A Polyglutamine Domain Enables Transcriptional Reprogramming in Response to pH Change

By

Juan Ignacio Gutiérrez

A thesis submitted in satisfaction of the requirements for the degree of Doctor of Philosophy in Molecular and Cell Biology in the Graduate Division of the University of California, Berkeley

Committee in charge:
Professor Douglas Koshland
Professor Rebecca Heald
Professor Donald Rio
Professor Arash Komeili

Spring 2017
Abstract

A Polyglutamine Domain Enables Transcriptional Reprogramming in Response to pH Change

By

Juan Ignacio Gutiérrez
Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley
Professor Douglas Koshland, Chair

pH is tightly controlled in the cell as it influences almost all processes. Intracellular pH in Saccharomyces cerevisiae changes according to nutritional state, acidifying as the cell runs out of nutrients. We found that cytosolic acidification plays a role in transcriptional reprogramming during carbon starvation. A polyglutamine domain in the SWI/SNF complex senses this pH change, enabling the induction of about 180 glucose-repressed genes. We propose that polyglutamines undergo phase transitions in a pH dependent manner, which allows glutamine transactivation domain proteins to engage and disengage in transcription. This could be a general mechanism for transcriptional reprogramming events within cells.
To the memory of my father Pedro Gutiérrez.
Acknowledgements

I would like to acknowledge everyone who helped me through this process.

First, my mentor Liam Holt, for the guidance and support provided during these years, as well as the members of the Holt lab at UC Berkeley and NYU: Conor Howard, Shyam Bhakt, Kris Kennedy, Greg Poterewicz, Morgan Delarue, Dajun Sang and especially to Greg Brittingham and Sud Pinglay for the feedback provided. Also thank you to my co-mentor Doug Koshland without whose support this would have not been possible, as well as my undergraduate mentor Claudia Saavedra, whom first gave me the opportunity to work in a research laboratory. Thanks also to the people from UC Berkeley, especially Vinny Guacci, Hugo Tapia, John Gerhart and Chris Killian for the countless protocols and advice provided. To the people of the Institute for System Genetics at NYU, particularly to: Neta Agmon, Emily Adney, David Truong, Marta Chiodin and Julie Oaks, for making my time in New York City such a pleasant experience. Also thanks to our collaborators Sanghyuk Lee, Antony Lee, Carlos Bustamante, Simon Alberti, Titus Franzmann, Andrii Kopach, David Fenyő and Xuya Yang.

These years would not have been the same without the friendship of my classmates at UC Berkeley, especially: Andy Navarrete, Andrew Birnberg, Kevin Tsui, Emilia Esposito and Arielle Woznica. And my Berkeley roommates and friends: Carolina Muñoz, Adriana Ramirez, Seth Saltiel, César Díaz, Matias Negrete, Rodrigo Araya, Tito Guerra, Lucho Fuentealba, José Parraguez, Carloclo Lizama, Fred Kinana, Simón Vidal, Reiley Reed, Ian O'Donohue, Nuria, Ramón and Aida De la Fuente. Also I would like to acknowledge my good friend Ivan Tarride for helping me in the art of figure design.

Finally, thanks to my family for their unconditional love: my mother Isabel and my siblings Pedro, Javiera and Daniela.
# Table of Contents

## Chapter 1: Introduction.

Figures:

1. Intracellular pH regulation.  
2. Acidification of the cytosol during starvation.

## Chapter 2: Material and methods.

Tables:

1. Table 1, strains used in this study

## Chapter 3: A polyQ domain enables transcriptional reprogramming in response to a pH change.

Figures:

3. Deletion of the Snf5 polyglutamine leads to loss of viability in carbon starvation, especially at neutral pH.
4. ADH2 expression requires an acidic environmental pH and the SNF5 Polyglutamine.
5. Histidines are important for the function of the SNF5 polyglutamine domain.
7. Transient acidification of the cytosol is sufficient for ADH2 expression.
8. Transient cytosolic acidification also occurs upon Re-addition of glucose to starved cells.
9. RNA-seq analysis reveals widespread polyglutamine dependence in the expression of glucose-repressed genes.

Tables:

2. Table 2, top 20 highly expressed genes upon glucose-starvation not upregulated in ΔQ-snf5 or HtoA-snf5.
Chapter 4: Characterization of SNF5 mutations and Exploration of a mechanism for polyQ gene expression activation.

Figures:
10. Deletion of SNF5 Polyglutamine creates a dominant negative allele. 46
11. C-terminally tagged SNF5 is a hypomorphic allele. 47
12. SNF5 as well as the mutant alleles are not degraded upon glucose-starvation. 48
13. Phenotypes of SNF5 mutants in various stress conditions. 49
14. snf5Δ strains have an extended lag phase after switching to alkaline pH media. 50
15. Fluorescent microscopy of SNF5 suggests an aggregation mechanism. 51
16. A synthetic mini-Spidroin protein aggregates upon glucose-starvation at pH 5.5 but not at pH 7.5. 53
17. A synthetic spidroin domain partially rescues pH-dependent induction of transcription. 54
18. The GEPM peptide that aggregates upon temperature does not recover ADH2 expression. 56
19. SS18, a glutamine rich, member of the Human SWI/SNF Complex can partially complement the SNF5 polyglutamine domain. 57

Chapter 5: Future perspectives.

Figures:
20. The SNF5 polyglutamine domain forms liquid droplets and filamentous aggregates in a pH-dependent manner. 60
21. Osmotic shock stress improves the relative fitness of the snf5Δ strain. 60

References. 61
Chapter 1: Introduction

Life has to constantly respond and evolve with the ever-changing environments on planet Earth. Without this capacity, life would have hardly produced the diversity we observe today. The adaptation mechanisms are diverse, but the capacity to rapidly modify gene expression plays a central role. The groundbreaking work of Jacques Monod and Francois Jacob in 1961 (Jacob and Monod, 1961) showed that *Escherichia coli* cells constantly repress the *lac* operon which contains the genes required for lactose transport and breakdown, while they have access to their preferred carbon source - glucose. This repression is maintained even in media that contains a mixture of glucose and lactose. However, as soon as glucose is depleted, the *lac* operon is de-repressed, allowing synthesis of the enzymes required for lactose metabolism. This was the first example of genetic regulation, which has since become one of the hottest fields of molecular and cellular biology.

Multiple forms of genetic regulation have been described, from single gene or operon, such as the one described by Monod and Jacob, to more extended reprogramming events where a large number of genes are either turned on or off. For instance, during the development of multicellular organisms, all cells come from a single cell, therefore multiple waves of transcriptional reprogramming are required for proper development. A single cell example of the same reprogramming would be starvation in baker’s yeast, where, upon glucose depletion, a whole survival program must be activated to ensure viability during starvation. We can hypothesize that such major reprogramming events would require a more penetrant signal, which can rapidly reach every protein involved in the response, informing them that a reprogramming event is taking place.

My dissertation is an exploration of pH as that major signal to inform yeast cells of nutrient deprivation. We identified polyglutamines (polyQ), an intrinsically disordered domain with a crucial role in the transcriptional machinery, as a receptor of the pH signal. Proper interpretation of this pH signal by a polyQ domain is required for cells to respond properly and thus ensure survival in a perilous and competitive environment. To contextualize this work, I will first introduce pH regulation of the cytosol and its coordinated acidification during starvation. Next, I will describe polyQs in the contexts where they have been studied the most: disease and as part of the transcriptional machinery. Finally, I will introduce the *Saccharomyces cerevisiae* SWI/SNF complex, which we have used a model system for this study.
Intracellular pH regulation

pH is tightly regulated in the cell. pH influences almost all biological processes by affecting the structure and function of nearly all proteins (Whitten, Garcia-Moreno E. and Hilser, 2005). Therefore, pH is one of the most important variables for the cell to control. To produce metabolic fuel, Eukaryotic cells generate a proton gradient across the inner membrane of the mitochondria, thus placing the hydrogen ion at the center of energy metabolism. Multiple organelles also contain membrane pumps that produce proton gradients and help to regulate pH. Indeed, the necessity to provide optimal pH conditions for the metabolic pathways that take place in each organelle, likely played a critical role in the evolution of compartmentalization.

Due to its importance, pH is regulated (and protected) at different levels. The fastest response to pH change comes from the buffering power of bicarbonate together with a minor help of weak acids and bases of amino acid lateral chains. Because of respiration, cells are always exposed to CO\(_2\), an uncharged gas that crosses biological membranes. CO\(_2\) is hydrated and subsequent deprotonation of carbonic acid produces bicarbonate. The bicarbonate – carbonic acid buffer system plays a central role in the regulation of blood pH. However, while this buffering mechanism is fast, it is also limited. Therefore, the cell requires another level of pH protection for continuous pH stresses. For instance, *Saccharomyces cerevisiae*, strongly prefers to grow in acidic environments but its cytosolic pH is robustly maintained at pH 7.4. It is well-known that cytosolic pH is regulated by the coordinated work of the plasma membrane proton pump PMA1, along with the V-ATPases of the vacuole (Martínez-Muñoz and Kane, 2008) which constantly extrude protons from the cytosol to the extracellular space and vacuole respectively. Proton-exchanging transporters as well as proton coupled transporters such as the NHE family (Kojima *et al.*, 2012), also play important role in pH homeostasis of mammalian cells. Indeed, NHE regulation controls cell proliferation, it is known that a mild alkalization of the cytosolic pH is required for exit of G0 (Quiescence), this increase in pH is specifically needed for mitogen activation and is facilitated by the NHE proton coupled transporter (Pomyssegur *et al.*, 1985) (Fig. 1).

The development of a genetically encoded pH sensor, pHluorin, by Miesenbock et al 1998 (Miesenböck, De Angelis and Rothman, 1998) has facilitated the study of cytosolic pH. pHluorin is a GFP-derived protein, which does not need any external input, besides excitation light, enabling direct pH measurements in live cells. pHluorin has allowed a deeper exploration of the genetic components of intracellular pH regulation. It has been shown that pH control of the cytoplasm and organelles occurs at a systems-level with more participants than the V-ATPases and PMA1. Cytosolic pH is very robust and in a screen of the yeast deletion collection, no single mutant was shown to have a pH deviating lower than 0.3 or higher than 0.5 units compared to wild-type (Orij *et al.*, 2012). Among the genes that had an effect, various mitochondrial proteins were identified (Orij *et al.*, 2012). Mitochondria have an alkaline lumen of pH 8, due to the constant extrusion of protons to the intermembrane space to create the electrochemical gradient required for ATP synthase. Hence, one may hypothesize that the electron transport chain plays a role in pH regulation. However, in the presence of oxygen and glucose, baker’s yeast strongly prefer fermentation over respiration to obtain energy, a phenomenon known as the Crabtree effect (Crabtree, 1928). Because of the Crabtree effect, the screen was
conducted in yeast cells with little to no respiration. Suggesting that mitochondria plays a regulatory pH role, independent of respiration, and therefore the proton gradient that fuels ATP synthase.

For proper pH regulation, it is essential that the cell must sense pH changes. Extracellular pH is sensed through the RIM pathway. As mentioned previously, Baker’s yeast prefers acidic environments and the RIM pathway plays an essential role in responding to alkaline pH stress. Alkaline pH activates the pathway through the plasma membrane proteins Dfg16 and Rim21 which, through a cascade of reactions, activate the Rim101 and Crz1 transcription factors which are required for the transcriptional response to alkaline pH stress (Viladevall et al., 2004).

It has been hypothesized that cytosolic pH oscillates during the cell cycle. Although there has been much discussion regarding this idea, it is not widely accepted. It is very clear that different organisms regulate their pH differently. For instance in the case of Schizosaccharomyces pombe, no fluctuation in pH is seen during the cell cycle, while various common stresses, such as osmotic stress or heat shock, did have an impact (Karagiannis and Young, 2001). Cell cycle pH variations are thought to be about 0.3 to 0.5 pH units in Saccharomyces cerevisiae, where the lowest pH is reached at Start and the highest during the G2/M transition. This oscillation is thought to be required for proper microtubule function during the cell cycle, particularly the pH drop during cytokinesis, required for proper chromosome segregation (Gagliardi and Shain, 2013).

**Acidification of the cytosol is a hallmark of starvation**

Although the presence of pH changes during the cell cycle are still controversial, changes in cytosolic pH of Saccharomyces cerevisiae cultures during the 3 distinctive stages of the exponentially growing culture, have been clearly shown. Early exponential phase cultures have a characteristic cytosolic pH of 7.4. However, during mid-to-late exponential phase, glucose concentration is lower than in the early culture, resulting in a mild acidification of ~0.3 pH units in the cytoplasm, this mild acidification is rapidly reverted by re-addition of glucose. When the culture reaches stationary phase as a result of glucose depletion, a bigger pH drop is observed, bringing cytosolic pH below 6 (Orij et al., 2012) (Fig. 2).

Acidification of the cytosol upon glucose-starvation is a very interesting phenomenon. Several nutritional changes have an effect on cytosolic pH (Orij et al., 2009), but the one produced by glucose-starvation is the largest in magnitude, affecting the biggest compartment in the cell. The pH drop upon glucose-starvation is due to inactivation of the transmembrane proton pump at the plasma membrane PMA1. As mentioned before, PMA1 is essential for cytosolic pH regulation (Portillo, 2000). However, maintaining a cytosolic pH of 7.4 in an acidic environment has a high energetic cost which has been calculated to be up to 20% of the cellular ATP (Morsomme, Slayman and Goffeau, 2000). It is precisely because of its high energy consumption that upon glucose-starvation, PMA1 is rapidly inactivated (Portillo, 2000). It is unclear if V-ATPases collaborate in the acidification of the cytosol, however, it has been shown that nutrient starvation produces a rapid disassembly of the two major V-ATPase subcomplexes, V1 and V0, resulting in inactivation of the transporter (Kane, 1995). Therefore a “hold” in the transport of protons into the vacuole could also affect cytosolic pH. Initially, pH drop upon glucose-starvation
was thought to be a mere consequence of energy saving by inactivating PMA1. However, recently there has been increasing evidence that this pH drop is actually required for proper response to starvation. Indeed, from our own results we see that even in glucose-starvation at neutral pH, the cytoplasm is still slightly more acidic than the environment. Suggesting a systemic effort to acidify upon glucose-starvation. What is the role of pH drop during glucose-starvation?

