun1p protein kinase pathway (Elledge et al., 1993; Wahl and Carr, 2001; Zhou and Elledge, 2000) which ultimately leads to Crt1 phosphorylation and release from repressed promoters.

In addition to repression, Crt1 is also thought to function as an activator of the RNR genes during MMS induction by directly or indirectly facilitating the
recruitment of TFIID and SWI/SNF (Sharma et al., 2003; Zhang and Reese, 2007; Zhang and Reese, 2005), which in turn promote chromatin remodeling and full assembly of the pre-initiation complex. Such retention of Crt1 would seem to contradict findings that Crt1 dissociates when target genes are activated. Conceivably Crt1 is retained at the promoter at least until the commitment to recruit TFIID and/or SWI/SNF is made. Thus, whether the release of Crt1 from the model RNR3 promoter is immediate or delayed upon MMS treatment is unclear at least at the early stages of induction.

While the repression of RNR genes by Crt1 has long been considered a model for the regulation of DNA damage inducible genes, very few MMS-induced genes have been associated with Crt1-directed repression (Huang et al., 1998). This may be due in part to limited genome-wide studies that specifically address this problem. Using custom-made oligo arrays, I have now generated moderate resolution tiling array occupancy data for Crt1 before and after MMS treatment. I also examined the genome-wide distribution of its co-repressor SSN6-TUP1, its co-activators SWI/SNF and TFIID, and other general transcription factors and sequence-specific regulators that may be part of the MMS response.

By comparing the genome-wide distribution of a wide range of factors in response to DNA damage to that of the heat shock response, I took some initial steps towards understanding whether distinct environmental response pathways interface with the transcription cycle in the same or distinct manner. Surprisingly, I found that while the heat shock pathway involves the simultaneous employment of
the TFIID-dominated and SAGA-dominated PIC assembly pathways, response to MMS initially involves SAGA recruitment and not TFIID. However, SAGA is ultimately replaced by TFIID, suggesting that genes utilize the stress-related SAGA pathway to mount an immediate transient response, and then follow up with the TFIID housekeeping pathway for a long-term response.

In addition to DNA damage, MMS is destructive towards proteins, resulting in their turnover and ultimately the biosynthesis of amino acids to resynthesize the proteins. Gcn4 is a master regulator of amino acid biosynthetic genes (Natarajan et al., 2001). Indeed I found that Gcn4 is recruited to amino acid biosynthetic genes in response to MMS but not heat shock, and this is linked to the recruitment of the SAGA complex in accord with known interactions between Gcn4 and SAGA (Brown et al., 2001; Drysdale et al., 1998; Fishburn et al., 2005; Swanson et al., 2003).
2.3 Results:

2.3.1 Differential recruitment of TFIID to MMS-induced vs. heat shock-induced promoter.

 Genome-wide measurements of factor occupancy levels were measured by ChIP-chip using a custom array platform in which three oligos were arrayed for each gene: one in the vicinity of where UASs tend to reside, one in the vicinity of the core promoter, and one internal to the ORF, near its 3’ end (Venters and Pugh, 2009a). In addition, several hundred locations were arrayed in intergenic regions between two convergently transcribed genes, and were used as background controls. The array contained a total of ~20,000 probes. The two promoter proximal probes allowed detection of factor binding from about -400 to about +100 relative to the start of each open reading frame.

 Occupancy was compared under normal (YPD media, 25°C) conditions as well as after a 30 min treatment with 0.03% MMS (This treatment condition was found to be optimal for triggering the DNA damage stress response without being toxic to the yeast cells). mRNA expression profiles were also generated, which were essentially identical to those published earlier (Gasch et al., 2001). Occupancy and mRNA changes were also compared under acute heat shock conditions (15 min. at 37°C). Genes that responded to the two stresses were largely distinct (Fig. 2.1A), although there was significant overlap in the two responses, in accord with previous
studies (Gasch et al., 2001; Gasch et al., 2000; Gasch and Werner-Washburne, 2002).

I first focused on induced genes, then on repressed genes, and finally those that were neither induced nor repressed but contained significant factor occupancy under one condition or both.

Fig. 2.1. Changes in factor occupancy in response to 30 min MMS treatment in comparison to a heat shock response. A, Venn diagram illustrating the gene membership overlap for those in equivalently labeled clusters in panel B. B, Cluster
plots of changes in factor occupancy and gene expression (as indicated). MMS-treated and heat shock-treated datasets were clustered separately. Data were filtered to retain only those having 100% data present and >1.5 fold changes in occupancy in at least one dataset. The numbers of genes meeting such criteria are indicated. Data were clustered by K means (K = 4 for MMS, and K=3 for heat shock). BY4741 represents a negative control in which the untagged parental strain was processed through the standard TAP ChIP procedure. Shown is data for the “UAS” and “TSS” microarray probes. The Spt3 ChIP dataset used the “UAS” probe, whereas the remaining factors used the “TSS” probe. C, Median log2 values for each dataset in cluster 1 are plotted.

I first compared the utilization of the TFIID vs. SAGA pathway, and included other representatives of the GTFs. As shown by the cluster plot in Fig. 2.1B, genes could be separated in accordance with changes in gene expression, and matched with corresponding changes in occupancy levels. Genes induced by MMS (cluster 1, left panel) acquired GTFs (TFIIB/Sua7, TFIIH/Ssl2, and pol II), as expected. Surprisingly, these genes acquired SAGA (Spt3) but not TFIID (TAF1), despite apparent results to the contrary at the model MMS-response gene RNR3 (Zhang et al., 2008; Zhang and Reese, 2007). This also contrasts with the response obtained from heat shock (Fig. 2.1B, cluster 1 right panel), in which heat shock-induced genes acquired both TFIID and SAGA, as shown previously (Zanton and Pugh, 2004, 2006). Quantification of occupancy changes comparing MMS-induced vs. heat shock-induced is shown in Fig. 2.1C. In comparison to the other GTFs and SAGA, it is clear that TAF1 was not substantially recruited to MMS-induced genes, and this was not due to some
inability to detect TAF1, because recruitment to heat shock-induced genes was apparent.

The general lack of acquisition of TFIID upon MMS induction might have a number of explanations: 1) The previously observed phenomenon might have been specific to a few genes such as RNR3, 2) TFIID might already be present, and so no additional acquisition would occur, or 3) the timing of TFIID acquisition is slow, in which I failed to detect its presence at 30 min. of MMS treatment. Previous studies examined TFIID occupancy after 2 hr of treatment with MMS (Zhang et al., 2008; Zhang and Reese, 2007).
Fig 2.2: Factor occupancy levels at RNR genes and MMS-induced genes in comparison to heat shock induced genes. (A-E) Each panel tracks the indicated GTF, and reports the median occupancy level for the indicated gene or set of genes. Values represent medians for Cluster 1 data, represented as log2 fold over background in the presence or absence of MMS or heat shock, as indicated. The same data used to plot fold changes in occupancy in Fig. 1 are used here. “Low” and “High” represent the bottom and top 10 percentiles of transcription frequency as defined by Holstege et. al (Holstege et al., 1998). Measurements were done under both induced (black bars) and normal growth conditions (gray bars). “RP” denotes ribosomal protein genes. The “TSS” probe was used for all data except for SAGA in which the “UAS” probe was used, which is where SAGA is thought to bind (Larschan and Winston, 2001; Venters and Pugh, 2009a). For the RNR4 gene the “UAS” probe was used instead of the “TSS” probe because it was located closer to the known core promoter region. This places it out of range to where SAGA might be expected to bind. The absence of significant signal from the RNR3 TSS probe (used for Sua7, Taf1, and Rpo21) may be due to a defect in the probe. Values for the RNR genes represent the average of two measurements, whereas values for gene sets represent median values of >200 measurements (two replicates of >100 genes), and thus is more robust.

To address whether TFIID is loaded specifically on the RNR genes in response to MMS, or is constitutively present at most MMS-induced genes and unchanging, I quantified TFIID (TAF1) and other GTF occupancy at RNR2, 3, and 4, as well as at the MMS-induced cluster 1 genes, before and after MMS treatment (Fig. 2.2). These occupancies represent the component values by which changes in occupancy were calculated in Fig. 1. For purposes of assessing dynamic range within the dataset,
median occupancy levels for lowly and highly expressed genes (bottom and top 10 percentile of transcription frequency (Holstege et al., 1998)), as well as the highly active RP genes are shown. In addition, for purposes of comparison to an expected level of occupancy for inducible genes, I examined the collection of heat shock-induced genes in response to heat shock (heat shock cluster 1).

*RNR2, RNR4*, and the median of MMS-induced genes displayed a similar level of GTF occupancy before induction by MMS and a similar magnitude of change after induction as seen at heat shock induced genes in response to heat shock ([Fig. 2.2 A-C](#)), note the background levels associated with the *RNR3* "TSS" probe might reflect a probe defect). SAGA appeared to be constitutively high at heat shock induced genes, and at intermediate levels at MMS-induced genes, in comparison to the dynamic range ([Fig. 2.2 D](#), note the "UAS" probe was used, and it may be improperly positioned at *RNR4* as illustrated in [Fig. 4C](#)). Its level increased modestly in response to MMS or heat shock induction. In contrast, TFIID levels were low prior to induction of both classes of genes ([Fig. 2.2 E](#)), but remained low at the MMS-induced genes after induction, whereas relatively high levels of recruitment were apparent at heat shock induced genes upon induction. Thus, SAGA may be constitutively present at inducible genes, whereas TFIID is immediately recruited to the heat shock genes but not to MMS-induced genes.

I next compared TFIID occupancy levels at 0.5 and 2 hr of induction, the latter being the time point where TFIID occupancy had been previously detected (Zhang et al., 2008; Zhang and Reese, 2007). Interestingly, TFIID occupancy
increased over the time course, whereas SAGA occupancy increased initially then decreased back to the uninduced levels (Fig. 2.3). Thus, my inability to initially detect TFIID occupancy may be due to delayed recruitment. Importantly, these results suggest that MMS-induced promoters may transition from an early SAGA (Spt3)-regulated state, to a more constitutive TFIID (TAF1)-regulated state later on. This contrasts with their apparently simultaneous recruitment at heat shock-induced genes. This might be due to the fact that the heat shock response is more transient and more robust than the response to MMS, and thus heat shock genes may need to deploy SAGA and TFIID more quickly. The MMS response, being more sustained, might therefore deploy the SAGA “emergency” response early, and the sustainable TFIID response over the long haul.

Fig 2.3: Delayed acquisition of TFIID at MMS-induced genes. Genome-wide Spt3 and Taf1 occupancy measurements were made after 0.5 and 2 hr of MMS induction as indicated, for MMS-induced genes (cluster 1), and normalized to occupancy level occurring in the absence of MMS. Changes in mRNA levels are also shown. “*”
indicates that the expression data are from Gasch et. al, 2001.