A few examples from the literature indicate that pH may play a regulatory role upon starvation. The metabolic enzyme Glutamine synthase (Gln1) aggregates into a filament that inactivates the enzyme. Acidification of the cytosol as a result of glucose depletion is required for this phenotype. When glucose is added back, cytosolic pH is reestablished to 7.4 and the filament dissolves releasing active enzyme (Petrovska et al., 2014). This storage is thought to allow energy conservation during quiescence without having to degrade large amounts of metabolic enzymes. Indeed, multiple screens in Saccharomyces cerevisiae have shown that over 200 proteins aggregate upon nutrient depletion, all major metabolic pathways have members among these proteins (Narayanaswamy et al., 2009; O’Connell et al., 2014). However, is not known how many of these proteins require a pH drop to aggregate.

Mobility of chromatin and mRNPs are reduced during glucose-starvation. To test if this was a pH dependent phenotype, a weak organic acid, potassium sorbate, which crosses the plasma membrane allowing manipulation of the cytoplasmic pH was used. Acidification of the cytosol alone, without nutrient depletion, only partially recapitulates the mobility reduction of chromatin and mRNPs. Interestingly, glucose-starvation at pH 7.4, where pH drop is prevented (given that the cytoplasm is already in equilibrium with the environment), does not reduce mobility of chromatin or mRNPs, indicating that pH drop is required but not sufficient for this phenotype (Joyner et al., 2016).

The effect of pH on chromatin and mRNPs, two macromolecules of very different sizes, suggests that pH could well be affecting most macromolecules in the cell. Munder et al, 2016 proposed that acidification produces a phase transition in the cytoplasm, changing from fluid to solid-like as a result of a major reorganization of proteins into higher order structures (Munder et al., 2016). Under this solid-like state of the cytoplasm, organelle mobility is reduced and metabolism halted (Fig. 1). However, studies in our own laboratory have shown that macromolecular mobility under various conditions strongly depends on the size of the molecule. The mobility of small molecules, like GFP, are not affected by glucose-starvation as bigger macromolecules like chromatin or mRNPs are (Delarue-Brittingham et al 2017, submitted).

Nevertheless, nutrient starvation not only turns things off, but activates some genes to ensure continued viability. For instance, during glucose-starvation, while transcription is generally downregulated, a subset of genes are highly upregulated. These genes, collectively known as glucose-repressed genes, are distributed in two major groups: stress-resistance and metabolic pathways to oxidize poor carbon sources (Zid and O’Shea, 2014). Therefore, expression of these genes is absolutely essential for viability under starvation. However, whether acidification of the cytoplasm plays a regulatory role in their transcription remains unknown.
One example of direct pH regulation of transcription is through the transcription factor Opi1. Opi1 is part of an Endoplasmic Reticulum (ER) lipid-sensor complex which binds to the ER-anchored protein Scs2 and phosphatidic acid (PA), but only when the head group of PA is deprotonated. Therefore, upon glucose depletion and cytosolic pH drop, the binding affinity of Opi1 to PA decreases, releasing Opi1 from the ER to the nucleus where it represses the Ino2/4 transcriptional activator complex, coupling nutrient availability with membrane biogenesis (Young et al., 2010).

However, the transcriptional machinery does not need to leave the nucleus to be subjected to these pH changes. The nuclear pore complex allows passage of molecules up to 9 nM in diameter, therefore the nuclear membrane does not act as a pH barrier between the cytoplasm and the nucleus. This means that nuclear pH undergoes the same fluctuations that cytoplasmic pH does upon nutrient depletion.

One interesting characteristic of transcriptional machinery is that the function of most DNA-binding proteins is divided in two distinct domains within the protein, a Trans-Activating Domain (TAD) generally found at the N-terminus and a DNA-Binding Domain (DBD) at the C-terminus (Latchman, 1997). As suggested by their names, TADs are required for regulation and DBDs provide specificity and binding to DNA. A general characteristic of TADs is that they lack a defined structure. Encoded by low complexity sequences, TADs are intrinsically disordered domains within proteins (Wells et al., 2008)(Arai et al., 2015). There are four types of TADs described in the literature: Acidic activators, such as GAL4 and GCN4, Proline rich, domains such as C-Jun, Isoleucine rich domains such as NTF-1 and finally Glutamine rich TADs; the first Glutamine rich TAD described was Sp1 at Robert Tijan’s laboratory (Kadonaga et al., 1987, 1988). Multiple glutamine rich TADs have been described since, including several members of the SWI/SNF and the Mediator complex.

Several studies has been recently published showing that Intrinsically disordered proteins (with a high enrichment for DNA and RNA-binding proteins) form multivalent structures that phase separate from liquid solutions (Brangwynne et al., 2009; Kato et al., 2012). This is a growing field and many processes in the cell are thought to require formation of these phase separated structures. However, regulation of the assembly and disassembly of these liquid droplets or hydrogels is still poorly understood. We hypothesize that the weak binding affinities required to form phase separated structures are highly susceptible to environmental variables such as pH, molecular crowding or redox state.

How do the TADs or activation sequences of the transcriptional machinery respond to pH changes? Does pH play a regulatory role in transcription of glucose-repressed genes? To address these questions we investigated the relationship between pH and polyQ TADs.
**Polyglutamines**

Polyglutamines (polyQs) have been traditionally studied in the context of neurological diseases. This is because toxic aggregates seeded by polyQ containing proteins are believed to be the molecular basis of 9 neurological diseases including Huntington's disease (Bates and Benn, 2002). All these diseases are both extremely debilitating and age dependent, therefore in a world where life expectancy continues to increase, the study of polyQ diseases has attracted a lot of attention. Glutamine is a polar amino acid; homo repeats form insoluble aggregates in polar solutions. However, the exact aggregation mechanism has yet to be discovered (Nakano, Ebina and Tanaka, 2013).

Despite that polyQ aggregation mechanism is unknown, it is very clear that the context where polyQ are have a strong effect on the aggregation degree. Therefore, the residues within glutamine rich regions or in the boundaries of pure polyQs can either induce stronger aggregation or decrease it. A proline homo repeat at the C-terminal of the Huntingtin protein-polyQ prevents aggregation while 17 residues with an overall positive charge at the N-terminal have the opposite effect, inducing stronger aggregation (Thakur et al., 2015; Shen et al., 2016). These results suggest that the amino acids around polyQs may serve as a regulatory unit of the aggregation state of polyQs.

Despite polyQs being studied as seeders of aggregation, polyQ-containing proteins are quite common in nature, beyond the 9 that cause neurological diseases. PolyQs are even found to be conserved between evolutionary distant species such as Humans and *Saccharomyces cerevisiae*. In contrast, polyQs are almost completely absent in *Schizosaccharomyces pombe* or bacteria (Schaefer, Wanker and Andrade-Navarro, 2012). Evidence for purifying selection to maintain polyQ repeats comes from the *Drosophilid* lineage. Some *Drosophila* species with a low GC content encode their polyQs with the CAA codon, instead of CAG most commonly found in species with high GC content. Therefore, despite the differences at the genomic sequence, polyQs in *Drosophilids* are conserved at the protein level, (Huntley and Clark, 2007) suggesting that polyQs are strongly conserved through evolution.

PolyQs have been hypothesized to function as flexible spacers in between protein domains as for other low complexity sequences (Huntley and Golding, 2000; Faux et al., 2005). However, a more important insight into the role of polyQs comes from sequence analysis; nuclear proteins and components of transcriptional machinery are more likely to have polyQs than proteins from other Gene Ontology (G.O.) groups (Albà, Guigó and Guigo, 2004).

**PolyQs of the transcriptional machinery**

How polyQs regulate gene expression remains unknown. However, the literature provides some insights. *Bicoid* a maternal determinant required for development of the *Drosophila* embryo, contains a polyQ activation domain. In an experimental framework that now has been used multiple times, similar to the two hybrid system, the authors cloned a reporter gene downstream of a *GAL4* DNA-binding domain. The *GAL4*-binding domain alone failed to induce expression of the reporter gene. However, *GAL4*-binding domain fused to *Bicoid* polyQ did activate expression of the reporter gene. Interestingly, increasing the length of the polyQ also increased the expression of the reporter gene (Janody et al.,
Laurent B. et al 1990, did a similar experiment where the polyQ of the yeast SNF5 is fused to the DNA-binding protein Lex A. Instead of increasing the length of the polyQ, as in the previous example, they cloned multiple contiguous binding sites of Lex A upstream of the reporter gene. They observed that induction is 3 fold higher in the strain with multiple sites compared to single site. Geng et al 2001 did a similar experiment with similar results (Laurent, Treitel and Carlson, 1990; Geng, Cao and Laurent, 2001). These results indicate that increasing the local concentration of polyQs at the promoter increases transcriptional activation, suggesting that polyQs may be forming a higher order structure for which multiple units are required.

This mechanism of transcriptional activation resembles the one described for the FET proteins. Some Sarcomas are produced by the mutation of the low complexity sequences of FUS (Fused in sarcoma), EWS (EWING sarcoma) and TAF15 which are translocated and fused to various DNA-binding proteins resulting in transformation to cancer cells. These proteins are collectively called FET proteins (Hoell et al., 2011). The proposed mechanism for driving aberrant gene expression relies on the formation of a hydrogel that is required for RNA PolII recruitment. These proteins form hydrogels at very high concentrations, unlikely to be found on a cell. However, when fused to a DNA-binding protein, multiple contiguous binding sites increase the local concentration past a certain threshold to produce a hydrogel at the promoter site. The hydrogel formation is critical for RNA PolII recruitment and therefore for transcription (Kato et al., 2012; Kwon et al., 2013).

PolyQs may have the capacity to phase separate; single molecule microscopy shows that polyQs made of 34 to 96 Qs are found to coexist in three different states: as soluble protein, in small rapid-diffusing clusters and within large slow-diffusing aggregates (Li et al., 2016). The small rapid-diffusing clusters resemble the liquid droplets and hydrogels described by Brangwynne et al., 2009 and Kato et al., 2012. However, these are structures seen in cells overexpressing the polyQ peptide. It remains to be demonstrated if this occurs with the polyQs of the transcriptional machinery during transcription. This hypothesis of phase separation for transcriptional control is becoming popular in the scientific community, however the difficulty of imaging phase separation in live cells has made it difficult to test experimentally (Hnisz et al., 2017).

To investigate how polyQs, part of transcriptional activation sequences, respond to pH changes produced by nutrient limitation in Saccharomyces cerevisiae, we used the SWI/SNF complex where both major regulatory subunits SNF5 and SWI1 contain a polyQ activation sequence, critical for expression of glucose repressed genes (Prochasson et al., 2003; Ferreira et al., 2005; Biddick et al., 2008; Biddick, Law and Young, 2008).

The SWI/SNF complex and its polyQ activation sequences.

The SWI/SNF (SWIched/Sucrose Non-Fermentable) also known as BAF, is a chromatin remodeling complex made of 15 subunits. It was the first described chromatin remodeling complex. Originally, discovered in yeast, SWI/SNF is highly conserved from yeast to mammals (Peterson and Herskowitz, 1992; Chiba et al., 1994; Peterson, Dingwall and Scott, 1994). The SWI/SNF complex regulates gene expression by evicting nucleosomes in an ATP-dependent manner, to then recruit the transcriptional machinery to the pre-
initiation complex. It acts as an inducer and repressor, regulating up to 10% of the genes in *Saccharomyces cerevisiae* (Sudarsanam *et al.*, 2000). Despite being a chromatin remodeling complex, SWI/SNF is also required for maintenance of transcription, suggesting that it also plays a role when RNA PolII has left the pre-initiation complex (Sudarsanam *et al.*, 1999).

Various mutations of SWI/SNF subunits have been linked to cancer. Indeed, 20% of tumor malignancies have SWI/SNF mutations (Shain and Pollack, 2013). However, not all mutations produce the same phenotype. In yeast, deletion of different subunits of the SWI/SNF impairs gene expression differently: some deletions prevent assembly of the complex at the promoters but others do not impair loading of the complex into the promoter, however gene expression is still altered (Dutta *et al.*, 2017). These results indicate that each subunit provides some degree of specificity to the complex. Null mutations of the SMARCB1 homolog of yeast SNF5, is the most characteristic mutation of childhood malignant rhabdoid tumors. Deletion of the human SMARCB1 as well as the yeast SNF5, impairs complex formation, causing disassembly of the complex at most promoters or enhancers typical of genes involved in lineage specification. However, SWI/SNF is not dissolved at super-enhancers, which are involved in cell identity, suggesting that null mutations on SMARCB1 impairs the balance between differentiation and cell renewal inducing tumorigenesis (Versteege *et al.*, 1998; Xi *et al.*, 2009).

The polyQs of SWI/SNF have also been linked to cancer. 100% of synovial sarcomas present a mutation called SS18-SSX, in which SS18 a glutamine rich subunit of the SWI/SNF is translocated and fused to the SSX (1 or 2) gene. SSX genes are transcriptional repressors and while the entire SS18 gene is translocated and fused, only the 78 last residues of the SSX gene are included (Jones *et al.*, 2016). This mutation is sufficient to drive sarcoma formation in mouse models (Haldar *et al.*, 2007). The translocation of the SS18 gene leads to the transformation, resemblances that caused by the FET proteins in other types of sarcomas. However, it has not been demonstrated that SS18-SSX fusion proteins form hydrogels at the promoter, as the FET proteins do.

The work presented in this dissertation stablishes a functional link between cytoplasmic pH oscillations and transcriptional regulation through the polyglutamines of the transcriptional machinery. We present evidence that a pH-driven transcriptional reprogramming through polyglutamines could be a general mechanism for transcriptional control.
Figure 1 Intracellular pH regulation. Cytosolic pH is robustly maintained at 7.4 in optimal conditions. A summary of the mechanisms for such maintenance are listed in the figure, bicarbonate – carbonic acid buffer system, the membrane H-ATPase pump PMA1 and the NHE family, the V-ATPases of the vacuolar membrane, the electron transport chain of the mitochondria and the RIM pathway.
Figure 2 Acidification of the cytosol during starvation. Upon glucose-starvation the PMA1 pump is inactivated and, possibly aided by V-ATPase inactivation, the cytosol is acidified as protons enter from the acidic external environment. If the environmental pH is neutral, no drop in the cytosol is observed. Top right panel: pH-driven phenotypes observed during starvation.
Chapter 2: Material and Methods

Cloning and yeast transformations

Yeast strains used in this study are S288c strain-background (BY4743). The sequences of all genes in this study were obtained from the *Saccharomyces cerevisiae* genome database (http://www.yeastgenome.org/).

We cloned the various SNF5 mutants into plasmids from the Pringle collection (Longtine *et al.*, 1998). We assembled plasmids by PCR or gene synthesis (IDT gene-blocks) followed by Gibson cloning (Gibson *et al.*, 2009). Then, plasmids were linearized and used to replace the WT locus by sigma homologous recombination at both ends of the target gene.

The ΔQ-SNF5 gene lacks the N-terminal 282 amino acids that comprise a polyQ rich low complexity domain. Methionine 283 serves as ATG for the ΔQ-SNF5 gene. In the HloA-SNF5 allele, Histidine 106, 109, 213 and 214 were replaced by alanine using mutagenic primers to amplify three fragments of the polyQ region and then reassemble the gene by Gibson assembly into a SNF5 parent plasmid linearized with BamH1 and Sac1.

We noticed that the slow growth null strain phenotype of the snf5Δ was partially lost over time, presumably due to suppressor mutations. Therefore, to avoid these spontaneous suppressors, we first supplemented WT S288c with a CEN/ARS plasmid carrying the SNF5 gene under its own promoter and the URA3 auxotrophic selection marker. Then a KanMX resistance cassette, amplified with primers with homology to the N-terminal and C-terminal of the SNF5 gene was used to delete the entire chromosomal SNF5 ORF by homologous recombination. We cured the CEN/ARS plasmid carrying WT SNF5 by negative selection against its URA3 locus by streaking for single colonies on 5-FOA plates immediately before each experiment to analyze the snf5Δ phenotype.

The ADH2 reporter was cloned into pRS collection plasmids for integration. URA3 (pRS306) and LEU2 (pRS305) were used as auxotrophic selection markers. The 835 base pairs upstream of the +1 of the ADH2 gene and the mCherry ORF were amplified by PCR and assembled into linearized pRS plasmids (Sac1/Asc1) by Gibson assembly. These plasmids were cut with Sph1 in the middle of the ADH2 promoter and integrated into the endogenous ADH2 locus by homologous recombination.

The pHluorin gene was cloned on a pRS collection plasmid for integration. URA3 (pRS306) and LEU2 (pRS305) were used for selection. The plasmid with the pHluorin gene was obtained (Orij *et al.*, 2009). We amplified the pHluorin gene and the strong TDH3 promoter and used Gibson assembly to clone these fragments into pRS plasmids linearized with Sac1 and Asc1.
A C-terminal TAP tag was used to visualize Snf5 and Snf2 proteins in Western blots. pRS plasmids were used but the cloning strategy was slightly different. A C-terminal region of the SNF5 and SNF2 genes were PCR amplified without the Stop codon. This segment does not contain a promoter or an ATP codon for translation initiation. The TAP tag was then amplified by PCR and cloned together with the 3' of SNF5 and SNF2 by Gibson assembly into pRS plasmids with linearized Sac1 and Asc1. The plasmids linearized in the snf5 3' and snf2 3' with Stul and Xbal respectively done to linearize the plasmid allowing integration it into the 3' of each gene locus by homologous recombination. Therefore, after transformation the WT promoter is upstream of the WT gene without the stop codon and fused to the TAP tag.

GFP strains, the SNF5-GFP strain was obtained by the yeast GFP collection (Huh et al., 2003), a gift of the Drubin/Barnes laboratory at UC Berkeley. The SNF2-GFP fused strain was made by the same approached used for the TAP tagged strain above.

**Culture media**

Most experiments, unless indicated, were performed in synthetic complete media (13.4 g/L yeast nitrogen base and ammonium sulfate; 2 g/L amino acid mix and 2% glucose). Carbon starvation media was synthetic complete without dextrose and supplemented with sorbitol, a non-fermentable carbon source to avoid osmotic shock during glucose-starvation (6.7 g/L YNB + ammonium sulfate; 2g/L Amino acid mix and 100 mM Sorbitol). pH was adjusted using 5 M NaOH.

**Glucose starvation**

Cultures were incubated in a rotating incubator at 30°C and grown overnight (14-16 h) to an OD between 0.2 and 0.3. Note: it is extremely important to prevent culture OD from exceeding 0.3 – and results are different if cells are allowed to saturate and then diluted back. Thus, it is imperative to obtain log phase cultures directly to obtain reproducible results. 3 milliliters of OD 0.2-0.3 culture were centrifuged at 6000 RPM for 3 minutes, re-suspended in SC Sorbitol at different pHs and washed 2 times. Finally, cells were re-suspended in 3 milliliters of SC Sorbitol. For flow cytometry, aliquots of 200 uL were taken for each time point in 96 well plates. During the course of time lapse experiments, culture aliquots were set aside at 4°C. LSR II – HTS were used for all measurements. 10,000 cells were measured for each time point.

**Cytosolic pH measurements**

The cytosolic pH measurements were made with flow cytometry or microscopy. pHluorin was used to measure pH based on the ratio of fluorescence from two excitation wavelengths. In our cytometry we used the settings for AmCyan (excitation 457, emission 491) and FITC (excitation 494, emission 520). While the AmCyan emission increases with pH, FITC emission decreases. A calibration curve was made in each experiment for each strain. To generate a calibration curve, glycolysis and respiration were poisoned using 2-deoxyglucose and azide leading to equilibration of the cytosolic pH to the extracellular
pH. Although many buffers are available, we used the calibration buffer published by Patricia Kane’s group (Diakov, Tarsio and Kane, 2013): 50 mM MES (2-(N-morpholino)ethanesulfonic acid), 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid), 50 mM KCl, 50 mM NaCl, 0.2 M ammonium acetate, 10 mM sodium azide, 10 mM 2-Deoxyglucose. Buffers were titrated to pH with HCL and NaOH to the desired pH. Sodium Azide and 2-deoxyglucose was always added fresh.

**qPCR and RNA sequencing**

For qPCR and RNA seq, RNA was extracted with the “High pure RNA isolation kit” (Roche) following the manufacturer instructions. Three biological replicates were done for qPCR and RNAseq. cDNAs and qPCR were made with iSCRIPT and iTAQ universal SYBR green supermix by Bio-Rad, following the manufacturer instructions. Samples processed were: Exponentially growing culture (+Glu) and glucose-starvation at pH 5.5 and 7.5 for 4 hours. Primers qPCR were taken from Biddick et al 2008. For ADH2 and FBP1 genes are: Forward (GTC TAT CTC CAT TGT CGG CTC)/ Reverse (GCC CTT CTC CAT CTT TTC GTA) and Forward (CTT TCT CGG CTA GGT ATG TTG G)/ Reverse (ACC TCA GTT TTC CGT TGG G). ACT1 was used as control amplification: Forward (TGG ATT CCG GTG ATG GTG TT)/ Reverse (TCA AAA TGG CGT GAG GTA GAG A).

**Western blot**

Strains containing SNF5 and SNF2 fused to the TAP tag were used. Given the low concentration of these proteins, they were extracted with Trichloroacetic acid (TCA): 3 mL or a colony (re-suspended in water) were pelleted by centrifugation for 2 min at 6000 RPM and then frozen in liquid nitrogen. Pellets were thawed on ice and re-suspended in 200 uL of 20% TCA, ~0.4 g of glass beads were added to each tube. Samples were lysed by bead beating 4 times for 2 min with 2 min of resting in ice in each cycle. Supernatants were extracted using a total of 1 mL of 5% TCA and precipitated for 20 min at 14000 RPM at 4 C. Finally, pellets were re-suspended in 212 uL of Laemmli sample buffer and pH adjusted with ~26 uL of Tris buffer pH 8. Samples were run on 7-12% gradient acrylamide gels with Thermo-Fisher PageRuler Prestained protein ladder 10 to 18 KDa. Once transferred, the membrane was blocked with 5% milk and incubated with a rabbit IgG primary antibody (which is bound by the protein A moiety of the TAP tag) for 1 hour and then with goat anti-rabbit secondary antibody. Membranes were visualized using LiCor Odyssey CLx scanner with Image studio 3.1 software. Membranes were scanned at 700 and 800 nM excitation light.
Data fitting into Gaussian curves

The ADH2 expression inferred from a fluorescent reporter ($P_{ADH2}$-mCherry) and cytosolic pH were fitted with a single or double Gaussian curve for statistical analysis, with a Matlab function developed in the lab. The choice of single or double Gaussian fit was determined by assessing which fit gave the least residuals. The height of the single or double Gaussian were used to determine the fraction of cells in each peaks. For simplicity, we quantified the height of the Gaussian rather than the area because peaks overlapped in many conditions.

Sequence analysis of polyQs

A polyQ structure in protein sequences was defined as a polypeptide sequence containing at least ten glutamines allowing any number of single or double amino-acid insertions, but broken by any interruption of three or more non-glutamine amino acid residues. For example, QQQQQAAQQQQQ and QAQAQAQAQAQAQAQAQAQ both count as polyQ, but QQQQQAAQQQQQ does not. Saccharomyces cerevisiae genome and protein sequences (S288c) were downloaded from SGD (www.yeastgenome.org). S288c transversion and transition rates were obtained from Zhu et al., 2014. We ran a computational evolution experiment with a single nucleotide mutation probability of 0.012 per nucleotide. This simulation leads to a non-synonymous change in roughly 20% of the glutamines, thus maintaining the overall integrity of the polyQ structure but introducing enough changes for statistical analysis. We ran this simulation ten thousand times using the polyQ-encoding portion of the S288c reference genome. Additionally, ten thousand simulations were performed on artificial genomes consisting of pure ‘CAA’, ‘CAG’ or ‘CAGCAA’ until we obtained the same non-glutamine frequency within the polyQ structure as found in S288c protein sequences. We defined an enrichment score as the relative frequency of a specific amino acid residue within or surrounding a polyQ normalized by the overall frequency of this amino acid residue present in the whole reference genome. Besides Saccharomyces cerevisiae, enrichment scores within and around polyQ were calculated for Drosophila melanogaster, Homo sapiens and Dictyostelium discoideum reference protein sequences (downloaded from http://www.ebi.ac.uk) (Zhu et al., 2014).
Table 1: Strains used in this study.