Given the differences in the physiological and kinetic nature of the two responses, where heat shock is a very rapid response and DNA damage is a delayed response it can be very well argued that the comparison between TFIID and SAGA was done when heat shock response was at its decline (at 15 minutes) whereas DNA damage response was on the rise (at 30 minutes). Thus the differences seen in the recruitment of TFIID might actually be due to the differences in timing at which the localization was being monitored. So to negate this issue, ChIP of Taf1 and Spt3 was done using a more detailed kinetic approach followed by parallel sequencing using SOLiD. For DNA damage response, the test samples were exposed to MMS for three time points of 30 minutes, 1hr and 2 hrs along with no treatment for the same amount of time for the mock samples. For heat shock, two time-points of 5 minutes and 15 minutes were used along with mock heat shock for the same time points. Biological replicates were generated for each condition and every time point. To determine the changes in occupancy in response to MMS and heat shock the total tags were calculated within 500 bp upstream of the ORF start of the yeast genes. The median tag counts was then calculated for ~ 1089 T-T regions (convergent genes) where not much binding was expected (as it is end of the genes). The median tag counts for the T-T regions can be assumed to indicate background binding and the different median values seen for different samples can be accounted for due to the difference in the total tag counts for individual samples. For eg: Sample A with 4 fold
more total tag counts than Sample B, the expected median tag counts of the T-T regions would be 4 fold different as well (within limitations of ChIP efficiency and DNA looping). With this assumption, the median tag counts of the T-T regions for all the samples and their replicates were made to be equal by dividing the total tag counts by a factor (calculated based on median T-T tag counts). Once the data was normalized to the T-T region, Log2 ratios of test (sum of total tags 500 bp upstream of ORF) over mock (sum of total tags 500 bp upstream of ORF) conditions for a given factor were calculated for all the genes. The biological replicates correlated well and were averaged. In response to MMS, SAGA recruitment increases after 30 minutes and recruitment stays robust until 1 hr followed by drastic decrease in occupancy within 2 hrs of MMS induction (Fig. 2.4A). Likewise a completely opposite pattern of TFIID recruitment was observed. There was less recruitment of TFIID during 30 minutes and 1 hr following MMS treatment followed by robust recruitment after 2 hrs of MMS treatment. Heat shock on the other hand showed simultaneous recruitment of TFIID and SAGA within 5 minutes and 15 minutes of exposure (Fig. 2.4B). Additionally 5 minutes of heat shock showed more recruitment of both SAGA and TFIID when compared with 15 minutes whereas transcript levels revealed an opposite pattern (more inducible transcripts at 15 minutes compared to 5 minutes of heat shock). Thus in agreement with the moderate-resolution tiling arrays we see parallel recruitment of both SAGA and TFIID following heat shock and sequential recruitment of SAGA followed by TFIID during DNA damage response.
Fig. 2.4: Simultaneous acquisition of TFII D and S AGA at heat shock induced genes as opposed to MMS-induced genes. Changes in occupancy of Spt3 and Taf1 monitored after 30 minutes, 1 hr and 2hr of MMS induction (panel A) at the DNA-damage inducible genes. Corresponding changes in mRNA levels are shown. Changes in occupancy of Spt3 and Taf1 following heat shock induction for 5 minutes and 15 minutes at the heat shock inducible genes (panel B). Corresponding changes in mRNA are shown as well. “*” indicates expression data are from Gasch et al, 2000 and 2001.
2.3.2 Unexpected mobilization and retention of factors in response to stress that are not coupled to a transcriptional response.

I next examined MMS-repressed genes. Like heat-shock repressed genes, which tend to be TFIID-dominated, we saw a loss of the GTFs (including TFIID) upon MMS treatment, although their initial levels were relatively low as is typical of housekeeping genes (Fig. 2.5A). This is consistent with the notion of a common PIC pathway for housekeeping genes that is shut down during a stress response. I also did not see a loss of SAGA from repressed promoters, whether responding to MMS or to heat shock. Conceivably, the retention of SAGA might contribute to the re-establishment of transcription of these genes, once the stress has been alleviated, although this remains to be tested.
Cluster 2: Repressed genes

Dynamic range
- Upper
- Lower

Cluster 3: SAGA recruited

Dynamic range
- Upper
- Lower
Fig 2.5: Occupancy levels at genes that are repressed or unaffected by MMS or heat shock. Occupancy levels are reported for genes in cluster 2 (panel A, repressed or down-regulated) and cluster 3 (panel B, genes that acquire SAGA but are not immediately activated), as described in Fig. 2.2.

For both MMS-treated and heat shocked cells, I observed a third type of gene cluster (cluster 3) in which SAGA (Spt3) occupancy increased upon treatment (Fig. 2.5B). These SAGA-recruited genes were generally distinct in the two responses (Fig. 2.1A). They were lowly expressed and had low to moderate levels of Pol II and TFIIH. The heat shock-defined set had relatively high levels of TFIID and TFIIB that did not change with MMS treatment. This may be equivalent to partial PIC assembly described previously (Zanton and Pugh, 2006). The same was not observed for cluster 3 genes defined by MMS treatment. Gene Ontology analysis did not reveal any functional grouping of these genes. While I do not know the basis for these seemingly innocuous changes in SAGA occupancy, it does not seem to be coupled to transcription, nor is the phenomenon stress-specific. As reported before (Zanton and Pugh, 2006), the genome undergoes a number of coordinated factor occupancy changes without any apparent or immediate effect on transcription. Nonetheless, the genome is plastic, and often rearranges factors that could help the cells contend with other stresses that might be linked to and occur subsequent to the primary stress.
2.3.3 Crt1 is not linked to MMS-induced genes

Given the historical focus on a select set of model genes (RNR2, 3, 4) to decipher MMS derepression mechanisms, I mapped the genome-wide location of other factors connected to MMS derepression. In particular, I examined Crt1 and its co-repressor Tup1 (Elledge et al., 1993; Huang et al., 1998; Li and Reese, 2001). Crt1 reportedly binds to the sequence T(C/T)GCCATGGAAC (Zaim et al., 2005), and recruits the chromatin co-repressor Tup1 (Zhou and Elledge, 1992). At least at the RNR genes, Crt1 and Tup1 ultimately dissociate upon DNA damage by MMS (Elledge et al., 1993; Huang et al., 1998; Li and Reese, 2001; Zhang and Reese, 2005), and this results in derepression of RNR gene expression. This mechanism deduced at the RNR genes is a model for the regulation for DNA damage inducible genes.

I addressed the validity of our Crt1 and Tup1 ChIP-chip data in four ways. First, I examined all regions significantly bound by Crt1 using MEME, and found that it identified the canonical Crt1 site as a highly enriched motif (E value = 7.4 x 10^-7; Fig. 2.6A). Thus, our Crt1 ChIP dataset has high specificity for its previously defined cognate site. Second, I determined the genome-wide overlap of the top 100 loci of significantly bound Crt1 and the top 250 loci of significantly bound Tup1 (Fig. 2.6B). A very strong overlap of bound regions was observed (P < 10^-65). Thus, Crt1 linkage to Tup1 is widespread and robust. Third, I examined Crt1 and Tup1 occupancy at the RNR genes. As shown in Fig. 2.6C, Crt1 and Tup1 were detected by ChIP-chip at the RNR2 and RNR3, as expected. Little was detected at RNR4, but this might be due
to the fact that the probes for \textit{RNR4} were not at an optimal location (illustrated in Fig. 2.6C). Fourth, upon MMS induction, \textit{Crt1} is reportedly depleted.
Fig. 2.6: Crt1 is not linked to MMS induced genes. (A) Crt1 motif using MEME for the significantly bound genes. (B) Venn diagram illustrating the overlap between Crt1 and Tup1 enriched genes genome-wide. (C) Bar graph representing Crt1 and Tup1 occupancy represented as log2 fold over background in the presence or absence of MMS at RNR genes. UAS and TSS probe distribution along with Crt1 X-box sites at the RNR2, RNR3 and RNR4 genes have been shown. (D) Bar graph illustrating Crt1 and Tup1 occupancy represented as log2 fold over background in the presence or absence of MMS at MMS inducible genes. Shown is data for the “UAS” microarray probes. (E) Venn diagram representing the overlap between MMS inducible genes with the Crt1 and Tup1 enriched genes.

at the RNR genes (Huang et al., 1998; Zhang et al., 2008; Zhang and Reese, 2007). At RNR2 and RNR3, I indeed saw a reduction of Crt1 upon MMS treatment, although the amount of depletion was rather modest (1.5-2 fold). Together these findings validate the Crt1 and Tup1 data sets (beyond the standard metrics of reproducibility).

Given the connection of Crt1 occupancy at repressed RNR genes and its loss upon induction, I examined the top 100 Crt1-bound sites for occupancy loss upon MMS treatment. Surprisingly, the median occupancy level of Crt1 at MMS-induced genes was rather low, and this did not change with MMS induction (Fig. 2.6D). Similarly, Tup1 levels were low at MMS-induced genes and did not change upon MMS-induction. When I examined the overlap between Crt1-bound genes and MMS-induced genes only four genes overlapped (P = 0.8, Fig. 2.6E). A similar low level of overlap was seen with Tup1. The few overlapping genes included the RNR genes.
Thus, Crt1 and Tup1, despite being at $RNR$ genes and behaving as previously reported, do not appear to have a widespread linkage to the MMS-induced DNA damage response.

2.3.4 Chromatin remodelers are present constitutively at DNA damage inducible genes.

ATP-dependent chromatin remodelers play important roles in repression and activation of DNA damage inducible genes (Cairns, 2004; Morrison et al., 2007; Sharma et al., 2003). I therefore examined the genomic distribution of several remodeling complexes, including RSC (Rsc2), SWI/SNF (Snf2), INO80 (Ino80), and ISW1, at bulk MMS-inducible genes as well as the $RNR$ genes in response to MMS (Fig. 2.7). For purposes of setting the dynamic range, we also plotted the apparent occupancy level at the top 5th and bottom 50th percentile of all genes (more than half of all genes are expected to have background levels of occupancy). In general, moderately high levels of RSC, SWI/SNF, and INO80 remodeling complexes were present at these genes prior to and subsequent to induction. The ISW1 complex appeared to be absent from promoter regions of MMS-induced genes. With the exception of SWI/SNF, these remodelers appeared absent from $RNR4$, although we cannot exclude the likely possibility that the $RNR4$ probes were not in a proper position to detect occupancy. I also detected high occupancy of ISW1 at $RNR3$. Since the bulk MMS-induced genes have substantial gene membership, the collective bulk...
assessment of occupancy at these genes should not be affected by a small number of out-of-position probes.