<table>
<thead>
<tr>
<th>Strain Code</th>
<th>Genotype Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>LH2145</td>
<td>ura3Δ0 his3Δ0 leu22Δ0 met15Δ0 SNF5</td>
</tr>
<tr>
<td>LH2090</td>
<td>ura3Δ0 his3Δ0 leu22Δ0 met15Δ0 ΔQ-SNF5::KAN</td>
</tr>
<tr>
<td>LH2974</td>
<td>ura3Δ0 leu22Δ0 met15Δ0 HtoA-SNF5::HIS3</td>
</tr>
<tr>
<td>LH2574</td>
<td>ura3Δ0 leu22Δ0 met15Δ0 SpidroinSNF5::HIS3</td>
</tr>
<tr>
<td>LH2991</td>
<td>his3Δ0 leu22Δ0 met15Δ0 SNF5 pADH2-mCherry::URA3</td>
</tr>
<tr>
<td>LH2992</td>
<td>his3Δ0 leu22Δ0 met15Δ0 ΔQ-SNF5::KAN pADH2-mCherry::URA3</td>
</tr>
<tr>
<td>LH2993</td>
<td>leu22Δ0 met15Δ0 HtoA-SNF5::HIS3 pADH2-mCherry::URA3</td>
</tr>
<tr>
<td>LH2994</td>
<td>leu22Δ0 met15Δ0 SpidroinSNF5::HIS3 pADH2-mCherry::URA3</td>
</tr>
<tr>
<td>IG0001</td>
<td>his3Δ0 met15Δ0 SNF5 pADH2-mCherry::URA3 pHluorin::LEU2</td>
</tr>
<tr>
<td>IG0003</td>
<td>his3Δ0 met15Δ0 ΔQ-SNF5::KAN pADH2-mCherry::URA3 pHluorin::LEU2</td>
</tr>
<tr>
<td>IG0005</td>
<td>met15Δ0 HtoA-SNF5::HIS3 pADH2-mCherry::URA3 pHluorin::LEU2</td>
</tr>
<tr>
<td>IG0011</td>
<td>met15Δ0 SpidroinSNF5::HIS3 pADH2-mCherry::URA3 pHluorin::LEU2</td>
</tr>
<tr>
<td>LH3486</td>
<td>ura3Δ0 his3Δ0 LEU22Δ0 met15Δ0 SNF5::KAN (CEN/ARS-SNF5::URA3)</td>
</tr>
<tr>
<td>LH3513</td>
<td>ura3Δ0 his3Δ0 met15Δ0 snf5Δ::KAN (CEN/ARS-SNF5::URA3) pADH2-mCherry::LEU2</td>
</tr>
<tr>
<td>LH3632</td>
<td>ura3Δ0 his3Δ0 met15Δ0 snf5Δ::KAN (CEN/ARS-SNF5::URA3) pADH2-mCherry::LEU2 pHluorin::NAT</td>
</tr>
<tr>
<td>LH3647</td>
<td>leu22Δ0 met15Δ0 pADH2-mCherry::URA3 SNF2-TAP::HIS3</td>
</tr>
<tr>
<td>LH3649</td>
<td>leu22Δ0 met15Δ0 ΔQ-SNF5::HIS3 pADH2-mCherry::URA3 SNF2-TAP::KAN</td>
</tr>
<tr>
<td>LH3652</td>
<td>leu22Δ0 met15Δ0 HtoA-SNF5::HIS3 pADH2-mCherry::URA3 SNF2-TAP::KAN</td>
</tr>
<tr>
<td>LH3662</td>
<td>his3Δ0 met15Δ0 (CEN/ARS-SNF5::URA) SNF5::KAN pADH2-mCherry::LEU2 SNF2-TAP::NATMX</td>
</tr>
<tr>
<td>IG0015</td>
<td>his3Δ0 met15Δ0 SNF5::KAN pADH2-mCherry::LEU2 (CEN/ARS-SNF5::URA3)</td>
</tr>
<tr>
<td>IG0019</td>
<td>his3Δ0 met15Δ0 ΔQ-SNF5::KAN pADH2-mCherry::LEU2 (CEN/ARS-SNF5::URA3)</td>
</tr>
<tr>
<td>IG0022</td>
<td>met15Δ0 HtoA-SNF5::HIS3 pADH2-mCherry::LEU2 (CEN/ARS-SNF5::URA3)</td>
</tr>
</tbody>
</table>
Chapter 3: A polyQ domain enables transcriptional reprogramming in response to a pH change.

Introduction

We chose the SWI/SNF chromatin-remodeling complex as a model to study the biological function of glutamine-rich low-complexity sequences (LCS). This complex is composed of 11 subunits, of which four contain poly-glutamines (polyQ) or glutamine-rich LCS (Fig. 3A, B). The second largest glutamine-rich LCS is found in the SNF5 regulatory subunit of the SWI/SNF complex, which contains 121 glutamine residues within the 282 N-terminal LCS (Fig. 3B).

We chose acute carbon starvation as a model for SWI/SNF function. The SWI/SNF complex is important for the expression of ~10% of the genes in Saccharomyces cerevisiae (~500 genes (Sudarsanam et al., 2000)) during normal vegetative growth, and is also required for efficient transcriptional reprogramming upon carbon starvation where transcription of most genes is down-regulated, but a set of glucose-repressed genes important for adaptation to nutritional stress are strongly induced (Zid and O'Shea, 2014). The SWI/SNF complex is required for the efficient expression of several hundred stress-response and glucose-repressed genes (Sudarsanam et al., 2000; Biddick et al., 2008).

The N-terminal glutamine-rich domain of SNF5 affects the viability of carbon starved yeast cells in a pH-dependent manner

First, we compared the fitness of wild-type yeast to strains with a complete deletion of the SNF5 gene (snf5Δ), or with a precise deletion of the N-terminal glutamine-rich domain of SNF5, referred to as ΔQ-snf5. The snf5Δ strain has a severe growth defect (Fig. 3C), but the ΔQ-snf5 grows at a similar rate to wild-type control cells. Next, we compared the viability of these strains after an acute switch to synthetic complete medium with no dextrose and maintenance in these carbon starvation conditions for 24 hours. Cells were plated before and after starvation and colonies were counted to assess viability. No difference was apparent; all strains maintained 100% viability after starvation (Fig. 3C-pH6.5 and below). Therefore, we reasoned that we would need to increase the sensitivity of our assay to reveal a phenotype for our mutants.

Budding yeast are adapted to an acidic environment, and growth media is typically at pH 4.0 – 5.5. The PMA1 plasma membrane proton pump maintains the cytosolic pH at around 7.4. During glucose-starvation, PMA1 is inactivated, possibly to conserve ATP (Bracey et al., 1998), and as a consequence the cytosol partly equilibrates with environmental pH conditions. The resultant cytosolic acidification has been reported to lead to changes in the biophysical properties of the cytosol, reducing rates of cytoplasmic diffusion and possibly driving phase-transition events (Joyner et al., 2016; Munder et al., 2016). Cytosolic acidification is important for viability during carbon starvation: when we buffered the environmental pH to 7, wild-type cell viability dropped to 50% (Fig. 3A). A
strong phenotype was revealed for the $\Delta Q$-snf5 allele under these conditions: this strain lost all viability during starvation at pH$_{env}$ 7. Furthermore, $\Delta Q$-snf5 only maintained 25% viability at pH$_{env}$ 6.75, while the wild-type strain was fully viable. In contrast, the snf5$\Delta$ strain behaved similarly to wild type in terms of viability during starvation at all pH values tested.

**Induction of the SWI/SNF target ADH2 in response to carbon starvation requires an acidic environment**

Alcohol dehydrogenase 2 ($ADH2$) is tightly repressed in the presence of glucose, and is strongly induced upon carbon starvation (Biddick et al., 2008). The SWI/SNF complex is required for the efficient and timely induction of $ADH2$ (Young et al., 2008) By RT-Q-PCR, we found that both the N-terminal glutamine rich domain of Snf5 and an acidic environment are required for transcription of $ADH2$ (Fig. 4A). After 2 hours of acute carbon starvation at pH 4, wild-type cells induce $ADH2$ transcription approximately 500-fold, while $\Delta Q$-snf5 did not induce detectable levels of transcript. Neither wild-type nor $\Delta Q$-snf5 strains induced any $ADH2$ when the starvation media was adjusted to pH 7.4.

To facilitate the study of the relationship between environmental pH (pH$_{Env}$) and $ADH2$ induction, and to enable the study of gene regulation at a single-cell level, we generated a reporter gene with the mCherry fluorescent protein under the control of the $ADH2$ promoter ($P_{ADH2}$-mCherry). We then used flow cytometry to quantify levels of red fluorescence. Time course experiments showed that full transcription and translation of the $mCherry$ protein took 24 hours, but robust induction was apparent at 6 hours. Induction of $P_{ADH2}$-mCherry was clearly bimodal for wild-type cells, indicating substantial cell-to-cell variability in $ADH2$ promoter activity under our conditions. This bimodality of $ADH2$ induction was lost in the snf5$\Delta$ null strain, and this strain induced 4-fold lower levels of $mCherry$ compared to wild-type. This data indicates that the polyQ of SNF5 is required for bimodal induction of $ADH2$. The phenotype of the $\Delta Q$-snf5 strain was far stronger, this strain failed to induce and $P_{ADH2}$-mCherry, indicating that the Snf5 subunit of SWI/SNF plays both a positive and negative regulatory role and the c-terminal domain of Snf5 plays a role in maintaining repression of $ADH2$ and the glutamine-rich N-terminus is required both for de-repression and full activation.

We next explored the effect of pH$_{Env}$ on $P_{ADH2}$-mCherry expression. Wild-type cells robustly induced $P_{ADH2}$-mCherry at all pH$_{Env}$ below 6.5, but almost completely failed to induce the reporter at pH$_{Env}$ of 7.0 and above (Fig. 4B). In contrast the $\Delta Q$-snf5 strain failed to induce $P_{ADH2}$-mCherry at any pH$_{Env}$. The snf5$\Delta$ strain showed a decrease of induction efficiency in control conditions (pH$_{Env}$ 5.5) as previously reported (Biddick et al., 2008; Biddick, Law and Young, 2008), however $P_{ADH2}$-mCherry induction was much stronger in the deletion strain than in the $\Delta Q$-snf5 strain. Furthermore, the snf5$\Delta$ strain still induced some $P_{ADH2}$-mCherry at pH$_{Env}$ 7.0, while the wild-type strain failed to induce. Taken together, these results suggest that SNF5 is a pH-sensitive regulator of SWI/SNF
function. In the complete absence of the Snf5 subunit, ADH2 is incompletely repressed even in the presence of glucose, and induction of ADH2 can occur in less acidic environments. Both the glutamine-rich N-terminus of Snf5 and an acidic environment are required for robust regulation of ADH2 induction upon carbon starvation.

**Histidines are enriched after poly-glutamine sequences in multiple species**

We analyzed amino acid enrichments around glutamine-rich low complexity sequences. We defined polyQ sequences as a polypeptide sequence containing at least ten glutamines interrupted by no more than two non-glutamine amino acid residues. Within the S. cerevisiae proteome, there is a strong enrichment (>3 fold) for histidine and proline within polyQ sequences compared to the global frequency of these amino acids (Fig. 5A). For histidine, this enrichment is even more pronounced immediately C-terminal to polyQ repeats (>4 fold). This enrichment of histidines within and immediately after polyQs is also apparent in the human, Drosophila melanogaster and Dictyostelium discoideum proteomes. This enrichment for certain amino acids could be an indication of functional importance, however the codons encoding glutamine (CAG and CAA) are similar to those for histidine (CAT and CAC). Therefore, we tested a null hypothesis that the observed structure of polyQ sequence is due to generation of polyQ repeats and then mutation of the underlying DNA sequence in the absence of selection. We therefore ran simulations where we allowed CAG, CAA and mixed CAGCAA repeats (artificial polyQ genomes) to randomly mutate at the nucleotide substitution frequencies that have been empirically described for S. cerevisiae (Zhu et al., 2014) until the polyQ had degenerated to have around 20% non-Q amino acids (the same frequency as in the true S. cerevisiae polyQ structures). The average results of 10,000 simulations are presented as grey bars in figure 3A. While this model does predict enrichment of histidines, the real polyQ structures are significantly more enriched suggesting that there is a general selective pressure to embed histidine residues within these structures.

**pH changes are sensed by histidines in the N terminal glutamine-rich domain of Snf5**

The imidazole side chain of histidine has a pKa of around 6, and chemical interactions within the polypeptide can readily shift this value up or down. Therefore, the charge of this amino acid could change as a function of pH changes similar to those reported to occur in the cytosol of budding yeast under carbon starvation (Orij et al., 2009). The N terminal glutamine-rich domain of Snf5 contains 6 histidines (Fig. 3B). Therefore, we reasoned that these amino acids might play a pH sensing role. To test this hypothesis with generated mutants by converting histidines to alanine, the side-chain of which will not change protonation state under any condition. Technical difficulties related to PCR and gene-synthesis within low complexity sequence frustrated attempts to replace all 6
histidines, however, we were able to mutate 4 of 6 histidines to alanine, as illustrated in figure 5B. We refer to this allele as $H_{toA}snf5$.

The $H_{toA}snf5$ allele was almost as severe as a complete deletion of the glutamine rich N-terminal domain. Under optimal pH conditions ($pH_{\text{Env}}$ 5.5) $H_{toA}snf5$ both the expression level and fraction of cells that induce is reduced compared to wild-type (Fig. 4A). Under less favorable conditions of $pH_{\text{Env}}$ 6.5 where WT is still strongly induced, $H_{toA}snf5$ shows no induction at all. These results indicate that removal of histidines from the $Snf5$ N-terminus desensitizes the system to pH change.

**Single cell analysis reveals a bimodal pH response to glucose-starvation, and that recovery of a subset of cells to neutral pH precedes $ADH2$ expression**

Intracellular pH can be studied using pHluorin, a GFP derivative that has been engineered as a pH indicator (Miesenböck, De Angelis and Rothman, 1998). Previous studies with pHluorin reported that the cytosolic pH ($pH_{\text{cyt}}$) of *S. cerevisiae* drops to around 6 during carbon starvation (Miesenböck, De Angelis and Rothman, 1998; Orij et al., 2009). However, these were average population measurements, partly because the original constructs were expressed from a *CEN/ARS* plasmid which leads to cell to cell variability in expression levels. We reengineered pHluorin using the strong $TDH3$ promoter and integrated this construct into the *URA3* locus. Our reengineered pHluorin gave strong expression and was less noisy, enabling single cell pH measurements. We generated strains with both the $P_{ADH2}$-$mCherry$ and $P_{TDH3}$-$pHluorin$ reporters, allowing us to investigate the dynamics of both pH change and $ADH2$ expression (Fig. 6B).

In wild-type cells, the $pH_{\text{cyt}}$ of the entire population dropped to ~6.5 immediately after glucose was removed. Then, after 30 minutes, the culture began to split into 2 subpopulations: around half of the cells further acidified to $pH_{\text{cyt}}$ 6 and this population never induced the $P_{ADH2}$-$mCherry$ reporter, while the other half recovered their $pH_{\text{cyt}}$ to ~7 and strongly induced $P_{ADH2}$-$mCherry$ (Fig. 6B, first panel, quantified in Fig. 6C, first panel).

**$SNF5$ mutants have an altered pH response upon glucose-starvation**

We also found that $Snf5$ plays a role in pH regulation during acute glucose-starvation. The $snf5\Delta$ strain did not acidify as strongly as the wild-type strain, nor did the bifurcation in pH and $ADH2$ expression occur (Fig. 6B and C, second panel). In contrast, the $\Delta Q$-$snf5$ mutant acidified to $pH_{\text{cyt}}$ 6 but then no cells neutralized back to pH 7 (Fig. 4B and C, third panel). The $H_{toA}$-$snf5$ mutant acts as a hypomorph – mostly phenocopying $\Delta Q$-$snf5$ but a small fraction of cells neutralize and induce $ADH2$ at lower levels than wild-type cells (Fig. 6B and C, fourth panel). Together, these results indicate that $SNF5$ is involved in pH regulation upon glucose-starvation, and that the N-terminal glutamine-rich domain, with histidines intact, is required for recovery to neutral pH. Thus, $SNF5$ appears
to both contribute to pH regulation and respond to pH changes through its N-terminal glutamine-rich domain and histidines therein.

**A transient decrease in cytosolic pH is sufficient for ADH2 induction**

The dynamics of pH$_{\text{cyt}}$ and $P_{\text{ADH2-}\text{mCherry}}$ suggested that transient acidification could be enough to induce ADH2 expression. We manipulated the environmental pH to test this hypothesis. Wild type cells were subjected to glucose-starvation for 2 hours at pH$_{\text{Env}}$ 5 and then the pH$_{\text{env}}$ was alkalinized to 7.5, thus forcing neutralization of the cytosol. pHluorin measurements confirmed that pH$_{\text{cyt}}$ of the entire population was forced to 7 by this treatment. Nevertheless, the transient acidification during the first 2 hours of starvation was sufficient to allow robust induction of $P_{\text{ADH2-}\text{mCherry}}$. The reverse experiment, where cells are transiently glucose-starved at pH$_{\text{Env}}$ 7.5 for 2 h and then switched to pH$_{\text{Env}}$ 5 did not lead to very inefficient induction of $P_{\text{ADH2-}\text{mCherry}}$, despite the pH$_{\text{cyt}}$ drop when switched to pH$_{\text{Env}}$ 5 after the 2$^{\text{nd}}$ hour. Together, these results indicate that ADH2 induction requires a transient drop in pH$_{\text{Cyt}}$ at the beginning of glucose-starvation treatment (Fig. 7).

**Transient acidification also occurs in starved cells upon re-addition of carbon source**

We hypothesized that cytosolic pH drop may serve as a general signal for transcription reprogramming. If this were true we hypothesize that another drop would be required when glucose is added back to the culture prior to cells and transcription switches to support normal growth. Indeed, this fast drop has been reported in the literature (Orij et al., 2009) but these results were population averages. To gain better insights into this phenomenon we performed single cell analysis of this cytoplasmic pH drop. We observed that, indeed there is a rapid, transient pH$_{\text{cyt}}$ drop that precedes recovery to neutral pH$_{\text{cyt}}$ (Fig. 8A right panel for WT). However, this drop only occurs in the subpopulation of starved cells that induced ADH2 and recovered their pH$_{\text{cyt}}$ to 6.8, and does not occur in the subpopulation that fails to induce ADH2 and maintains a low pH$_{\text{cyt}}$ of 6.0 (Fig. 8B right panel for WT). All cells do recover back to neutral pH after 30 minutes. We speculate that the acidic cells pH$_{\text{cyt}}$: 6 are in an intermediate state where reprogramming to induce glucose-repressed genes has not yet occurred, therefore another pH$_{\text{cyt}}$ drop is not required to return to the rich media program. Therefore, only the cells that actively inducing glucose-repressed genes require the second pH$_{\text{cyt}}$ drop to return to the rich media genetic program. In concordance with these results, when glucose is added back to a culture starved at pH$_{\text{Env}}$ 7, and therefore there is no inducing population, there is no observable pH decrease when glucose is added back to the medium (Fig. 8C, D).
Transcriptomic analysis indicates that the SNF5 polyQ is required for the efficient induction of a large number of glucose-repressed genes

We performed RNA sequencing analysis to determine the extent of the requirement for the SNF5 polyQ domains in the activation of glucose-repressed genes. Total RNA was extracted from WT, ΔQ-snf5 and HtoA-snf5 strains during exponentially growth (+Glu) and after 4 hours of glucose-starvation.

All three strains showed some degree of upregulation of glucose-repressed genes. However, ΔQ-snf5 and HtoA-snf5 strains upregulated less genes overall (Fig. 9A, 9B, 9C). We confirmed that ADH2, as well as FBP1, whose induction requires the same molecules as ADH2, are highly upregulated in WT but not in ΔQ-snf5 and HtoA-snf5 strains. We also observed that the mRNA levels of our fluorescent reporter mCherry are very similar to ADH2, confirming it as a valid reporter of ADH2 expression. Over 300 genes are upregulated in WT strains including ADH2 and FBP1. ΔQ-snf5 and HtoA-snf5 overlap in most of the genes upregulated, 91 genes are shared among them and only 9 are upregulated in ΔQ-snf5 and WT but not in HtoA-snf5, 56 are shared among HtoA-snf5 and WT but not ΔQ-snf5 (Fig. 9D). From the >300 genes upregulated in WT, 180 were not upregulated in ΔQ-snf5 or HtoA-snf5. Indicating that the majority of genes upregulated during glucose-starvation require the polyQ of SNF5, and presence of histidine in the SNF5. Table 2 lists the 20 most highly induced genes in WT that failed to induce in ΔQ-snf5 and HtoA-snf5 strains. As expected, many genes are involved in carbon metabolism or stress response pathways. Indeed, we were able to identify plasma membrane transporters of carbon sources (JEN1, HXT5), several mitochondrial proteins required for metabolism of poor carbon sources (SFC1, ODC1), cytoplasmic metabolism proteins (BAT2, RGI2), members of the transcriptional machinery required for expression of glucose-repressed genes (CAT8, YGR067C) and stress related (HSP26). Therefore, the polyQ of SNF5 acts as a pH-responsive switch to regulate the expression of a large number of genes.
**Discussion**

Here, we explored pH as a signal for transcriptional reprogramming of cells under carbon starvation. Hydrogen ion concentration in the cytoplasm is maintained, and can be rapidly altered by an intricate network of membrane pumps at the vacuole and plasma membrane. We showed, and it has been reported in the literature, that acidification of the cytosol upon glucose-starvation is critical for viability. When the pH of starvation media is increased, thus preventing cytosolic acidification, viability crashes (Fig. 3). It has also been shown that the ATP content of the cell crashes when cells are starved at neutral pH, potentially explaining the reduced viability.

We found that cytosolic acidification below pH 6.5 is required for expression of the glucose-repressed genes *ADH2* and *FBP1* upon glucose-starvation, which would explain the viability results. We also showed that the polyQ of *SNF5* is required for expression of these genes at all pHs. Performing single cell pH analysis, we discovered that acidification of the cytosol upon glucose-starvation is relatively homogeneous during the first ~30 minutes: all cells decrease their cytosolic pH to ~6.3 (Fig. 6). However, after this initial drop, the culture divides into two subpopulations, one with cytosolic pH ~6 and a second that increases its cytosolic pH to ~6.8. This second subpopulation then induces the glucose repressed gene *ADH2*, while the acidic population does not. We also showed *ADH2* expression fails if the environmental pH is above 6.5, suggesting that a transient cytosolic pH decrease to below 6.5 is required for induction. We propose Histidine as the residue in charge of sensing such pH drop. The imidazole ring in the lateral chain of Histidine has a pKa of 6. However, this pKa is sensitive to the local environment and could be close to 6.5 in the context of the *SNF5* polyQ. Finally, our RNA-seq analysis suggests that this pH-responsive polyQ switch regulates the majority of glucose-repressed genes that require SWI/SNF for their activation.

We propose that proper downregulation of general transcription and upregulation of glucose-repressed genes requires a pH$_{\text{Cyt}}$ drop. In this scenario, a rapid pH drop would help to disassemble transcriptional complexes, during the 1st to 2nd hour of glucose-starvation. The transcriptional machinery must then re-localize and reassemble on the promers of glucose-repressed genes. Finally, an increase in pH$_{\text{Cyt}}$ allows transcriptional activation of glucose repressed genes (Fig. 6).

We found that all strains generally decrease transcription upon glucose-starvation (Fig. 6). Considering the fact that induction of glucose repressed genes in the ΔQ-*snf5* strain is not upregulated to the same extent as in WT, transcription in ΔQ-*snf5* is therefore considerably more inhibited than in *SNF5* after 4 hours of glucose-starvation. Therefore, cells in the acidic state (pH$_{\text{Cyt}}$:~6) could be in an intermediate state, after transcriptional repression but before activation of glucose repressed genes. It has been reported that another pH$_{\text{Cyt}}$ drop takes place upon re-addition of glucose to cells under glucose-starvation (Orij *et al.*, 2009) suggesting that a transient pH change is required for reprogramming in both directions. However, these are population averages; according to our data, the pH$_{\text{Cyt}}$ of cultures under glucose-starvation is bistable, with one group at
pH_{cyt} \sim 6 and a second group at pH_{cyt} \sim 6.8 (Fig. 6). According to our hypothesis, only the latter group would need a second pH drop, in order to disengage from transcription of glucose-repressed genes, while the acidic group does not need this drop because those cells are in an intermediary state (Fig. 9). Indeed, we observed that only cells expressing ADH2 with a cytosolic pH of \sim 6.8, decrease their cytosolic pH upon glucose addition, before all cells increase their cytosolic pH to 7.4 and resume growing (Fig. 8). Together, these results suggest that the transient pH_{cyt} drop may only be required for cells that have engaged an active transcription program, not those in an intermediate repressed state.

We also had an unexpected result regarding regulation of cytosolic pH by SNF5. A Screen conducted to identify genes required for pH homeostasis, (Orij et al., 2012), did not identify SNF5. This is not surprising as the screen was performed in optimal growth conditions in which SNF5, snf5Δ and ΔQ-snft5 have the same cytosolic pH (Fig. 6A “Glu”). However, upon glucose-starvation the pH_{cyt} dynamics of these strains are very different. While ΔQ-snft5 decreases its cytosolic pH and then fails to recover, the snf5Δ strain does not acidify to the same extent as WT, and most cells stay above pH_{cyt} 6.5 during the whole course of the experiment (Fig. 6A). These results indicate that SNF5 plays a role in pH regulation upon glucose-starvation: The C-terminus appears to be required for full acidification upon starvation and the polyQ N-terminal domain is required for the partial recovery of pH in cells that induce ADH2. This latter result could be explained by the requirement for transcriptional events to drive pH recovery, but the time-scales of recovery are probably too rapid and, there is no clear candidate for a gene that would drive the pH recovery in our RNA-seq data. However, a more careful dissection of the upregulated genes could provide insights in this phenomenon.