Fig. 2.7: Constitutive presence of many chromatin remodelers at DNA damage inducible genes. Each panel tracks the indicated chromatin remodeler, and reports the median occupancy levels at the “UAS” probe for the MMS and heat shock
inducible genes along with the RNR genes. Median occupancy values are represented as log2 fold over background in the presence or absence of MMS or heat shock, as indicated. The occupancy level at the top 5th and 50th percentile of all genes was also plotted to set the dynamic range.

In contrast to MMS-induction, heat-shock induced genes had comparatively little SWI/SNF and RSC, but contained INO80 and ISW1, and their occupancies changed little with heat shock. This comparatively low SWI/SNF enrichment, nevertheless, was similar in magnitude to previous observations especially at the promoter region since only the occupancy at the UAS probes were used for the remodelers (Shivaswamy and Iyer, 2008). Thus, classes of genes that respond to distinct environmental stimuli appear to involve the constitutive presence of different but also overlapping sets of chromatin remodeling complexes.

2.3.5 MMS-induced accumulation of Gcn4 at amino acid biosynthetic genes

In addition to DNA damage, MMS damages proteins (Boffa and Bolognesi, 1985; Jelinsky and Samson, 1999; Lee et al., 2007; Natarajan et al., 2001). Sufficient damage triggers protein turnover and a requirement for additional protein biosynthesis that places demands on the free amino acid pool. MMS treatment is known to induce amino acid biosynthetic genes, which are under the control of Gcn4 (Jelinsky and Samson, 1999). I therefore examined whether MMS treatment lead
Fig. 2.8. Increased occupancy of Gcn4 and SAGA at MMS-induced amino acid biosynthetic genes. (A) Frequency distribution plot illustrating changes in occupancy of Gcn4 at the ‘UAS probes’ in response to heat shock and MMS. (B) Bar graph representing changes in expression of the 31 amino acid biosynthetic genes in response to MMS or heat shock. Changes in expression of MMS or heat shock
inducible genes (Cluster 1 genes from Fig1) have been used to set the dynamic range. (C) Each panel tracks the indicated GTF, and reports the median changes in occupancy in response to MMS or heat shock at the amino acid biosynthetic genes as well as the MMS or heat shock inducible (Cluster 1) genes.

to increased occupancy of Gcn4 at the 31 amino acid biosynthetic genes that contain a Gcn4 binding site. I examined Gcn4 occupancy under normal conditions, in the presence of MMS, and also after heat shock (as a control). In the presence of MMS, Gcn4 was recruited to most of the 31 genes (Fig. 2.8A). This level of robust recruitment was not observed during heat shock, demonstrating its specificity towards MMS. As reported previously (Jelinsky and Samson, 1999; Natarajan et al., 2001), I also found the expected increase in expression of the amino acid biosynthetic genes in response to MMS treatment but not heat shock (Fig 2.8B). For the 31 genes, I observed preferential recruitment of SAGA (Spt3) and not TFIID (Taf1), which is consistent with my observations of other MMS-induced genes (Fig. 2.8C). This Gcn4 recruitment at amino acid biosynthetic genes in response to MMS also agrees with earlier work which has shown how UV irradiation activates Gcn4 through a Ras2 dependent and Gcn2-independent pathway (which is also an the amino acid starvation independent response) at HIS3 and HIS4 genes (Engelberg et al., 1994). Thus, Gcn4 is recruited to amino acid biosynthetic genes in response to MMS, likely in response to protein damage, and this leads to assembly of the transcription machinery that likely involves the SAGA complex.
2.4 Discussion:

2.4.1: Serial vs. parallel implementation of the SAGA and TFIID PIC assembly pathways

The goal of this study was to evaluate the similarities and differences in PIC assembly when sets of genes respond to distinct environmental stimuli. The basic question is whether different genes responding to different stimuli set up their PIC in basically the same way. I found this not to be the case, although commonalities exist. I chose to compare the DNA damage and heat shock response because prior expression profiling studies suggested that they elicit similar (ESR) as well as distinct stress-specific responses (Gasch et al., 2001; Gasch et al., 2000). Also there are differences between the two responses, where one (heat shock) is transient and the other (DNA damage) more prolonged (Gasch et al., 2001). Previously, we reported evidence of two distinct pathways for PIC assembly (Huisinga and Pugh, 2004), one involving the SAGA complex whose function in TBP delivery tends to predominate at highly regulated stress-induced genes, and the second involving TFIID whose function in TBP delivery tends to operate at housekeeping genes which tend to be down-regulated in stress. These so called stress-induced and housekeeping genes were originally classified as the ‘inducible’ and ‘constitutive genes’ (Harbury and Struhl, 1989; Kuras et al., 2000; Struhl, 1986). A lot of work has shown how the inducible genes are more tightly regulated through activators
whereas the constitutive genes are less dependent on activators. Additionally, studies have shown that for constitutive or house-keeping genes which are predominantly the TFIID dependent genes mediator is dispensable for TBP recruitment unlike the SAGA dependent inducible genes (Li et al., 2000). Thus there are variations in PIC assembly at SAGA and TFIID dependent genes and can be attributed primarily to the different activators responsible for their activation. But these activators in turn could be orchestrated by different stress conditions thereby further contributing to the differences seen in the complex regulation of genes.

When cells are exposed to an abrupt shift in temperature from 25°C to 37°C, not only do they respond by inducing genes important for stress tolerance, but they also increase their metabolic rates, which necessitates increased expression of metabolic genes (Gasch, 2007). Thus many heat shock induced genes may be partially active at 25°C and thus have moderate levels of GTFs occupying their promoters. Similarly, many DNA-damage inducible genes may be lowly expressed and have moderate levels of GTF occupancy in the absence of DNA damage.

My study suggests that SAGA promotes rapid PIC assembly in response to stress, whereas the timing of the response through TFIID is more stimulus-specific, with heat stress creating a rapid response and DNA damage a slower but more sustained response. Heat-shock genes tend to be more functionally dependent on SAGA than on TFIID, perhaps because the response is so transient that the contribution from TFIID is rather limited. MMS-induced genes tend to be more TFIID-dominated, despite the presence of SAGA, perhaps because the SAGA
contribution to PIC assembly is rather short and the TFIID contribution rather prolonged during sustained activation of these genes. Once the TFIID pathway takes charge, the presence of SAGA may not be required (at least for PIC assembly).

Conceivably, genes that are regulated to have their assembly pathways in constant flux may be more SAGA-dominated due to the rapidity by which SAGA can promote PIC assembly (van Werven et al., 2009) compared to slower TFIID recruitment, and possibly slower TFIID-directed PIC assembly (although slower TFIID recruitment may be gene-specific or specific for types of environmental responses rather than an inherent limitation of TFIID in that recruitment is rapid at heat shock induced genes). Thus, loss of TFIID would have less of an impact than loss of SAGA. Genes that are regulated by a more sustained recruitment of the PIC may nevertheless recruit both SAGA and TFIID, but the constitutive presence of TFIID might be manifested by a greater dependence on it.

Both heat shock and DNA damage induce a common set of about 25 genes, based upon stringent filtering criteria used in my study. In regard to the timing of SAGA/TFIID recruitment, they followed the MMS-induced serial recruitment pattern when responding to MMS, but followed the heat shock induced parallel pattern when responding to heat shock (data not shown). Thus, the type of environmental stress and by implication the respective stress-specific signal transduction cascade may dictate the predominant PIC assembly pathway (and timing) for a given gene.

So far studies have already shown how cells when exposed simultaneously to two different stress conditions like hypo-osmotic shock along with mild heat shock
lead to a combined genomic response which is usually a sum of both the responses (Gasch and Werner-Washburne, 2002). This shows how efficiently cells can respond independently to two different conditions simultaneously and is another example of how multiple diverse regulatory pathways can orchestrate a response to a distinct environmental condition. It would therefore be interesting to study the PIC assembly under varied stress conditions, independently as well as in combination to dissect all the complex mechanisms employed by cells to ensure survival.

2.4.2: Apart from RNR genes, Crt1 does not play a widespread role at DNA-damage inducible genes

The finding that Crt1 is not present at MMS-inducible genes, except for RNR genes was surprising since the notion that Crt1 negatively regulates the RNR genes has been a model for DNA damage inducible gene regulation. The question that remains to be addressed is what other factors (if any) are playing a role in repression of those genes under normal growth conditions. Additionally it would be interesting to study the genes which Crt1 binds to and also address what function or role it is playing at those genes and its relation (if any) to growth under different stress conditions.
2.4.3: **Constitutive presence of chromatin remodeling complexes**

I found that both heat-shock and DNA damage inducible genes involve multiple chromatin remodeling complexes that appear to be constitutively present before and after induction. The two classes of genes utilize the same and different chromatin remodeling complexes. While the basis for this distinction is unclear, differential chromatin remodeler use might be related to lower levels of uninduced expression in one class vs. the other, or it may be related to transient vs. sustained induction perhaps in some way linked to serial vs. parallel use of the SAGA and TFIID PIC assembly pathways.
2.5 Materials and methods:

2.5.1 Yeast strains, growth conditions and public datasets

_Saccharomyces cerevisiae_ S288C strains were obtained from the Yeast TAP-Fusion Library (Open Biosystems). The C-terminally TAP tagged proteins were immunoprecipitated using IgG antibodies. Untagged strain BY4741 was used for the null ChIP and as a control for occupancy normalization.

Each strain was grown to OD<sub>600</sub> 0.8 at 25°C in 500 ml of YPD. Methyl methanesulfonate (MMS) was then added to a final concentration of 0.03% for 30 minutes, 1hr or 2 hrs for DNA damage growth conditions. For heat shock conditions cultures were heat shocked at 37°C for 15 minutes (or 5 minutes) or mock-treated at 25°C for 15 (or 5 minutes) minutes by adding volumes of appropriately heated media to achieve the desired final temperature.

Where indicated, published gene expression data for DNA damage were from Gasch et al. (2001) (expression datasets for only 30 min. and 2 hr exposure to 0.02% MMS were used). Gene expression data for 15 minutes heat shock were from Zanton and Pugh (2004). Expression data sets for 5 and 15 minutes heat shock were from Gasch et al. (2000). ChIP-chip data for heat shock condition for the factors Rpo21, Taf1, Sua7, Ssl2, Spt3, Ino80, Isw1, were from Venters and Pugh (2009).
2.5.2: ChIP-chip and microarray expression assays

ChIP assays were performed as described previously (Zanton and Pugh, 2004). Following treatment (respective stress and mock-stress) cells were cross-linked with 1% formaldehyde for 15 min. at 25°C and then quenched for 5 min. with glycine. After harvesting the cells were lysed with zirconium beads, the washed chromatin pellet was sheared by sonication using a Bioruptor (Diagenode) generating on average 200-300bp sized DNA fragments. Sheared chromatin was immunoprecipitated with IgG-Sepharose and the ChIP-enriched DNA was then amplified using ligation-mediated PCR (Harbison et al.) 75-300bp of LM-PCR amplified DNA fragments were gel purified using the Qiagen protocol. DNA yield following gel purification was determined using the Nanodrop ND-1000 spectrophotometer.