It is clear that pH has a role in maintenance of viability and induction of glucose repressed genes during starvation (Fig. 3, 4, 5, 6, 7 and 8). However, exactly what occurs to polyQs upon acidification is still unclear. One hypothesis is that polyQs undergo phase transitions in a pH-dependent manner. Multiple low complexity domains have been shown to form hydrogels and liquid droplets (DNA and RNA-Binding) (Brangwynne et al., 2009; Kato et al., 2012). PolyQs are likely to form similar structures (Li et al., 2016). The phase transition behavior of polyQs can be strongly impacted by surrounding amino acids. The addition of two negatively charged residues at the amino end and two positively charged residues at the carboxyl end of the polyQ domain of the yeast Sup35 solubilizes the polyQ at acidic pH and aggregates at neutral pH while a polyalanine flanked by the same charged amino acid is soluble at all pH (Perutz et al., 2002). These results reinforce the hypothesis that amino acids within and at the boundaries of polyQs could play a regulatory role. We observed that the SNF5 polyQ required its histidine residues for proper regulation (Fig. 5). Therefore a dynamic state of protonation of histidine depending on cytosolic pH could drive a phase transition. Two different hypotheses could be tested: First that polyQs form a phase separated structure during transcription, at cytosolic pH 7.4 and is dissolved, and therefore disengaged from transcription at acidic pH to then be formed again at the promoter of glucose repressed genes when pH increases and transcription is activated. A second hypothesis would consider the polyQs “soluble” at pH
7.4 while Snf5 is engaged in transcription to then aggregate at acidic pH as a way to disengage from transcription. Further increase of pH would then dissolve the aggregate allowing re-distribution of the polyQs of the transcriptional machinery into glucose repressed genes. Experiments to test these hypotheses are discussed in Chapter 4.
Figure 3: Deletion of the Snf5 polyglutamine leads to loss of viability in carbon starvation, especially at neutral environmental pH. A- The 288 amino acids of the polyQ domain of SNF5, glutamine is shown in green and histidine in red/orange. B- Schematic representation of the SWI/SNF according to the available structural data, in red are the subunits that contain polyQ sequences, in the bottom of B is a schematic representation of the SNF5 gene with its polyQ domain indicated at the N-terminus (Green). C - Viability after 24 h of glucose-starvation at pH_{Env} 6.5 (Yellow), 6.75 (Light Green) and 7.0 (Dark Green). D- Growth curves for WT, ΔQ-snf5 and snf5Δ strains.
Figure 4: ADH2 expression requires an acidic environmental pH and the SNF5 polyglutamine. A- RT-QPCR experiments to determine ADH2 mRNA content of each strain after 4 h of glucose-starvation at pH_{Env} 6.0 (orange) and pH_{Env} 7.5 (blue). ACT1 mRNA was used as a control, Y axis corresponds to fold induction of ADH2 mRNA on glucose-starvation over exponentially growing cells. B- Cytometry data histograms of P_{ADH2-mCherry} induction. Environmental pH is indicated by color codes in B and C: pH 5.5 (red), pH 6.0 (orange), pH 6.5 (yellow), pH 7.0 (green), pH 7.5 (blue) and pH 8.0 (purple). C- Expression of P_{ADH2-mCherry} fluorescent reporter integrated at the ADH2 locus, upon starvation at different pH. Raw data was quantified by fitting to single or double Gaussian models (see material and methods) to quantify relative levels of gene induction (bar height) and the fraction of cells inducing (bar shade).
Figure 5: Histidines are important for the function of the *SNF5* polyglutamine domain. A- Sequence analysis results for all *Saccharomyces cerevisiae* polyglutamines, showing frequencies of each amino acid within polyQ structures (blue), and immediately C-terminal to the structures (the +1 position, green). We also show expected frequencies generated for a null-hypothesis where all sequences are generated from a pure polyglutamine, followed by neutral evolution – results are averages of 10,000 simulations (grey). B- Cytometry data showing levels of induction of $P_{ADH2}$-mCherry after 6 h glucose-starvation in a strain harboring an allele of *SNF5* in which 4 of 6 histidines in the polyglutamine are mutated to alanine ($H^4toA$snf5). C- Data from B was quantified by fitting to a single or double Gaussian model. Expression levels are relative to wild type expression.
Figure 6: Single cell analysis reveals a bifurcation in behavior upon glucose-starvation: One subpopulation acidifies to pH 6 and fails to induce, while a second subpopulation transiently acidifies and then induces ADH2. A- Dot plots of cytometry data quantifying cytosolic pH (x-axis) and induction of $P_{ADH2}$-$m$Cherry (y-axis). Columns show WT, snf5Δ, ΔQ-snf5 and HtoAsnf5 strains (left to right). Rows are time points starting from with logarithmically growing cells prior to dextrose wash out “Glu”. 10,000 cells were sampled at each time point. B- Schematic representation of the two fluorescent reporters used for this experiment. C- Raw data from A was fitted to single or double Gaussian models (see materials and methods). Top panels show quantification of cytosolic pH, each time point has one or two peaks and where there are two lines (black and grey), these represent the median pH values for the more neutral and more acidic populations respectively. Bottom panels show levels of induction of the $P_{ADH2}$-$m$Cherry expression. Bar height indicates to the intensity of mCherry fluorescence (normalized to maximum values for the WT control strain) and the color code represents the percentage of the population that induces the $P_{ADH2}$-$m$Cherry reporter above background levels.
Figure 7: Transient acidification of the cytosol is sufficient for ADH2 expression.
A- Outline of the experiment, exponentially growing cultures are divided in four: two of them washed into glucose-starvation media at pH 5 (red), and the other two into starvation media at pH 7.4 (green). After two hours of glucose-starvation, one culture at pH 5 was switched to 7.4 and one culture that was at pH 7.4 was switched to pH 5. B- $P_{ADH2}$-mCherry reporter expression, x-axis is time in hours (pH switch occurred at 2 h), y-axis is the percentage of cells that induced the $P_{ADH2}$-mCherry reporter. C- Cytosolic pH histograms, culture in complete media is labeled “Glu”, red histograms indicate cultures starved at pH 5, and green histograms indicate cultures starved at pH 7.4. y-axis is time in hours. D- Histograms of $P_{ADH2}$-mCherry reporter expression over time, same color code as C.
Figure 8: Transient cytosolic acidification also occurs upon re-addition of glucose to starved cells. A- Histograms of cytosolic pH of the strains used in this study, starting from a 6 h starved culture in environmental pH 5.5 (black). Glucose was added at time 0. B- Dot plots showing cytosolic pH (x-axis) and $P_{ADH2-mCherry}$ reporter expression (y-axis) and cytosolic pH histograms of the same samples of A. C and D show the same experiment but starting with cultures starved for glucose at pH_{env} 7.5 (black).
Figure 9: RNA-seq analysis reveals widespread polyQ-dependence in the expression of glucose repressed genes. A- Volcano plot of WT in “Glu” vs carbon starvation: x-axis is the natural logarithm of the ratio of expression in dextrose versus carbon starvation with repressed genes to the left of the origin and induced genes to the right, y-axis is the p-value for difference (negative log
scale). Red dots indicate genes expressed to significantly different levels. B and C show the same volcano plots for ΔQ-
strains. D- Venn diagram of the genes upregulated during glucose-
starvation.
Table 2, top 20 highly expressed genes upon glucose-starvation not upregulated in ΔQ-snf5 or HtoAsnf5.

<table>
<thead>
<tr>
<th>Systematic Name</th>
<th>Standard name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YLR377C</td>
<td><em>FBP1</em></td>
<td>Fructose-1,6-bisphosphatase; key regulatory enzyme in the gluconeogenesis pathway, required for glucose metabolism; undergoes either proteasome-mediated or autophagy-mediated degradation depending on growth conditions; glucose-starvation results in redistribution to the periplasm; interacts with Vid30p</td>
</tr>
<tr>
<td>YLR174W</td>
<td><em>IDP2</em></td>
<td>Cytosolic NADP-specific isocitrate dehydrogenase; catalyzes oxidation of isocitrate to alpha-ketoglutarate; levels are elevated during growth on non-fermentable carbon sources and reduced during growth on glucose; IDP2 has a paralog, IDP3, that arose from the whole genome duplication</td>
</tr>
<tr>
<td>YKR097W</td>
<td><em>PCK1</em></td>
<td>Phosphoenolpyruvate carboxykinase; key enzyme in gluconeogenesis, catalyzes early reaction in carbohydrate biosynthesis, glucose represses transcription and accelerates mRNA degradation, regulated by Mcm1p and Cat8p, located in the cytosol</td>
</tr>
<tr>
<td>YMR303C</td>
<td><em>ADH2</em></td>
<td>Glucose-repressible alcohol dehydrogenase II; catalyzes the conversion of ethanol to acetaldehyde; involved in the production of certain carboxylate esters; regulated by ADR1</td>
</tr>
<tr>
<td>YKL217W</td>
<td><em>JEN1</em></td>
<td>Monocarboxylate/proton symporter of the plasma membrane; transport activity is dependent on the pH gradient across the membrane; mediates high-affinity uptake of carbon sources lactate, pyruvate, and acetate, and also of the micronutrient selenite, whose structure mimics that of monocarboxylates; expression and localization are tightly regulated, with transcription repression, mRNA degradation, and protein endocytosis and degradation all occurring in the presence of glucose</td>
</tr>
<tr>
<td>YER065C</td>
<td><em>ICL1</em></td>
<td>Isocitrate lyase; catalyzes the formation of succinate and glyoxylate from isocitrate, a key reaction of the glyoxylate cycle; expression of ICL1 is induced by growth on ethanol and repressed by growth on glucose</td>
</tr>
<tr>
<td>YJR095W</td>
<td><em>SFC1</em></td>
<td>Mitochondrial succinate-fumarate transporter; transports succinate into and fumarate out of the mitochondrion; required for ethanol and acetate utilization</td>
</tr>
</tbody>
</table>
| YJR148W         | *BAT2*       | Cytosolic branched-chain amino acid (BCAA) aminotransferase; preferentially involved in BCAA catabolism; homolog of murine ECA39; highly expressed during stationary phase and repressed during logarithmic
<table>
<thead>
<tr>
<th>Gene ID</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>YIL057C</td>
<td>Protein of unknown function; involved in energy metabolism under respiratory conditions; expression induced under carbon limitation and repressed under high glucose; RGI2 has a paralog, RGI1, that arose from the whole genome duplication</td>
</tr>
<tr>
<td>RGI2</td>
<td></td>
</tr>
<tr>
<td>YAR035W</td>
<td>Outer mitochondrial carnitine acetyltransferase; minor ethanol-inducible enzyme involved in transport of activated acyl groups from the cytoplasm into the mitochondrial matrix; phosphorylated</td>
</tr>
<tr>
<td>YAT1</td>
<td></td>
</tr>
<tr>
<td>YHR096C</td>
<td>Hexose transporter with moderate affinity for glucose; induced in the presence of non-fermentable carbon sources, induced by a decrease in growth rate, contains an extended N-terminal domain relative to other HXTs; HXT5 has a paralog, HXT3, that arose from the whole genome duplication</td>
</tr>
<tr>
<td>HXT5</td>
<td></td>
</tr>
<tr>
<td>YBR072W</td>
<td>Small heat shock protein (sHSP) with chaperone activity; forms hollow, sphere-shaped oligomers that suppress unfolded proteins aggregation; long-lived protein that is preferentially retained in mother cells and forms cytoplasmic foci; oligomer activation requires heat-induced conformational change; also has mRNA binding activity</td>
</tr>
<tr>
<td>HSP26</td>
<td></td>
</tr>
<tr>
<td>YBL015W</td>
<td>Protein with CoA transferase activity; particularly for CoASH transfer from succinyl-CoA to acetate; has minor acetyl-CoA-hydrolase activity; phosphorylated; required for acetate utilization and for diploid pseudohyphal growth</td>
</tr>
<tr>
<td>ACH1</td>
<td></td>
</tr>
<tr>
<td>YPL134C</td>
<td>Mitochondrial inner membrane transporter; 2-oxodicarboxylate transporter, exports 2-oxoadipate and 2-oxoglutarate from the mitochondrial matrix to the cytosol for lysine and glutamate biosynthesis and lysine catabolism; suppresses, in multicity, an fmc1 null mutation; ODC1 has a paralog, ODC2, that arose from the whole genome duplication</td>
</tr>
<tr>
<td>ODC1</td>
<td></td>
</tr>
<tr>
<td>YDL215C</td>
<td>NAD(+)-dependent glutamate dehydrogenase; degrades glutamate to ammonia and alpha-ketoglutarate; expression sensitive to nitrogen catabolite repression and intracellular ammonia levels; genetically interacts with GDH3 by suppressing stress-induced apoptosis</td>
</tr>
<tr>
<td>GDH2</td>
<td></td>
</tr>
<tr>
<td>YNL117W</td>
<td>Malate synthase, enzyme of the glyoxylate cycle; involved in utilization of non-fermentable carbon sources; expression is subject to carbon catabolite repression; localizes in peroxisomes during growth on oleic acid, otherwise cytosolic; can accept butyryl-CoA as acyl-CoA donor in addition to traditional substrate acetyl-CoA</td>
</tr>
<tr>
<td>MLS1</td>
<td></td>
</tr>
<tr>
<td>Gene</td>
<td>Symbol</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>YHR137W</td>
<td>ARO9</td>
</tr>
<tr>
<td>YGR067C</td>
<td>YGR067C</td>
</tr>
<tr>
<td>YMR280C</td>
<td>CAT8</td>
</tr>
<tr>
<td>YOL126C</td>
<td>MDH2</td>
</tr>
</tbody>
</table>
Chapter 4: Characterization of SNF5 mutations and Exploration of a mechanism for polyQ gene expression activation.

Introduction

In this chapter, we will review further characterization of the SNF5 mutant alleles. Together with a look at other types of stresses, which may or may not induce a pH change, in order to test the hypothesis of transcriptional regulation by pH described in Chapter 3. We will also discuss conflicting results from GFP tagged strains, and how these fusion proteins may be affecting the architecture of the SWI/SNF complex.

Finally, we test an aggregation hypothesis for polyQ TADs, using an orthogonal induction of aggregation of the mutant SNF5 allele. The objective of this experiment is to provide insights into the mechanistic role of polyQs upon glucose-starvation, serving as a link to with the future perspectives in Chapter 5.

ΔQ-snf5 is a dominant negative allele

The effect of ΔQ-snf5 and HloA-snf5 on expression of glucose repressed genes is more severe than the null mutation (Fig. 4, 6). To test if ΔQ-snf5 is a dominant negative mutation we made a haploid strain that co-expresses both the WT SNF5 and ΔQ-snf5 alleles. Both genes were integrated at the endogenous SNF5 locus on chromosome II. After 6 h of glucose-starvation, the strain carrying both alleles failed to induce ADH2 (Fig. 10A). As a second approach, we expressed higher amounts of WT SNF5 from a non-integrating CEN/ARS plasmid. ΔQ-snf5 and snf5Δ strains with a CEN/ARS plasmid carrying SNF5 were compared to their counterparts after plasmid curing by selection against the plasmid-encoded URA3 marker using 5-FOA. The snf5Δ strain containing the SNF5 CEN/ARS plasmid induced the P_{ADH2-mCherry} reporter to WT levels, and then reverts to snf5Δ levels after curing (Fig. 10 compare to Fig. 4). In contrast, the SNF5 CEN/ARS plasmid fails to complement a strain containing the ΔQ-snf5 allele, and only has a minor increase (5%) in ADH2 expression compared to same strain after plasmid curing (Fig. 10B). Thus, the ΔQ-snf5 allele is dominant negative even when present at levels substoichiometric to wild type. According to the literature, the SWI/SNF contains a single Snf5 protein (Dutta et al., 2017; Sen et al., 2017). Therefore, this result suggests that multiple copies of the SWI/SNF complex operate at the ADH2 promoter. Thus, a combination the ΔQ-snf5 SWI/SNF complex acts to poison transcriptional activation even when the wild type complex is present in excess.

C-terminally tagged SNF5 is a hypomorphic allele

Structural biochemistry studies by crosslinking of the SWI/SNF, followed by proteolysis and mass spectrometry indicate that the C-terminal of Snf5 is buried within the SWI/SNF complex (Sen et al., 2017). Therefore, C-terminal fusion mutants might fail to incorporate
properly into the SWI/SNF complex. If this were the case, we would expect SNF5 and ΔQ-snf5 fused to GFP to have a null or a null-like phenotype. To test this hypothesis, we performed RT-qPCR to assess ADH2 induction during glucose-starvation. As mentioned in Chapter 3, ADH2 induction after 4 hours of glucose-starvation at pH_{Env} is ~1000 fold over the repressive condition; the null mutation reduces that induction to a half while in ΔQ-snf5 there is no induction (Fig. 4A). We found that ADH2 expression was reduced 3-fold in the SNF5-GFP strain, decreasing from 1000 fold induction to below 300 fold, almost phenocopying deletion of the SNF5 gene (Fig. 11). On the other hand, the ΔQ-snf5-GFP allele allowed a partial induction of ADH2, around ~30 fold indicating a partial failure of this dominant negative protein to integrate into the SWI/SNF complex (Fig. 11). These results are likely due to a failure of these C-terminally GFP-tagged alleles to assemble into the SWI/SNF complex.

SNF5 as well as the mutant alleles are not degraded upon glucose-starvation

Mutations can affect protein stability. Therefore, we performed Western blot analysis on TAP tagged strains to determine levels of Snf5, ΔQ-snf5 and HtoA-snf5 protein during glucose-starvation, It has been previously shown that TAP tagged Snf5 at the C-terminal immune precipitates the whole SWI/SNF complex indicating that this tag is tolerated (Yang et al., 2007). We observe similar protein levels of all alleles during glucose-starvation at both pH_{Env} 5.0 and 7.5 (Fig. 12).

Phenotypes of SNF5 mutants in various stress conditions.

As discussed in Chapter 3, the phenotypes of the WT and mutant alleles of SNF5 were markedly different from the null phenotype. To explore the extension of the conclusions from Chapter 3, we undertook growth assays with the ΔQ-snf5 and null mutant strains under various nutrient and stress conditions. As expected the response of these two mutants was very different (Fig. 13). The ΔQ-snf5 strain was more sensitive to peroxide and Dithiothreitol (DTT) stress (Fig. 13A, B). Peroxide stress also led to acidification of the cytosol to pH ~6.5 (Fig. 12C). This pH_{Cyt} drop didn’t occur in the whole population: about 40% of the population acidified after 30 min incubation with 2 mM peroxide (Fig. 13C). Total recovery to neutral pH occurred after about an hour (Fig. 13C), similar to the kinetics of acidification of cells that eventually induce ADH2 following carbon starvation. The degree of acidification upon peroxide treatment is dose-dependent with an EC50 around 1 mM – at which point the pH response of the population is bimodal, similar to carbon starvation (Fig. 13D). Whether the Snf5 polyQ domain responds to these pH changes in the context of oxidative stress remains to be determined.

The snf5Δ null mutant showed distinct phenotypes. It has been reported that snf5Δ has decreased resistance to multiple stresses, including acidic and alkaline pH (Sambade et al., 2005; Schmidt et al., 2012). Indeed, rich media at pH 4 to 7.5 does not affect the
growth rate of WT or ΔQ-snf5 (Fig. 14A). In contrast, the snf5Δ strain showed an increasing lag phase after changing growth conditions to media of progressively more neutral pH (from pH 4 to pH 7.5, Fig. 14). Maximum exponential growth rate remained approximately the same across these conditions. This result suggests that SNF5 is important for adaptation to changes in environmental pH.

**Early microscopy and polyQ aggregation**

The first experiments done for this project were microscopy of SNF5 and ΔQ-snf5 fused to a GFP molecule at the C-terminus. WT and ΔQsnf5-GFP strains look similar during exponential growth. Snf5 copy number is ~300 molecules per cell, which is sufficient to produce a faint but clear nuclear GFP signal (Fig. 15). During glucose-starvation at pHEnv 5.5, the nuclear signal of Snf5-GFP vanished (Fig. 15A) and other bright spots were visible in the cell. Originally we thought that these spots represented aggregates of Snf5-GFP. However, an untagged control strain showed similar a signal from auto-fluorescence, most likely arising from metabolic intermediates in the mitochondria. Upon glucose-starvation, respiration is upregulated to optimize energy production, and it is thought that intermediaries from respiration are auto-fluorescent in the GFP channel. Indeed, when we replaced the original eGFP tag with a strong non-dimerization mutant (A206K known as GFPmut3), we no longer saw this phenotype (data not shown).

However, an important reason why we were misled is that the SNF5-GFP and ΔQ-snf5-GFP strains showed distinct phenotypes during starvation. While WT fusion protein vanishes, the ΔQ-snf5-GFP does not and is still clearly visible as nuclear signal even during carbon starvation. The GFP signal does slightly decrease in intensity during glucose-starvation due to acidification of the cytosol. This is due to the pKa of the eGFP fluorophore which leads to protonation and loss of fluorescence below pH 7 (Patterson et al., 1997) (Fig. 15B). Consistent with this idea, starvation in media buffered to pH 7.5 prevented changes in the nuclear signal of both strains. Thus, we hypothesized that a pH dependent aggregation process was quenching the GFP signal for SNF5-GFP but not ΔQsnf5-GFP. Given the limitations of fluorescence microscopy to study an aggregation process like this, we decided to explore an orthogonal way to test the hypothesis. By using the Spider protein Spidroin.

**Spidroin-driven aggregation of SNF5 partially rescues the ΔQ-snf5 phenotype.**

If the SNF5 polyQ domain undergoes an aggregation process during the pHcyl drop upon glucose-starvation, then orthogonal induction of the pH-dependent aggregation process should rescue the phenotype of the ΔQ-snf5 mutant strain. To test this hypothesis, we took advantage of the spidroin protein which is known to oligomerize in response to pH changes. Spidroin, the main component of spider silk, is made and stored in the silk glands of spiders in soluble form at pH ≥7. During the silk production process, the protein
is excreted through the silk duct, which progressively acidifies from around pH 7 at the proximal end to ≤6.3 at the distal end. This pH decrease causes spidroin to convert into a solid silk fiber (Askarieh et al., 2011) (Fig. 17A). Thus, spidroin undergoes a pH-dependent aggregation process over a similar pH range to that which occurs in the S. cerevisiae cytosol during carbon starvation.

We generated a synthetic mini-spidroin allele of 376 amino acid in length. First, we tested if acidification of the cytosol upon glucose-starvation, induces aggregation of spidroin. Exponentially growing cells with a cytosolic pH of 7.4 do not show any spidroin aggregation (Fig. 16). We manipulated the cytosolic pH by poisoning glycolysis and the electron transport chain with 2-deoxyglucose and sodium azide respectively. This causes complete depletion of ATP and consequently causes the cytosol to equilibrate to the pH of the extracellular buffer. All cells buffered to pH 5.5 showed aggregation while, in contrast, no visible aggregation takes place at pH 7.5. These results indicate that the mini-spidroin domain aggregates in the yeast cytoplasm at acidic pH as expected.

We next tested the effect of glucose-starvation on the spidroin peptide. Spidroin cultures grown overnight to an OD_{600} of 0.3, were subjected to acute glucose-starvation in media at pH_{ext} 5.5 or 7.5. At pH_{ext} 5.5, spidroin aggregation was variable: around 40% of the cells had clear aggregates and 60% did not (Fig. 16C). This result probably reflects the variable pH dynamics that occur during carbon starvation (Fig. 6). When cells were subjected to glucose-starvation at pH_{ext} 7.5, only about 5% of cells have aggregates (Fig. 16B). This small amount of aggregation is likely to be due to the slight acidification of the cytosol that occurs even when cells are starved at neutral pH. Thus, cytosolic acidification that occurs during carbon starvation is sufficient to drive spidroin aggregation.

Having confirmed that at least 40% of the culture would induce spidroin aggregation upon glucose-starvation at pH 5.5, we fused the mini-spidroin domain to the N-terminus of the \( \Delta Q\)-\( snf5\) gene, thereby replacing the polyQ of \( SNF5\) with the spidroin peptide, a strain that we term \textit{Spidroin-}\( \Delta Q\)-\( snf5\). This strain allowed us to evaluate the hypothesis that a pH-driven aggregation process was important to enable the induction of the \( P_{ADH2}\)-\textit{mCherry} reporter. Indeed, \textit{Spidroin-}\( \Delta Q\)-\( snf5\) partially reverted the defect observed in \( \Delta Q\)-\( snf5\) (Fig. 17B). We tested if spidroin dependent \( P_{ADH2}\)-\textit{mCherry} expression correlates with environmental pH. As shown in figure 17B, spidroin induction was maximal in media at pH 5.5_{ext} with around 50% of cells inducing, and then induction decreased at pH_{ext} 6 and was undetectable at pH_{ext} 6.5 or above. Therefore, the mini-spidroin domain partially rescues the pH-dependent induction of \( ADH2\) consistent with the hypothesis that pH changes drive a transient aggregation process during transcriptional reprogramming.

From the results presented in figure 6A, we observed that \( SNF5\) also seems to control cytosolic pH: \( snf5\)\( \Delta\) strains failed to acidify their cytosol, while \( \Delta Q\)-\( snf5\) strains didn’t recover neutral pH during starvation, but rather maintained an acidic pH throughout the course of the experiment. The cytosolic pH of the \textit{Spidroin-}\( \Delta Q\)-\( snf5\) strain dropped upon glucose-starvation at pH 5.5. Immediately after glucose removal, most of the \textit{Spidroin-}\( \Delta Q\)-
snf5 population dropped to pHcyt ~6.5 and then, similar to control cells, this population bifurcated to an acidic population at pHcyt ~6 and a population that recovered to pHcyt ~7 (Fig. 17D). pH neutralization and $P_{ADH2}$-mCherry induction was somewhat less bistable than in control cells (compare Fig. 8 and 17), but it is remarkable that an orthogonal pH responsive aggregation domain can partially rescue both gene induction and cytosolic pH regulation.

After 8 h under optimal ADH2 induction conditions, about 50% of the population induces $P_{ADH2}$-mCherry with an average of induction intensity of 65% of WT control levels (Fig. 17 C). As expected spidroin-$\Delta Q$-snf5 did not induce $P_{ADH2}$-mCherry at pHext above 6.5 consistent with a requirement for aggregation of the spidroin domain to enable transcription (Fig. 17A, B). ADH2 induction in spidroin-$\Delta Q$-snf5 is pH-dependent, where after acidification a neutralization of the cytosol precedes $ADH2$ expression (Fig. 17C, D). Spidroin-$\Delta Q$-snf5 partially recapitulates the $ADH2$ expression and the cytosolic pH dynamics of WT.

An increasing number of proteins involved in transcriptional control have been reported to undergo phase transitions. We therefore, wanted to test whether N-terminal fusion of other domains that drive phase separations could also improve the function of the $\Delta Q$-snf5 allele. First we tested the low complexity domains of the FUS and hnRNP, proteins. These domains are known to undergo phase transitions to hydrogels which can recruit RNA PolII to drive aberrant gene induction during oncogenesis following chromosomal translocation events that fuse these domains to DNA-binding proteins (Kato et al., 2012; Kwon et al., 2013). We fused the low complexity domains of FUS or hnRNP to the N-terminus of $\Delta Q$-snf5, generating alleles that we term $FUS$-$\Delta Q$-snf5 and $hnRNP$-$\Delta Q$-snf5. Neither of these domains have been reported to change their state depending on pH. When the strains carrying the $FUS$-$\Delta Q$-snf5 and $hnRNP$-$\Delta Q$-snf5 alleles were subjected to glucose-starvation at optimal conditions for $ADH2$ expression, no induction of $ADH2$ was seen (Data not shown). Neither of these domains have been reported to change their state depending on pH. Therefore, we hypothesized that a dynamic transition between phase separation states might be required for transcriptional activation. This idea is consistent with the pH-dependent behavior of the spidroin-$\Delta Q$-snf5 allele. To further test this idea, we took advantage of the synthetic elastin repeat domain GEPM peptide that dynamically aggregates and disaggregates upon temperature changes (Pastuszka et al., 2013).

GEPM is a synthetic peptide inspired by Elastin-like polypeptides (ELPs) and is comprised of tandem repeats of a VPGXG sequence. This peptide has been shown to undergo temperature-dependent phase transitions with the temperature determined by the number of repeats (Pastuszka et al., 2013). We engineered a GEPM domain that was soluble below 28 C and aggregated above and fused this domain to the N-terminus of $\Delta Q$-snf5 to create a GEPM-$\Delta Q$-snf5 allele. We grew GEPM-$\Delta Q$-snf5 cultures at 25 C, at OD$_{800}$: 0.3, the cultures washed with glucose-starvation media and incubated at 32 C, for 2 hours (to induce a phase separation) and then returned to 25 C. After 4 hours at 25 C total RNA
extractions were performed for qPCR. The WT control strain had a weaker than normal induction of the $P_{ADH2}$-mCherry gene of ~400 fold likely due to the reduced temperature, while $\Delta Q$-snf5 (negative control) as well as GEPM-$\Delta Q$-snf5 did not show any induction of ADH2 expression (Fig. 18).