DNA labeling and hybridization to the custom oligo-tiling arrays were performed as described previously (Zanton and Pugh, 2004). 100ng of gel-purified LM-PCR ChIP-enriched DNA was amplified by 15 PCR cycles. The +MMS and 37°C samples were then Cy5-labeled, and the −MMS and 25°C samples were Cy3-labeled and cohybridized to microarrays. A dye swap was performed with independent biological replicates for every factor.

For expression analysis, BY4741 cultures were treated as described above for the respective stress conditions (DNA damage or heat shock) and the cells were
harvested at the end of the treatment. Sample preparation and microarray assays were performed as described in previous publications (Chitikila et al., 2002; Huisinga and Pugh, 2004). Changes in gene expression were similar to published results (Gasch et al., 2001; Gasch et al., 2000).

2.5.3: Array design and analysis

Array design for the low-density tiling microarrays is the same as described previously (Venters and Pugh, 2009a). Occupancy data were filtered, normalized to corresponding probes in the null (no tag) control, then centered by dividing the dataset by the median normalized probe value for those probes located in intergenic regions between two convergent genes, and log₂ transformed as described previously (Zanton and Pugh, 2006). For changes in occupancy, ratios were calculated on filtered signal, then log₂ transformed, and centered by subtracting the median ratio for probes located in intergenic regions between two convergent genes. Raw data is accessible through ArrayExpress, and processed data is available in Table S1. CLUSTER and TREEVIEW were used for analysis and representation of microarray data (Eisen et al., 1998). P values were calculated by using the CHITEST function in EXCEL.
2.5.4 Significance of p-value:

High throughput experiments generate tons of data. Considerable variability in the data is additionally obscured by biological noise. The data is subjected to various filtering parameters and quality control steps. Finally, the use of both replicates and statistics helps to find biological meaning in our data. We usually compute p-values to find the significant changes in expression/occupancy of a transcription factor. The higher the p-value greater the probability that the differences is by chance. The smaller the p-value it is more likely they are significantly different and it is not by random chance. P-value of .05 means there is a 5% or a 1 in 20 times probability that our observations is by random chance. So it could be wrong and there is a risk involved. If we decrease the p-value the risk of an incorrect observation decreases, thus for a p-value of 0.01 (1%) 1 in 100 times the observation is by random chance; for 0.1% or 0.001, 1 in 1000 times it is by random chance. The lower the p-value the higher confidence one has in their data. Thus with a p-value of 0.0001, there is a 1/10000 chance of data being unreliable. On the other hand a p-value of 0.276, 276/1000 = 27.6% or 1 in 3.6 times its wrong. P-value tells us how confident one is of the data. P-value greater than 5% or 0.05 suggests differences between two datasets is non-significant and includes false-positives.

Additionally with more number of hypotheses being tested on the same datasets the chances of getting a significant p-value increases thereby generating false
positives which can be corrected using the Bonferroni correction. Thus if the acceptable p value cut-off is 0.05 (making the overall error rate 5%) and if 10 pairwise comparisons are to be made, the new cut-off applying the bonferroni correction method would be 0.05/10 = 0.005. Although this method likely decreases the number of false positives but with increasing number of hypotheses being tested with a given dataset it can increase the number of false negatives. This problem can be circumvented to some extent by doing replicate studies.

Within the same dataset, the bonferroni correction can be applied to individual genes as well. A p-value of 0.05 suggests there is a 5% probability that expression of the gene in one condition vs a second condition is different by chance alone. As the number of genes increases say to 10,000, 500 genes could be significantly different by chance alone. On applying the bonferroni correction, the p-value of each genes is multiplied by the total number of genes. So with 1000 genes the new corrected p value cut-off becomes 0.05 * 1000 = 0.00005.

2.5.4: ChIP-Seq

ChIP assays were performed as described previously (Zanton and Pugh, 2004). Library preparation was done according to Applied Biosystems SOLiD protocol using ChIP enriched DNA where individual samples were bar-coded using multiplexed 4-bp adaptors. They were then pooled together and sequenced using SOLiD sequencer. 31 bp sequence reads were then aligned back to the reference
yeast genome using SHiMP software allowing upto 3 mismatches (Rumble et al., 2009). Raw tags are available at NCBI Trace Archives (Accession number: SRA024321).
2.6: References:


Drysdale, C.M., Jackson, B.M., McVeigh, R., Klebanow, E.R., Bai, Y., Kokubo, T.,
activation domain interacts specifically in vitro with RNA polymerase II holoenzyme,

Ehrenhofer-Murray, A.E. (2004). Chromatin dynamics at DNA replication,
display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95, 14863-
14868.


response involving the Ras signaling pathway and AP-1 transcription factors is

transcription activator during the transcription cycle. Mol Cell 18, 369-378.
ascomycete fungi. Yeast 24, 961-976.


Workman, C.T., Mak, H.C., McCuine, S., Tagne, J.B., Agarwal, M., Ozier, O., Begley, T.J.,


Chapter 3

ChIP-exo mapping of Ribosomal Protein (RP) gene factors suggests differential binding at paralogous RP genes and potential role of Crt1 in down-regulation of RP genes.

3.1 Summary:

In *S. cerevisiae*, there are ~137 ribosomal protein (RP) genes that encode for 78 ribosomal proteins. 59 out of the 78 ribosomal proteins are each encoded by two RP genes arising from WGD (whole genome duplication). Given the coordinate regulation of RP genes in response to growth conditions the question I address here is how redundant is the transcriptional regulation of the RP paralogs. In particular, I examined the binding of seven known RP gene regulators (Ifh1, Fhl1, Hmo1, Sfp1, Rap1, Gcn4, Abf1) and Crt1 (novel regulator of RP genes) using ChIP-exo. The data suggests that RP paralogs bind the factors differentially. Interestingly most of these factors track together as a huge complex at RP promoters (localized around -200 to -400 bp upstream of TSS) except for Crt1 which is more localized in the coding region of RP genes. Also most of these factors have multiple binding sites (>4) in the RP gene promoters that could play a role in cooperative binding or release from the RP genes, for coordinate regulation in response to growth cues. Furthermore, motif analysis of these factors shows location linkage with the Rap1 ribosomal protein gene motif in addition to usage of selective motifs. Additionally
although Abf1 is known to regulate a few (11-12 out of 137) RP genes, which are not bound by Rap1, this study identified Abf1 to bind to 97 RP genes.

3.2 Introduction:

Ribosomal protein (RP) genes are known to be the most highly and coordinately expressed genes in the cell (Warner, 1999). As they are the major constituents of the ribosome, bulk of the cell’s energy is invested in regulating its expression since the demand to maintain the ribosomal protein levels at an optimal and equimolar amount in the cell is extremely high. Almost 50% of transcripts produced by RNA Pol II is accounted for by the RP genes (Warner, 1999; Warner et al., 2001). Additionally transcription of RP genes is very much entwined with the growth and environmental conditions of the cell and in response to any stress these are the first group of genes that are coordinately down regulated (Causton et al., 2001; Gasch, 2007; Gasch et al., 2000).

In *Saccharomyces cerevisiae*, there are 137 genes that encode for around 78 ribosomal proteins, of which 19 genes encode for a unique RP and the remaining 118 genes (59 pairs) each encode for a duplicated RP, and are known as the paralogous genes (Planta and Mager, 1998). This duplication of 75% of ribosomal proteins was the result of a whole genome duplication (WGD) event that took place one hundred million years ago (Wolfe and Shields, 1997). Most of the duplicated genes were lost and some mutated to other functional forms in the process of
evolution barring a few genes such as the 59 pairs of RP genes, protein kinases and genes involved in carbohydrate metabolism (Seoighe and Wolfe, 1999a, b). Out of the 78 RPs, 14 of them are not needed for viability of the cell at least under standard growth conditions; 4 out of the 14 are encoded by single genes and the remaining 10 are encoded by duplicate genes (Steffen et al., 2012). Surprisingly, 21 out of the 59 RP pairs encode for identical proteins (100% identity at amino acid level) and majority of the rest are highly similar (Steffen et al., 2012). Although few pairs have diverged significantly such that the codon adaptation index \(\text{CIA}(\text{CIA is used to quantitatively predict the expression level based on codon usage})\) differs considerably between the paralogs (Akashi and Eyre-Walker, 1998; Planta and Mager, 1998). An intriguing question, which has been addressed before, is what could be the reason for retention of the RP paralogs following WGD? It was initially thought that two copies were retained to meet the cells needs for ribosomal proteins i.e to always maintain a high dosage (Warner, 1999). But of the 19 unique RP genes 15 are essential for viability whereas out of 59 paralogous pairs, 5 are essential \(\text{(RPL15A, RPL18A, RPL42A, RPS28A and RPS30B)}\) (Steffen et al., 2012). That these 5 paralogous genes are essential suggests that the other pair does not support the protein dosage needed for the cell in the absence of the essential pair. It has been shown that RPL15B is not transcribed under normal lab growth conditions and hence does not contribute to RPL15 protein levels in the cell (Simoff et al., 2009). Also majority of the paralogs are non essential suggesting a functional redundancy of the paralogs at least for one function (meeting the required protein...
dosage level in the cell). But recent work has also alluded to the fact that paralogous pairs are not actually functionally redundant (Komili et al., 2007), particularly at the protein level. So the question that arises is what about regulation of the ribosomal paralogs at the transcriptional level? A lot of key regulators are already known to play a role in RP gene regulation. In particular Rap1 and Abf1 have been known to play a role in its regulation. Rap1 binding sites have been identified in virtually all the RP promoters, present mostly in pairs and at a preferred position relative to TSS. T-rich regions proximal to Rap1 sites have been shown to play a role in regulating RP gene transcription (Lascaris et al., 2000; Lascaris et al., 1999). In addition to Rap1 and Abf1, Ifh1, Fhl1, Sfp1, Gcn4 and Hmo1 have also been implicated as regulators of RP gene transcription (Hall et al., 2006; Joo et al., 2011; Jorgensen et al., 2004; Marion et al., 2004; Martin et al., 2004; Rudra et al., 2005; Schawalder et al., 2004; Wade et al., 2004; Zhao et al., 2006). Now most of the genome-wide studies related to dissection of the role of the above factors in regulation of RP genes relied heavily on in silico motif searches, ChIP-chip or ChIP-seq studies, which are all affected by stringency thresholds. Also role of certain factors like Sfp1 at RP genes has been debatable and some studies show that it binds at RP genes (Marion et al., 2004) whereas some studies refute this claim (Jorgensen et al., 2004; Kasahara et al., 2007; Lee et al., 2002) Likewise, in our previous work (Ghosh and Pugh, 2011) we had identified about 100 Crt1 bound genes of which 25 were RP genes (unpublished data), which too was limited by threshold levels. So, with the availability of ChIP-exo (Rhee and Pugh, 2011), a high resolution genome-wide factor mapping method we
tried to re-define the transcriptional regulatory network at RP genes. Some of the questions that arose were the following: As ChIP-chip studies of Crt1 identified 25 RP genes bound by Crt1. Could we identify more RP genes bound by Crt1 using ChIP-exo? Given that Crt1 is a sequence specific regulator (Elledge and Davis, 1989), what motif is used at RP genes and does it differentiate between paralogous pairs of RP genes? Additionally do the known bonafide RP factors have a bias for one RP paralogous pair over the other? How does the spatial organization of these factors work at RP genes? Most of the factors are known to be location linked to RP genes via Rap1 motif. Does that mean they do not bind directly to the DNA? Is there any other motif playing a role in the factor binding sites? Could we identify novel motifs?

Thus in this study, ChIP-exo was henceforth used to dissect the roles of Abf1, Rap1, Fhl1, Ifh1, Sfp1, Hmo1, Gcn4 (known regulators of RP genes) and Crt1 (potential regulator of RP genes) in the transcription of RP genes. In particular, the binding of the factors at paralogous RP genes (RP genes present in duplicates encoding the same ribosomal protein subunit) was addressed. The results show that Rap1, Hmo1, Sfp1, Ifh1, Abf1 co-occupy the same RP genes and cluster together thereby conglomerating at RP gene promoters. Also all the bonafide RP regulators were predominantly located in the promoters of RP genes (100-500bp upstream of TSS) whereas Crt1 was found to be located at the coding region. Crt1 was also found to bind to 88 RP genes and is involved in the down regulation of RP genes, which was not known previously. Additionally, majority of the paralogous RP genes bind the
factors differentially both with respect to occupancy levels (one paralog has lesser occupancy compared to the other) and also based on either presence of binding or no binding at RP genes. Also of all the factors, Sfp1 binding at RP genes has always been debatable and this study shows that it does bind significantly at the RP genes. Interestingly although majority of the RP factors show a tight location linkage to the RP genes via Rap1 ribosomal protein motif (Rhee and Pugh, 2011) but some like Hmo1, Gcn4, Crt1 and Ifh1 do show the presence of other signature motifs at RP genes.

3.3 Results:

3.3.1: Crt1 binds to 1535 Pol II genes genome-wide including 88 ribosomal protein genes suggesting a potential role in global gene regulation.

ChIP-exo analysis of TAP-tagged Crt1 identified 3223 peak pairs without pre-selection based on the presence of a motif. ChIP-exo applies a specialized λ-exonuclease (exo) digestion step along with chromatin immunoprecipitation (ChIP) thereby defining the binding location of the cross-linked sites (Rhee and Pugh, 2011). 3223 Crt1 bound locations (peak-pairs) in turn correspond to ~1535 Pol II encoded genes (assigned to genes whose TSS is within -750 to +1000 bp of the Crt1 bound location). Previous ChIP-chip studies based on p-value cut-offs have reported around 100 genes (Ghosh et. al, 2011) and 105 genes (Harbison et al, 2004) bound by Crt1. Zaim et. al 2005 used in silico methods to identify only 30 genes based on
the presence of Crt1 binding motif T(C/T)GCCATGGCAAC. This method has its own limitations as the initial PSSM for the motif search was generated based on the sequences of only 5 Crt1 bound genes (RNR2, 3, 4, HUG1, CRT1). Interestingly out of the 100 Crt1 bound genes identified previously using ChIP-chip (Ghosh and Pugh, 2011), 25 belonged to ribosomal protein genes (RP genes), which was significant (unpublished data).

Investigation of the distribution of the Crt1 bound sites (3,223 sites) around the TSS of Pol II encoded genes shows binding both at the promoters and coding regions of Pol II genes (Figure 3A). Also previously I had found that Crt1 does not bind to DNA damage (MMS) inducible genes although it is the predominant regulator of the bonafide DNA damage inducible genes, the RNR genes (Elledge and Davis, 1989; Ghosh and Pugh, 2011; Zhou and Elledge, 1992). With 15 times the number of genes found compared to ChIP-chip studies (Ghosh et. al, 2011) I wanted to re-confirm if Crt1 could be linked to DNA damage inducible genes. Out of the 1535 genes only 47 were found to overlap with 146 DNA damage inducible genes (p-value 0.42). Although it includes almost one-third of the DNA damage inducible genes but such an overlap could occur by random chance and hence was dismissed. Thus the lack of a robust widespread linkage between DNA damage inducible genes and Crt1 still holds true (Figure 3B). Now that I had identified new sets of genes bound by Crt1 I tried to investigate the functional classification of the genes involved. KEGG and MIPS analysis of the Crt1 bound genes showed metabolic pathways, ribosome and
Figure 3.1: Crt1 binds to 1535 Pol II encoded genes including 88 ribosomal protein genes suggesting a potential role in global gene regulation. A, Distribution of 3,233 Crt1 bound locations (peak-pair mid-points) around the TSS of Pol II genes. B, Venn diagram illustrating the overlap between 1,535 Crt1 bound genes and 146 MMS induced genes genome-wide. C, Functional classes of genes bound by Crt1 using MIPS and KEGG. D, Raw sequencing tag distribution around 3,223 Crt1 bound locations. Red and blue indicate the 5’ends of forward and reverse strand tags centered around the Crt1 motif mid-point. Rows were sorted based on Crt1 occupancy level. E, Crt1 motif (14bp) degeneracy at different sites and functional classes of genes.
synthesis of secondary metabolites as the predominant groups (Figure 3.1C). The identification of genes involved in central metabolism further strengthens the claims made by earlier studies of a regulatory link between metabolic pathways and binding at RNR genes involved in DNA damage (Zaim et al., 2005). Additionally, out of the 121 RP genes, Crt1 was found to bind at 88 RP genes.

Crt1 is also a sequence-specific regulator that binds to a 14 bp consensus sequence (Huang et al., 1998; Zaim et al., 2005). Immediate question was whether Crt1 ChIP-exo data showed enrichment of the same motif? Could I identify novel motif variants for Crt1? MEME analysis (Bailey et al., 2009) of top 500 bound sites resulted in identification of a Crt1 consensus motif as reported earlier (Ghosh and Pugh, 2011; Harbison et al., 2004). It resembles the Rfx1 (mammalian homolog of Crt1) binding site, which is an imperfect palindromic sequence (GTNRCC/N-N0-3-RGYAAC). It consists of two 6bp half-sites, separated by a region variable in length (0 to 3 bp) but is preferentially either 1 or 2 bp long. For 1 bp spacers there is a strong preference for either T or A, whereas for 2bp spacers the consensus is AT. For Crt1, I found 2bp spacer (AT) was the default consensus. It is also known that one half-site matching the consensus RGYAAC is enough for Rfx1 to bind as it can bind as a monomer (Emery et al., 1996a; Emery et al., 1996b; Siegrist et al., 1993) even though the other half-site maybe considerably more degenerate. FIMO (p-value threshold of $10^{-2}$) using the Crt1 consensus sequence generated from MEME, reported the presence of a Crt1 motif in all the 3,223 bound sites. Figure
3.1D shows tags distribution around the motif mid-point. Further characterization of the Crt1 motif at the bound sites (top 100, top 101-200 bound sites) and at different classes of genes (RP genes or tRNAs) showed that the second half-site is usually more conserved compared to the first half (Figure 3.1E).

3.3.2: Crt1 is linked to down-regulation of RP genes.

Low-resolution ChIP-chip data from previous work (Ghosh and Pugh, 2011) (unpublished data) had shown a significant overlap between Crt1-bound genes (P =10^{-68}) and Rap1-bound RP genes (Bulyk et al.) (Figure 3.2A). To further corroborate the previous observation, I performed high-resolution mapping of Crt1 and Rap1 to precisely characterize genome-wide binding events as ChIP-chip is plagued by both noise from low resolution and background binding. ChIP-exo identified 88 Crt1 bound RP genes and 97 Rap1 bound RP genes and like previous observation a robust overlap was found between the two factors at the RP genes, thereby suggesting that Crt1 could be a novel regulator of RP genes (Figure 3.2B). This observation can be further strengthened by previous work showing Rfx1 (mammalian homolog of Crt1) regulating mouse rpl30 gene (Safrany and Perry, 1993). Given the function of Crt1 as a repressor (Elledge and Davis, 1989; Huang et al., 1998), yet paradoxically present at many highly transcribed RP genes, I then investigated the extent to which Crt1 was regulating RP genes. I compared published gene expression changes (Gasch et al., 2001) in a crt1Δ strain at all genes
Figure 3.2: Crt1 attenuates expression of Ribosomal Protein genes. A, Venn diagram illustrating the overlap between Crt1 enriched genes and Tup1 enriched genes (Ghosh et.al, 2011) with Rap1 bound ribosomal protein genes (Bulyk et.al). B, Venn diagram illustrating overlap between Crt1 and Rap1 bound RP genes (obtained by ChIP-exo). C, Frequency distribution plot representing changes in expression of ribosomal protein genes in a *crt1Δ* strain (Gasch et.al, 2001). Top Crt1 bound genes include the top 100 genes.

vs. all RP genes. As shown in Figure 3.2C, expression of RP genes increased in the *crt1Δ* strain, demonstrating that Crt1 is directly inhibiting RP gene expression, despite these genes being highly expressed. Their intrinsically high expression likely limits the fold increase that is possible in the *crt1Δ* strain (compared to the repressed RNR genes). I also examined MMS-inducible genes and found that while
most were unaffected by the loss of Crt1, a subset were down-regulated rather than the expected up-regulation if Crt1 was repressing (Figure 3.2C). These genes were not particularly enriched for either Crt1 occupancy or its recognition site (not shown), suggesting that the positive regulation of these genes by Crt1 might be indirect.

### 3.3.3: RP gene regulators Abf1, Fhl1, Ifh1, Sfp1, Hmo1, Rap1, Gcn4 bind at the promoters of RP genes as opposed to Crt1, which binds primarily at the coding region.