**SS18, a glutamine rich member of the human SWI/SNF complex, can rescue the polyQ domain of SNF5**

We wanted to see if our studies in yeast might have relevance for mammalian transcriptional regulation and disease. SMACRB1, the human SNF5 homologue, does not contain a polyQ domain. However, from human SWI/SNF architecture assays, it is known that SMACRB1 and SS18 are interact with the same subunits that Snf5 contacts in the yeast complex (Kadoch and Crabtree, 2015). SS18 is an intrinsically disordered protein that contains 79 glutamines and 65 prolines within its 418 amino acids (Middeljans et al., 2012). Thus, we hypothesized that human SS18 might be functionally analogous to the polyQ domain of SNF5. To test this idea, we tested whether SS18 could complement the polyQ domain of SNF5 polyQ. We fused the SS18 coding sequence to the N-terminus of the $\Delta Q$-snf5 gene to create an allele that we term $ss18$-$\Delta Q$-snf5. We then introduced this allele into a strain carrying the $P_{ADH2}$-mCherry reporter. We subjected exponentially growing cultures to glucose-starvation at pH Env 5.5 to 8. We observed that about 45% of the $ss18$-$\Delta Q$-snf5 strain population induced the $P_{ADH2}$-mCherry reporter during glucose-starvation at pH Env 5.5 (Fig.18 left panel) compared to 80% of the control cells (Fig.19 right panel). An unexpected result was obtained at increasing pHs. While the induction of the $P_{ADH2}$-mCherry reporter decreased at more neutral pH Env for control strains and other mutants alleles tested in this work, we observed maximal expression peak for the $ss18$-$\Delta Q$-snf5 strain at pH Env 6.0 where about 60% of the culture induced while in SNF5 is about 78%. This result was consistent in 3 biological replicas. At pH 6.5 reporter induction in $ss18$-$\Delta Q$-snf5 strains dropped to the levels of the negative control and above pH Env 6.5 there is was no induction. Thus, $ss18$-$\Delta Q$-snf5 has a higher pH Env optimum than control cells. This could be because SS18 has evolved to activate transcription in response to a different pH Env range. Nevertheless, it is striking that SS18 partially complements the function of the SNF5 polyQ domain. These results support the idea that similar pH-dependent regulation could occur in human cells. This hypothesis remains to be tested in future work.
**Conclusions**

The $\Delta Q$-*snf5* allele is a dominant negative mutation, therefore equivalent concentrations of the WT and mutant protein in the cell produce the mutant phenotype. The idea that the polyQ domain forms aggregates upon glucose-starvation originated from microscopy observations using fluorescently tagged *SNF5-GFP* and $\Delta Q$-*snf5-GFP* strains. The *SNF5-GFP* signal vanished upon glucose-starvation which could be due quenching of fluorescence as a result of an aggregation. We now know that *SNF5* and the mutant alleles, fused to GFP at the C-terminus produce a null-like phenotype, possibly as a result of aberrant SWI/SNF assembly (Dutta *et al.*, 2017; Sen *et al.*, 2017). Despite the limitations of the microscopy experiments, we were able to test the aggregation hypothesis in an orthogonal way by forcing aggregation of $\Delta Q$-*snf5* upon glucose-starvation using *spidroin*. Spidroin is a spider web peptide that is soluble at pH 7.5 and aggregates at pH 6.3 (Askarieh *et al.*, 2011). We engineered a mini-spidroin domain that undergoes pH-dependent aggregation dynamics in the cytoplasm of *Saccharomyces cerevisiae* in response to carbon starvation. The replacement of the polyQ domain of *SNF5* with this domain ($spidroin-\Delta Q$-*snf5*) partially recovered *ADH2* expression upon glucose-starvation. Our working model is that the aggregation process induced by spidroin helps $\Delta Q$-*snf5* to disengage transcription of SWI/SNF from promoters that are active during exponential growth to allow redistribution to SWI/SNF-dependent glucose-repressed genes upon carbon starvation. The last chapter contains more insights into the possible mechanism for polyQ engaging and disengaging transcription, together with future perspectives.
Figure 10: Deletion of SNF5 polyQ creates a dominant negative allele. A- Histograms of P_{ADH2-mCherry} reporter expression, Y-axis is time in hours, starting immediately prior to glucose-starvation ("Glu"), WT/ΔQ-snf5 indicates a strain carrying both alleles integrated at the endogenous locus. B- Same experiment with strains carrying the extrachromosomal WT allele expressed from a CEN/ARS plasmid and the counterparts after selection against the CEN/ARS plasmid created a version cured of the WT extrachromosomal allele.
Figure 11: C-terminal GFP tagging of SNF5 creates a hypomorphic allele. ADH2 mRNA content of GFP and non GFP tagged WT and ΔQ-snF5. Strains were subjected to glucose-starvation at pH_{Env} 5.5 for 4 h and then RNA was collected. ACT1 mRNA was used as a control, Y axis corresponds to fold induction of ADH2 upon glucose-starvation compared to exponentially growing cells. The SNF5-GFP strain fails to properly induce while the ΔQ-snF5-GFP fails to fully repress expression.
Figure 12: SNF5 as well the mutant alleles are not degraded upon glucose-starvation. Western blots of the SNF5-TAP and TAP tagged mutant alleles. ΔQ-snf5 is 288 amino acids smaller than WT. Anti-glucokinase antibody was used as a loading control (Bottom band at ~50 kDa).
Figure 13: Phenotypes of SNF5 mutants in various stress conditions. A- Growth curves of WT, ΔQ-snf5 and snf5Δ strains. From left to right, WT (red), ΔQ-snf5 (yellow), WT + 1mM H₂O₂ (blue) snf5Δ (green) snf5Δ + 1 mM H₂O₂ (dark blue), ΔQ-snf5 + 1 mM H₂O₂ (grey). Exponentially growing cultures at an OD₆₀₀ of ~0.3 were diluted to a final OD₆₀₀ of 0.01 with and without 1 mM H₂O₂. B- Same strains and experimental procedure as A except that cells were incubated with or without 2 mM DTT. From left to right, WT (red), ΔQ-snf5 (yellow), WT + 2 mM DTT (blue), ΔQ-snf5 + 2 mM DTT (grey) snf5Δ (green) snf5Δ + 2 mM DTT (dark blue). C- Histograms of cytoplasmic pH of WT cells upon addition of 1 mM H₂O₂, “Glu” is pretreatment and time 0 is immediately after inoculation of H₂O₂. D- Cytosolic pH 10 minutes after addition of various concentrations of H₂O₂.
Figure 14: *snf5Δ* strains have an extended lag phase after switching to alkaline pH media. A- A WT strain was inoculated to a final OD600 of 0.05 into rich media at pH 4.2 (red), 5.5 (yellow) and 7.5 (blue). Y axis is OD$_{600}$ and X axis time. B- Same experiment with a *snf5Δ* strain.
Figure 15: Fluorescent microscopy of *SNF5* suggests an aggregation mechanism. A- *SNF5-GFP* strains taken from an exponentially growing culture (+Glu) and after glucose-starvation at environmental pH 5.5 and 7.5. B- *SNF5-GFP* and Δ*Q-snf5-GFP* upon glucose-starvation at pH_{env} 5.5.
Figure 16: A synthetic mini-spidroin protein aggregates upon glucose-starvation at $pH_{Env} 5.5$ but no at $pH_{Env} 7.5$. A- Spidroin-GFP on exponentially growing cells. B- Spidroin-GFP upon glucose-starvation at $pH_{Env} 5.5$ and $pH_{Env} 7.5$. C- Analysis of the data presented in B, Y axis showing the percentage of cells that have aggregates.
Figure 17: A synthetic spidroin domain partially rescues pH-dependent induction of transcription. A- Schematic representation of the aggregation behavior of spidroin. B- Induction of the $P_{ADH2}$-$mCherry$ reporter with varying environmental pH: pH 5.5 (Red), pH 6.0 (Orange), pH 6.5 (Yellow), pH 7.0 (Green), pH 7.5 (Blue) and pH 8.0 (dark Blue). C- Quantification of reporter expression from (B) at various time points after glucose withdrawal. The Y axis is the intensity of fluorescent reporter (normalized to the levels of wild type control cells), shading corresponds to the percentage of cells inducing at each time point. D- Quantification of cytosolic pH from B.
Figure 18: The GEPM peptide that aggregates upon temperature does not recover ADH2 expression. WT, ΔQ-snf5 and GEPM-ΔQ-snf5 were growth at RT (19°C). After glucose withdrawal at OD$_{600}$: 0.3, cells were incubated at 30°C for 2 h and then moved back to RT. Total RNAs were collected after total 6 h of glucose-starvation.
Figure 19: SS18, a glutamine-rich, member of the human SWI/SNF complex can partially complement the SNF5 polyglutamine domain. Cytometry data histograms of $P_{ADH2}$-mCherry induction. Environmental pH is indicated by color codes: pH 5.5 (Red), pH 6.0 (Orange), pH 6.5 (Yellow), pH 7.0 (Green), pH 7.5 (Blue) and pH 8.0 (dark Blue). Each panel corresponds to the indicated strain.
Chapter 5: Future perspectives

Perspectives

Low complexity sequences are highly enriched in nuclear proteins (Albà, Guigó and Guigo, 2004) and there is increasing evidence that many of these intrinsically disordered domains have the capacity to undergo phase transitions (Brangwynne et al., 2009; Kato et al., 2012). However, there aren’t yet many examples of a functional role for phase transitions in transcription. Hydrogels made of FET proteins, required for gene expression, shed some light on the possible role of these sequences (Kwon et al., 2013). However, it remains to be shown that phase transitions play a general role during transcription.

Our working hypothesis is that the SNF5 polyQ domain can undergo phase transitions in a pH dependent manner. To further investigate this idea we have started, through collaboration with Simon Alberti’s group in Dresden, to test the phase-separation behavior of purified SNF5 polyQ in vitro. We found that indeed, in a range from 100 to 300 M concentrations of NaCl, SNF5 polyQ can demix from aqueous solution to form a liquid droplet (Fig. 16). In addition, the protein can collapse into filamentous aggregates (Fig. 20). Therefore, the polyQ domain of SNF5 can exist in soluble, phase separated liquid droplet, and fibrous aggregate forms. We observed that the equilibrium among these states is partially governed by pH (Fig. 16). At pH 7 more liquid droplets are formed than at pH 6 where the protein transforms into soluble or fibrous aggregate forms (Fig. 20).

We hypothesize that SNF5 is required for liquid droplet formation at the transcription site, to properly package or recruit the molecules required for transcription in a sort of “fluid” aggregate. Evidence from the literature and a second piece of our data supports this hypothesis. Increased molecular crowding using osmotic pressure can partially revert the growth defect of a snf5Δ strain. The nucleolus, where transcription and assembly of the ribosomes takes place, is a densely packaged subdomain of the nucleus. It has been documented in the literature, that the SWI/SNF complex is required for RNA PolI transcription (Zhang et al., 2013). Deletion of SNF5 and SNF6 have a big impact on ribosomal production, interestingly deletion of SNF6 (another polyQ containing protein of the SWI/SNF) does not affect RNA PolI occupancy, therefore its processivity is affected by the absence of this protein. Performing growth assays in various carbon sources we found that in SCD supplemented with 1M NaCl WT, snf5Δ and ΔQ-snf5 strains show a clear growth defect (Fig. 21). However, the fitness order is altered: snf5Δ is the fastest doubling, then WT and finally ΔQ-snf5 (Fig. 21). A similar result is obtained with SCD supplemented with 1 M Sorbitol. A possible explanation for this phenomenon is that SNF5 is important to maintain a phase separated structure around rDNA. Crowding is an important driver of phase separation, thus the increase in crowding driven by osmotic shrinkage of cells could help complement a loss of phase separation in the snf5Δ strain (Kato et al., 2012).
The division rate in *Saccharomyces cerevisiae* is limited by the rate of ribosome biogenesis. An approximate calculation taking into account the elongation rate of RNA PolI, the number of rRNA loci and the footprint of the RNA PolI enzyme suggests that polymerase molecules are packed immediately adjacent to one another during exponential growth. If the transcriptional machinery is so highly packaged, the fluidity of the nucleolus will play a key role facilitating RNA PolI elongation. It has been hypothesized that to achieve such level of packaging without impairing RNA PolI function, and perhaps facilitate it, the nucleolus forms a phase separated structure (Denes Hnisz 2017). Indeed, the nucleolus has been demonstrated to be a phase-separated viscous liquid in *Xenopus* (Brangwynne, Mitchison and Hyman, 2011). This hypothesis could be tested using single molecule microscopy to evaluate the crowding and dynamics of the rDNA transcriptional machinery in live cells and compared to growth rates. One inconvenience of using fluorescent microscopy on the transcriptional machinery is that the GFP molecule can interfere with the proper folding of the transcriptional machinery as is the case for C-terminal tagging of SNF5 (Fig. 10). The ideal candidate to test this hypothesis, is RNA PolI. However, careful examination of the fitness on various conditions must be done. If tagging the polymerase affect fitness, a second candidate is SNF2, which is also part of the SWI/SNF complex and is required for RNA PolI activity. C-terminally tagged SNF2 has been used for multiple immunoprecipitation assays of the SWI/SNF complex (Smith *et al.*, 2003; Yang *et al.*, 2007; Dutta *et al.*, 2017). Therefore, is likely not to affect *Saccharomyces cerevisiae* fitness. We have also confirmed that SNF2-GFP strains have normal induction of our model genes during carbon starvation.

Failure to properly transcribe ribosomal DNA, has been recently reported to be the most important trigger for cell death in during aging of *Schizosaccharomyces pombe* (Spivey *et al.*, 2016) and over 40% of human cancers have mutations in the rDNA transcription machinery (Ajore *et al.*, 2017). Studying the mechanisms of nucleolus formation and function is essential to understand the role of rDNA in aging and disease.

The work present in this thesis provides data supporting the hypothesis that phase transitions play a role in transcription. This mechanism involves fluid aggregates of the members of the transcriptional machinery, in order to recruit the molecules required for transcription as well as mechanism to facilitate the elongation of the transcript by the polymerase.
Figure 20: The SNF5 polyQ domain forms liquid droplets and filamentous aggregates in a pH-dependent manner. The polyQ from SNF5 was purified and incubated at various pH values and salt concentrations. At pH 7 SNF5 polyQ forms liquid droplets from 0 to 250 mM NaCl, while at pH 6 it forms droplets over a narrower pH range from 0 to 150 mM.

Figure 21: Osmotic shock stress improves the relative fitness of the snf5Δ strain. Growth curves of WT, ΔQ-snf5 and snf5Δ under osmotic stress. Exponentially growing cultures OD$_{600}$: ~0.3 were inoculated (final OD$_{600}$: 0.01) with and without 1 M NaCl or 1 M Sorbitol.
References