To study the role of Abf1, Fhl1, Ifh1, Sfp1, Hmo1, Rap1, Gcn4 and Crt1 at RP genes, I carried out ChIP-exo using TAP-tagged strains of the above mentioned factors (Rhee and Pugh, 2011). The binding of the factors at an individual RP gene loci (RPS21B) has been shown in Figure 3.3A. Out of the 8 factors, robust binding could be seen for Ifh1, Hmo1, Crt1 and Sfp1 at the same loci at the RPS21B gene. Sfp1 binding at RP genes has always been debatable where one study indicates that it does bind to promoters of some RP genes (Marion et al., 2004), other studies show that binding of Sfp1 cannot be detected at significant levels at RP genes (Jorgensen et al., 2004; Kasahara et al., 2007; Lee et al., 2002). Genome wide ChIP-exo of the above-mentioned factors identified thousands of binding locations in the genome (Figure 3.3B). To identify the number of RP genes bound by each factor the peak-pair midpoints (factor binding site co-ordinate) was aligned within +/- 1000 bp to the TSS of 121 RP genes (only 121/137 RP gene TSS information is available). Majority
of the RP genes are short (mature mRNA after splicing) but the presence of one or two introns (400-500bp in length on average) in most of the RP genes was the rationale behind assigning factor binding to RP genes within 1000bp. Overall, with ChIP-exo I identified a lot more RP genes bound by the factors (except for Fhl1) than previously reported (Hall et al., 2006; Jorgensen et al., 2004; Marion et al., 2004; Martin et al., 2004; Rudra et al., 2005; Schawalder et al., 2004; Wade et al., 2004; Zhao et al., 2006). The factor bound locations (C-W peak pair midpoints) were then plotted around the TSS of 121 RP genes. All the factors except for Crt1 showed enhanced enrichment at the promoters of RP genes within 500 bp upstream of the TSS (Figure 3.3C). One fraction of Crt1 does bind at the promoters but majority of the binding sites were enriched in the coding region of RP genes. In addition to Crt1, Hmo1 also showed a much broader binding distribution around the TSS of RP genes. (Huang et al., 1998)
### B

<table>
<thead>
<tr>
<th>Factor</th>
<th>C-W paired peaks (genome-wide)</th>
<th>No. of RP genes bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ifh1-TAP</td>
<td>8,758</td>
<td>103 RP genes</td>
</tr>
<tr>
<td>Fhl1-TAP</td>
<td>2,648</td>
<td>64 RP genes</td>
</tr>
<tr>
<td>Rap1-TAP</td>
<td>3,426</td>
<td>97 RP genes</td>
</tr>
<tr>
<td>Sfp1-TAP</td>
<td>5,686</td>
<td>100 RP genes</td>
</tr>
<tr>
<td>Hmo1-TAP</td>
<td>14,227</td>
<td>113 RP genes</td>
</tr>
<tr>
<td>Abf1-TAP</td>
<td>6,426</td>
<td>97 RP genes</td>
</tr>
<tr>
<td>Gcn4-TAP</td>
<td>1,724</td>
<td>51 RP genes</td>
</tr>
<tr>
<td>Crt1-TAP</td>
<td>3,223</td>
<td>88 RP genes</td>
</tr>
</tbody>
</table>

### C

- Frequency vs. distance to TSS of RP genes for: Rap1, Ifh1, Sfp1, Abf1.
- Frequency vs. distance to TSS of RP genes for: Fhl1, Gcn4.
- Frequency vs. distance to TSS of RP genes for: H3 (Nucleosome) for Hmo1 and Crt1.
**Figure 3.3:** Crt1 binds to coding regions of RP genes as opposed to the bonafide RP factors. **A**, Genetrack screen shots of RP factor binding at RPS21B gene loci. **B**, Panel showing total number of factor binding sites identified genome-wide using ChIP-exo and the number of RP genes bound by the RP factors. **C**, Frequency distribution of RP factor peak-pair midpoints (binding location mid-points) around TSS of 121 RP genes.

### 3.3.4: Differential binding of RP gene regulators observed at paralogous RP genes with no robust correlation with RP gene promoter activity or TFIIB occupancy

Out of the 78 ribosomal proteins, 58 are encoded by duplicate RP genes. 49 out of the 58 paralogous pairs were used to monitor the RP factor binding between the two pairs. **Figure 3.4A** shows the percent rank occupancy of the RP factors at the two paralogous pairs of RP genes. To measure the occupancy of one factor at a RP gene, the total tags were summed up at every binding location for each factor at every RP gene. Thus tags were averaged across multiple sites for a given RP gene. The tags were then normalized to the median tag counts at all binding locations (genome-wide) for a given factor. Normalized tag counts were then log2 transformed and percent rank binding was computed for each RP paralogous pair. The paralogous groups (I and II) were separated based on Hmo1 occupancy. The occupancy (percent rank) was clustered using k-means =2 and visualized using Treeview (Eisen et al., 1998) **Figure 3.4A** clearly shows differential binding of the
RP factors at cluster K1. For K2 although the factors bind robustly at both the groups, overall occupancy at group I is still higher for all the factors except for Rap1, Sfp1 and Abf1. Also in both the clusters Hmo1, Ifh1, Sfp1, Rap1 and Abf1 seem to bind with same robustness. In contrast, Fhl1 and Gcn4 do not completely follow the core factors (Hmo1, Ifh1, Sfp1, Rap1 and Abf1). With Crt1 binding at the coding region of the RP genes it does follow other factor binding patterns at RP genes either. Now that majority of RP paralogs bind factors differentially I was interested to see if the differential binding of the factors at the paralogous pairs correlated with the transcriptional readout of the RP genes. I used TFIIB occupancy at promoters as a measure of transcriptional efficiency. TFIIB occupancy (% rank) was plotted at the paralogous RP genes (Rhee and Pugh, 2012). In addition to TFIIB occupancy the RP promoter activity from a recent study (Zeevi et al., 2011) was also used to see any correlation with differential binding. A significant correlation was not observed between TFIIB binding and differential binding of factors at RP genes. Although the work in Zeevi et al. 2011 did find a correlation of promoter activity with motif distribution (of Rap1, Fhl1 and Sfp1) at RP promoters and nucleosome disfavoring sequences, my occupancy results do not correlate with the promoter activity classification (high, medium and low). This suggests that motif based studies can sometimes lead to ambiguity.

In addition to reporting occupancies at the individual paralogous pairs I looked into the median occupancy levels at each cluster K1 and K2 for the two paralogous groups I and II (Figure 3.4B). The magnitude of differential binding at K1 group of
**Figure 3.4:** Differential binding of RP factors at RP paralogous pairs. **A,** Occupancy (percent rank) of RP factors at the RP paralogous pairs clustered using K-means=2. Each row represents one ribosomal protein, with one paralog present in group I and the other paralog present in group II. TFIIB occupancy (percent rank) data obtained from Rhee and Pugh, 2012. RP gene promoter activity classification into high, medium and low used from Zeevi et.al (2011). **B,** Each panel shows the indicated factor median occupancy (percent rank) at the specific paralogous group (I or II) and cluster (K1 or K2).

Paralogous pairs is much greater compared to K2. Additionally Fhl1, Gcn4 and Crt1 show differential binding at K2 group versus the K1 group.

**3.3.5: RP factors co-cluster at promoters of RP genes within a spatial distance of <20 bp of each other.**

In **Figure 3.4A** the occupancies reported at the RP genes was the average occupancy at multiple bound locations at every RP gene. On average all RP genes have >5 binding location for every factor at the promoter region clustered within 500 bp upstream of the TSS with some spreading into the coding region of RP genes as well. To get a more detailed view of the factor binding using the resolution provided by Chip exo all factor binding sites were mapped at the individual RP genes. **Figure 3.5A** and **Figure 3.5B** depicts a cartoon representation (drawn to scale) of the binding events at two paralogous pairs of RP genes (RPL14A, RPL14B and RPS0A, RPS0B). It can be seen that a large complex of all the factors conglomerates at the
promoters of RP genes. Such a clustered view of factor binding at RP genes has not been previously reported. To measure the distances of each factor from one another at the RP genes, distances between peak midpoints was computed in a pair-wise combination. Rap1 being the bonafide regulator of the RP genes was used as a lynchpin for formation of the huge conglomerate. The peak midpoint distribution of all the factors with respect to Rap1 was calculated and plotted. As seen in Figure 5C majority of the factors are within 20 bp of each other (except Crt1 which binds predominantly at the coding region of RP genes) suggesting co-occupancy of these factors at the RP genes. Given the coordinate regulation of RP genes multiple bound locations of these regulatory factors might be essential for the quick, drastic expression changes of RP genes under adverse conditions or for a more subtle and tighter fine tuning of its expression under normal growth conditions.
Figure 3.5: (A-B) Cartoon representation (drawn to scale) of factor binding at two paralogous pairs of RP genes (RPL14A, RPL14B and RPS0A, RPS0B). Each filed oval represents an RP factor. Degree of opacity of the oval represents occupancy levels. High being opaque and low being translucent. The width of the oval represents factor binding site size (C-W distance). The numbers denote binding locations with respect to TSS of the RP gene.
Figure 3.5C Distance between peak pair mid-points of all the factors with respect to Rap1 peak-pair midpoint at RP genes. The number in parentheses represents the mode of the distance distribution.
3.3.6: Selective motif usage by RP factors at RP genes.

To identify the motif usage by the RP factors I performed MEME (Bailey et al., 2009) on top 300 genome-wide bound sites for the factors Abf1, Fhl1, Ifh1, Sfp1, Hmo1, Rap1, Gcn4 and Crt1. Three to four sequence motifs were found to be strongly overrepresented in the genomic locations bound by the above factors.

PSSM’s of the overrepresented motifs identified are shown in Figure 6A-H (top panel of the 4 color plot motif plot). The PSSM’s were used to do a directed FIMO search on all binding locations for each factor at RP genes with a p-value threshold of 10^{-2}. Although the Rap1 ribosomal protein motif was one of the motifs identified for all the factors in addition to an A-rich motif known to be present at RP genes, some of the factors such as Abf1, Gcn4, Hmo1 and Crt1 showed enrichment of their own signature motifs at RP genes. The motif bound by Hmo1 has been shown in previous work to be linked to Fhl1, Ifh1 and Hmo1 binding at RP genes and is known as the IFHL motif (Hall et al., 2006; Wade et al., 2004). Surprisingly I found widespread binding of Abf1 at RP genes using its own consensus motif, given the fact that Abf1 was previously known to bind to few RP genes which were known to lack Rap1 motifs. To investigate the binding patterns around the motifs, 5’ends of raw tags on the forward and reverse strand were mapped -/+300 bp around the motif mid-point. Interestingly all the factors bind around the motif using a much broader region (spans 200-300 bp) which could be due to the huge complex getting assembled at the RP promoters.
**Figure 3.6(A-H)** Raw tag distribution around motif midpoints of all the RP factors at the RP genes. Red and blue denote 5’ ends of tags on the forward and reverse strand. Rows were sorted based on occupancy levels. Occupancy rows were linked to motif on the right panel where a four-color plot was used to present motif conservation at all binding locations.
3.4 Discussion:

3.4.1. Crt1 negatively regulates ribosomal protein genes, not DNA-damage inducible genes

Ribosome biosynthesis accounts for >50% of transcription in yeast and mammalian cells. The rate of the ribosome biogenesis in yeast is controlled by the rate of transcription of RNA and protein subunits of the ribosome (Warner, 1999). The rRNA and RP gene transcription in turn is very sensitive to the growth potential of the cell and is rapidly repressed in a variety of stress conditions (Gasch et al., 2000; Warner, 1999). The entire RP regulon comprises of 137 RP genes and is tightly regulated. The majority of the RP genes have binding sites for Rap1 at the promoter region and some have sites for Abf1. Rap1 acts both as an activator and silencer and it is been postulated that it depends on the combinations of associated factors. Also it is known that Rap1 binds to the upstream sequences of RP genes in order to activate, whereas for repression it does not have to bind directly to the RP genes (Li et al., 1999). Many RP specific transcription factors that cooperate with Rap1 to regulate RP genes have been identified such as Abf1, Fhl1, Ifh1, Sfp1, Sch9, Reb1, Hmo1 (Hall et al., 2006; Marion et al., 2004; Planta et al., 1995; Rudra and Warner, 2004; Rudra et al., 2005; Wade et al., 2004; Warner, 1999). Most of these factors play a role in activation of these RP genes. So far only Crf1 has been known to represses RP genes in response to stress conditions and it seems plausible that
there could be more factors playing a role in RP gene repression (Martin et al., 2004; Zhao et al., 2006). It was surprising to find Crt1 playing an inhibitory role at the majority of the RP genes. Just like Crt1, Rfx1 (mammalian homolog of Crt1) has been found to play a repressive role at c-myc and PCNA (Katan et al., 1997; Liu et al., 1999; Zajac-Kaye et al., 2000), and has been reported to bind to the promoter region of the mouse RPL30 gene (Safrany and Perry, 1993). This is in agreement with my observation of Crt1 binding at RP genes.

The finding that Crt1 negatively regulates RP genes rather than MMS-inducible genes was surprising since my data supports the notion that Crt1 negatively regulates the RNR genes which have been a model for DNA damage inducible gene regulation (although upon DNA damage Crt1 becomes a dual repressor, activator protein (Zhang and Reese, 2005)). How Crt1 negatively regulates RP genes remains to be explored, and the RNR genes may not provide a suitable model for such regulation.

3.4.2: Non-redundancy of ribosomal protein genes

Ribosomal protein genes need to be in equimolar amounts in the cell for efficient ribosome biogenesis and are under tight regulation both in terms of RP gene transcription as well as degradation of RP gene mRNA (RP mRNA has one of the shortest half-life in the cell, $t_{1/2} = 0.5-3.0$ min) (Planta and Mager, 1998; Warner, 1999). Ribosomal protein gene paralogs are indeed different, although they encode
for virtually identical proteins. Studies have shown that deletion of RPL7A affects growth to some extent and properties such as budding whereas RPL7B has no such effects (Mizuta et al., 1992). When I compared the factor occupancy at RPL7A I indeed found high occupancy of all the RP factors compared to RPL7B (cluster K1 Fig 3.4A). Similarly, yeast replicative life span (RLS) studies invariably have shown that on deletion only one paralog leads to an increased RLS but not the other (Chiocchetti et al., 2007; Mizuta et al., 1992; Steffen et al., 2008). In this work too, I found non-redundant binding (occupancy) of the RP factors (Hmo1, Sfp1, Ifh1, Rap1, Abf1, Fhl1, Gcn4, Crt1) at the RP genes suggesting that the two paralogs do behave differentially. Additionally, I found clustering together of these RP factors in a massive overlapping complex with most factors binding within 20 bp of each other. Why the need to cluster all the factors together is not known, but one can always speculate that clustering of binding sites might be needed for a graded response to fine tune the optimum levels of transcript in response to varied growth cues. Similarly as found in previous studies not all the factors are present at all the RP genes. Why do some genes have one combination of factors whereas the others have a different combination? Is there a ‘binding code’? (eg: RPL26B has moderately high occupancy of Hmo1, Ifh1, Sfp1, Rap1 and Abf1, but low occupancy of Fhl1, Gcn4; whereas RPL26A has low occupancy of Hmo1, Ifh1, Sfp1, Rap1 and Abf1 and a moderately high occupancy of Fhl1, Gcn4) One might argue that the distribution of factors might have some evolutionary link where RP genes need to back up some of the regulatory factors in anticipation of sudden changes in environment. Indeed
studies have shown that in the evolution of RP gene paralogs, one paralog is more ancestral the other is more derived (Kellis et al., 2004). One can also speculate that the more conserved RP paralog would be used under normal growth conditions whereas the more derived paralog would serve its purpose under alternative growth conditions. Irrespective of how important the role of the other factors, the role of Rap1 as a ‘superfactor’ of RP genes still holds true (although some RP genes do not bind Rap1). As new factors regulating RP genes are being discovered over time one cannot help but believe in the concept of multiple “parallel pathways” in the regulation of RP genes. Nevertheless it all converges to Rap1, because RP genes lacking Rap1 do not respond to environmental stimuli that regulate RP gene transcription.
3.5 Materials and methods:

3.5.1 Yeast strains and growth conditions:

*Saccharomyces cerevisiae* S288C strains were obtained from the Yeast TAP-Fusion Library (Open Biosystems). Each TAP-tagged strain was grown to OD$_{600}$ 0.8 at 25°C in 500 ml of YPD (yeast peptone dextrose). The cells were cross-linked with formaldehyde at a final concentration of 1% for 15 min followed by quenching with 0.125 M glycine.

3.5.2 ChIP-exo (ChIP followed by λ exonuclease digestion)

3.5.3 Chromatin immunoprecipitation (ChIP)

ChIP was performed as described in Zanton and Pugh 2004. Briefly, each TAP-tagged strain was grown to OD$_{600}$ 0.8 at 25°C in 500 ml of YPD (yeast peptone dextrose). The cells were cross-linked with formaldehyde at a final concentration of 1% for 15 min followed by quenching with 0.125 M glycine. Following harvesting the cells were disrupted using zirconium beads using a vortexer for 2 hrs at 4°C using FA-lysis buffer (50 mM HEPES (pH 7.5), 150 mM NaCl, 2mM EDTA, 1% Triton-X-100, 0.1% NaDeoxycholate, 0.2% SDS, Complete Protease Inhibitor (Roche)) and 0.1% SDS. The lysate was collected, spun at 14 krpm for 5 min and 4°C, and the supernatant was discarded. The chromatin pellet was then washed thrice with FA-
lysis buffer + SDS + CPI, resuspended in 1.5 ml of FA lysis buffer + SDS + CPI and transferred to a 15 ml polystyrene conical tube.

1.5 ml of chromatin from 500 ml cell equivalent was sonicated using a Bioruptor (Diagenode, power high, time: 30 sec on, 30 sec off for 30 minutes). The sheared chromatin was spun at 14 krpm for 15 min at 4°C. The supernatant was then transferred to a new 1.5 ml tube and spun at 14 krpm for 30 mins at 4°C. 1.5 ml of the solubilized chromatin was diluted using FA lysis buffer to get a final SDS concentration of 0.05%. It was then incubated using IgG-bound Dynabeads at 4°C for 1.5 hrs on a rototorque. Following immuno-precipitation, the Dynabeads were washed with the following buffers: 1ml each of FA-lysis buffer, High salt wash buffer (FA lysis buffer + 1M NaCl), Wash buffer 2 (FA lysis buffer + 0.5 M NaCl) and Wash buffer 3 (FA lysis buffer + 0.25 M LiCl, 1% NP-40, 1% sodium deoxycholate, 2mM EDTA and 10 mM tris-HCL (pH 8.0))

3.4.4 ChIP-exo

ChIP-exo was performed as described in Rhee and Pugh 2011, with slight modifications needed for the Illumina platform. Briefly, the immunoprecipitated DNA still bound to the IgG Dynabeads resin was subjected to the following enzymatic steps sequentially with intermittent steps of stringent washing: Firstly the chip’d DNA was polished using 1.5 U of T4 DNA polymerase (New England biolabs),150 uM dNTPs, 50ug/ml of BSA,1X NEBuffer 2 in a reaction volume of 30ul
at 12°C for 30 min. The samples were then kinased using 10U of T4 PNK (New England biolabs), 1X T4 ligase buffer in a reaction volume of 30ul, incubated at 37°C for 30 min. They were then A-tailed (or Illumina platform only) using 3’ to 5’ Klenow (5U/ul), 100uM dATP, 1X NEBuffer 2 in a reaction volume of 30 ul at 37°C for 30 min. This was followed by the first adaptor ligation (ExA2/Index, 5’AGACGTGTGCTCTTCCGATCT3’-5’GATCGGAAGAGCACGTCTGAACTCCAGTCAC3’, 5’CAAGCAGAAGACGGCATACGAGATINDEXGTGACTGGAGTTC3’), using 15 pmol of ExA2/Index adaptor, 800 U T4 DNA ligase (New England biolabs), 100ug/ml BSA 1X T4 ligase buffer in a reaction volume of 40 ul at 25°C for 1 hr. Fill-in reaction was then performed using 10 U of φ29 polymerase (New England biolabs), 165 uM dNTPs, 200ug/ml BSA, 1X φ29 reaction buffer in a 30 ul reaction volume at 30°C for 20 min. The filled-in DNA was then subjected to 5 U of λ exonuclease (New England biolabs) digestion using 1X λ exonuclease buffer in a reaction volume of 30 ul at 37°C for 30 min. This was followed by RecJf exonuclease treatment using 22.5 U of RecJf (New England biolabs) in 1X NEBuffer 2 in a reaction volume of 30 ul at 37°C for 30 min. The buffers used for washing the dynabeads following each enzymatic reaction were as follows: 10 mM Tris-EDTA buffer (1 mM EDTA, pH 8.0), High salt wash buffer, Wash buffer 2, Wash buffer 3, 10 mM Tris-HCl (pH 8.0, 7.5, dependent on the pH of the ensuing enzymatic step). After the last washing, the DNA bound to the dynabeads was eluted using 450 ul of ChIP elution buffer (25 mM Tris-Cl pH 7.5, 200mM NaCl, 2mM EDTA, 0.5% SDS) at 65°C for 15 min. The eluted DNA was then subjected to reverse-crosslinking and protein degradation using 1.5 ul of Proteinase.
K (Roche, 20mg/ml) at 65°C with an overnight incubation (6-12 hr). DNA samples were then subjected to Phenol:Choloroform:Isoamyl alcohol (PCIA) extraction and precipitated with ice-cold 100% ethanol. The DNA samples were then subjected to primer extension using 0.5uM of P7 PCR primer (3’GAGCATACGCGAGACGAAC 5’) in 20 ul reaction volume containing 1X φ29 reaction buffer, 200ug/ml BSA, 100 uM dNTPs, at 95°C for 5 min, followed by 52°C for 5 min and 30°C for 10 min. 10 U of φ29 polymerase was then added and the samples were incubated at 30°C for 20 min. Following heat inactivation of the enzyme at 65°C for 10 min, the samples were A-tailed (for Illumina platform only) using 3’ to 5’ Klenow (5U/ul), 100uM dATP, 1X NEBuffer 2 in a reaction volume of 30 ul at 37°C for 30 min. This is followed by the second adaptor (EXA1 adaptor, 5’GGACACTCTTTCCCTACACGACGCTCTTCCGATCT3’, 3’TGTGAGAAAGGGATGTGCGAGAAGGCTAG 5’) ligation 15 pmol of ExA1 adaptor, 800 U T4 DNA ligase (New England biolabs), 100ug/ml BSA, 1X T4 ligase buffer in a reaction volume of 40 ul at 25°C for 1 hr. The samples were then purified using Agencourt AMPure magnetic beads (Beckman Coulter Genomics) following the manufacturers instructions. This was followed by PCR amplification (18-21 cycles) of the samples using P7 and P1.2 primers (5’AATGA TACGGCGACCCGGAGA TCTA CACTCT TTCCCTA CAGACGC3’). Samples were then run on a 2% Agarose gel and 200-350 bp PCR products were size selected and then gel purified using QIAquick gel extraction columns (Qiagen) following the manufacturers instructions. The samples were then quantified using 2100 Bioanalyzer (Agilent) and sequenced using Illumina sequencer.
3.4.5 Peak calling, replicate matching, assignment of binding locations to genes and motif analysis:

The sequenced tags were aligned to *Saccharomyces cerevisiae* genome (build 19 Jan: 2007), using Bowtie alignment algorithm allowing upto 1 mismatches in the INDEX and 3 mismatches in the sequenced reads. The tag distribution was used to identify peaks in the forward (W) and reverse strand (C) using the peak calling algorithm of Genetrack ([Albert et al., 2008]). The mode of the peak-pair (C-W distance= D) distribution was then estimated followed by matching of replicates using python scripts (courtesy Pindi Albert). The C-W (D) modes, which in turn is the binding site size for the factors were as follows: Crt1 (28bp), Rap1 (26bp), Abf1 (34bp), Fhl1 (39bp), Gcn4 (43bp), Hmo1 (76bp), Ifh1 (19bp), Sfp1 (54bp). Two replicates for each factor were matched and were considered to be replicates when the binding sites were within distance D of each other. Also peak-pairs with total tags <=2 were removed from the analysis.

The final peak-pairs (bound locations) were estimated for all the factors. For the RP factors, the number of genes bound by each factor was determined by plotting the peak-pair midpoint around +/- 1000 bp of the TSS of the RP genes.

MEME and FIMO ([Bailey et al., 2009]) were used for motif analysis at the bound locations.
3.6 References


Chapter 4
Discussion

4.1 Summary of this study

4.1.1 Stress-specific transcriptional responses:

Unicellular organisms can sense different environmental signals. Genome-wide expression studies with model organisms have shown that the responses to those conditions are not only conserved among higher eukaryotes but a commonality in the response exists as well (ESR) (Causton et al., 2001; Gasch, 2002, 2007; Gasch et al., 2001; Gasch et al., 2000; Gasch and Werner-Washburne, 2002). The key question that arose from the expression studies was that if the genome was re-programmed by two different stress conditions would that invoke similar or varied PIC assembly pathways.

In the work presented in Chapter 2 I took some initial steps in trying to characterize the differences or similarities by which transcriptional regulators work in order to execute the ESR. The major emphasis was on the two co-activator-pathways for PIC assembly, namely TFIID and SAGA. Additionally other players involved in transcriptional regulation like the chromatin remodelers (INO80, SWI/SNF, Isw1, RSC) were tested as well to monitor if a bias for a particular stress response exists or not. The key finding from this study is that when cells are exposed to two different stress conditions (namely, heat shock and DNA damage) which differ both
in the kinetics as well as signaling pathways involved to trigger the response, we see
differences in the modes of co-activator recruitments at the genes induced by the
two conditions. With DNA damage we found that cells deploy the co-activator
pathways serially, with SAGA being recruited before TFIID. In case of heat shock
both the co-activator pathways were recruited parallely. Although both the stresses
are recruiting SAGA and TFIID at their induced genes suggesting that the regulation
of expression is fundamentally similar but the timing of their implementation varies,
which could be attributed to the characteristic differences between the two stress
conditions.

4.1.2 Non-redundant regulation of ribosomal protein genes:

RP genes are highly regulated cohort of genes that encode for 78 ribosomal proteins.
59 out of the 78 ribosomal proteins in S. cerevisiae are encoded by two RP genes.
Although a lot of work has been done to study the non-redundant roles of the
ribosomal proteins encoded by the paralogs (Baudin-Baillieu et al., 1997; Enyenihi
and Saunders, 2003; Komili et al., 2007), not much emphasis has been placed on the
transcriptional regulation of the paralogs. The recent development of high-
resolution genome-wide mapping methods such as ChIP-exo (Rhee and Pugh,
2011) led me to utilize its power leading to the work presented in chapter 3 of this
thesis where some initial attempts have been made in investigating the binding of
RP factors at the paralogous genes. 7 known RP factors (Rap1, Hmo1, Ifh1, Fhl1,
Gcn4, Abf1, Sfp1) (Fingerman et al., 2003; Hall et al., 2006; Idrissi et al., 1998; Joo et
al., 2011; Jorgensen et al., 2004; Lascaris et al., 2000; Lascaris et al., 1999; Lieb et al., 2001; Marion et al., 2004; Martin et al., 2004; Wade et al., 2004) were investigated along with a new potential regulator Crt1. Some of the key observations were as follows: Crt1 was found to bind predominantly in the coding region of ~ 88 RP genes unlike the other RP factors. What could be the rationale for binding in the coding region? One can speculate a link with introns of RP genes, since we know 100 out of the 137 RP genes contain introns (Spingola et al., 1999; Zhang et al., 2008). Previous reports have suggested novel factors like Arr1 (transcription factor involved in arsenite detoxification) to have binding sites at introns of RP genes (Zhang et al., 2008). Additionally Crt1 was not only found to bind at the RP genes but also seems to be involved in down-regulating its expression. Thus Crt1 has both a location and functional relationship at the RP genes. Also given how RP genes are coordinately regulated the expectation from this study was to find all the regulators at all the RP genes. Previous ChIP-chip studies on individual factors always reported binding at a fraction of the RP genes, and one can assume it was due to limitations of data thresholding with ChIP-chip that compromises the actual number of binding sites identified. But with ChIP-exo too a similar trend was observed where not all the factors were present at all the RP genes. Additionally the RP paralogs were found to bind the RP factors differentially. Also most of the players at RP genes cluster within 20 bp of each other. Additionally the factors were found to be location linked at RP genes via both the Rap1 motif and their own signature motifs. Does that mean the factors bind directly to Rap1 as well as bind
individually at the RP genes. If they indeed were being recruited via Rap1, one would expect similar ranges of C-W distances (estimate of binding site size) for all the factors, but that was found not to be the case.

4.2 Future directions:

The purpose of Chapter 2 was to investigate if different stress conditions re-programmed the genome in similar or different ways. The work presented in Chapter 3 provides insights about regulation of the paralogous RP genes and also the finding that Crt1 might play a role in regulation of RP genes. With the study involving the mobilization of the transcriptional machinery in response to two stress conditions, it is clear that differences exist in the co-activator recruitment. But given how only a handful of representative transcription factors were tested in contrast to ~400 factors involved in transcription there is a lot left to be discovered. Secondly, the study showed that SAGA is eventually replaced by TFIID in the MMS inducible genes, like a switch in the co-activator recruitment. Does that mean both SAGA and TFIID work independent of each other at those genes or is SAGA necessary for the recruitment of TFIID. This is a testable hypothesis where recruitment of TFIID can be investigated in Spt3, Spt20 and Gcn5 deleted mutant strains of SAGA. A possible outcome from such studies is that none of the mutants affect TFIID recruitment suggesting there is no inter-dependency. But if Spt3Δ does affect TFIID recruitment, it would suggest that delivery of TBP by SAGA is a pre-requisite for TFIID binding. If Spt20Δ affects TFIID recruitment, would suggest that
integrity of the whole SAGA complex is needed for TFIID to bind and if the Gcn5Δ affects TFIID recruitment, would suggest that the acetylation status of those genes affects TFIID recruitment (Bhaumik, 2011). Additionally, it would also be interesting to investigate the bias in the co-activator recruitment pathway in other stress conditions like oxidative stress, osmotic stress, and nutrient starvation (to name a few). Another characteristic effect of the stress responses is when cells are exposed to two stress conditions like mild heat shock followed by hypo-osmotic shock, the response is a summation of both the conditions (Gasch et al., 2000). Thus even when assaulted by two conditions cells can distinguish and mount a combined response. It would be interesting to investigate the re-programming of the genome when cells are exposed to dual stresses, as that would be closer to a more physiological state. Since in the environment yeast has to deal with a bunch of stresses at the same time instead of one stress at a time as tested in laboratories.

With respect to the study of RP genes regulation, investigating the RP factor binding at RP genes in different growth conditions would be a necessary step leading to a more complete understanding of the transcription factor dynamics at the RP genes. Although the role of one paralog over the other in one growth condition vs another has been much of a speculation, it would be interesting if mechanisms involving such regulation could be established. It would help in the understanding of why the cell has divided responsibilities among so many transcription factors for regulating
RP genes (though not at an absolute level since all factors do not bind the same RP genes).
4.3 References:


VITA

Sujana Ghosh

EDUCATION:

PhD in Biochemistry, Microbiology and Molecular Biology, May 2013
The Pennsylvania State University, University Park, PA
Dissertation title: “Genome-wide transcriptional responses in yeast
(Saccharomyces cerevisiae)”
Dissertation Advisor: Dr. B. Franklin Pugh

MSc in Microbiology, July 2005
University of Calcutta, Calcutta, India
Master’s thesis project: “Isolation of a CCCP resistant mutant of Vibrio
cholerae and characterization of its virulence properties”
Thesis project advisor: Dr Rukhsana Chowdhury (E1 Scientist, IICB, Calcutta,
India)

BSc with Honors in Microbiology (minor in Chemistry and Physics), July
2003
University of Calcutta, Calcutta, India

PUBLICATIONS:

• Ghosh, S., Pugh, B.F. (2011) Sequential recruitment of SAGA and TFIID in a
  genomic response to DNA damage in Saccharomyces cerevisiae. Mol Cell Biol
  39(21): 9155-66

  conserved domain in Swi2/Snf2 is required for SWI/SNF remodeling. Nucleic
  Acids Research 31(1):190-202

HONORS:

• Braddock Graduate Fellowship, BMMB Dept., PSU, 2006-2007, 2007-2008