The Snf1 Protein Kinase in the Yeast Saccharomyces cerevisiae

Usaite, Renata; Olsson, Lisbeth; Nielsen, Jens

Publication date:
2008

Document Version
Publisher's PDF, also known as Version of record

The Snf1 Protein Kinase in the Yeast *Saccharomyces cerevisiae*

Renata Usaite

Ph.D. Thesis
June 2008

Center for Microbial Biotechnology
Department of Systems Biology

To my family

The Snf1 Protein Kinase in the Yeast

Saccharomyces cerevisiae

Ph.D. Thesis

Renata Usaite

Center for Microbial Biotechnology

BioCentrum-DTU

Technical University of Denmark

2008

Preface

The work for this Ph.D. thesis was carried out at the Center for Microbial Biotechnology, at the Technical University of Denmark from June 2004 to June 2007, under the supervision of Prof. Lisbeth Olsson and Prof. Jens Nielsen. External 6-month project was carried out at the Proteomics Mass Spectrometry Lab, at the Scripps Research Institute, La Jolla, California from January to July in 2006, under the supervision of Prof. John R. Yates III. Financially, this Ph.D. work was supported by the Danish Research Agency for Technology and Production. In addition, proteome work was supported by the National Institute of Health. Travels throughout the period of Ph.D. study were supported by FEBS Youth Travel Fund and Otto Mønsteds Fund.

Foremost I would like to thank Prof. Lisbeth Olsson and Prof. Jens Nielsen for enabling me to conduct this Ph.D. study. I thank them for our discussions, their constant encouragement, and the freedom they gave me in designing and proceeding with the project. I am very much grateful to them for the given opportunity to learn about, practice and evaluate the strengths of fermentation physiology and Systems Biology. I thank them for the opportunity to go to many conferences and meetings. Two events: the 1st FEBS Advanced Lecture Course 'Systems Biology: From Molecules & Modelling to Cells', Gosau, Austria, and the RECOMB Satellite Conference on Systems Biology and Computational Proteomics at UCSD, La Jolla, California, were real breakthroughs for me in the field of Systems Biology. Taking part in the EU AMPKIN project was an inspiring experience to pursue on my project. I am grateful to my supervisors and Prof. John R. Yates III for the opportunity to do a 6-month project in a world leading Proteomics Mass Spectrometry laboratory. There I learned about modern proteomics techniques and contributed to the success of this Ph.D. project largely.

I thank many people within scientific, technical and office personnel at CMB and in Yates lab that coached, helped to succeed, became my friends and were in collaboration with me on the different projects. Sorry, for not being able to list all of you here. I especially thank coaches and collaborators Birgitte Regenberg, Lene Christensen, Kiran R. Patil, Vijayendran Raghevendran, Steen L. Westergaard, Thomas Grotkjær, Ana Paula Oliveira, Michael C. Jewett, Jie Zhang, James Wohlschlegel and John Venable, and my Master Students Miaojia Zhou and Jakob R.H. Nielsen.

I thank my whole family for their support along my path to complete this Ph.D. I also thank my close friends Birute, James, Jurga and Osvaldas for their patience and lasting faith through the long years apart. My big special thanks goes to James for his endless support and English language coaching.

Renata Usaite,
January 2008, Kgs. Lyngby, Denmark

List of publications prepared during my Ph.D. study:

This thesis is based on the following publications and manuscripts:

1. **Usaite R**, Nielsen J, Olsson L (2008) Physiological characterization of glucose repression in the strains with *SNF1* and *SNF4* genes deleted. *J Biotechnol*, 133(1):73-81.
2. **Usaite R**, Wohlschlegel J, Venable JD, Park SK, Nielsen J, Olsson L and Yates JR III (2008) Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods. *J Proteome Res*, 7(1):266-75.
3. **Usaite R**, Jewett MC, Oliveira AP, Yates JR III, Olsson L and Nielsen J (submitted) Reconstruction of the yeast Snf1 kinase regulatory network reveals its role as a global energy regulator.

Other collaborations were performed during my Ph.D. study:

4. **Usaite R**, Patil KR, Grotkjaer T, Nielsen J and Regenberg B (2006) Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. *Appl Environ Microbiol*, 72(9):6194-6203.
5. Chapman E, Farr GW, **Usaite R**, Furtak K, Fenton WA, Chaudhuri TK, Hondorp ER, Matthews RG, Wolf SG, Yates JR III, Pypaert M and Horwich AL (2006) Global aggregation of newly translated proteins in an Escherichia coli strain deficient of the chaperonin GroEL. *Proc Natl Acad Sci U S A*, 103(43):15800-15805.
6. Fazio A, Jewett MC, Daran-Lapujade P, Mustacchi R, **Usaite R**, Pronk JT, Workman CT, Nielsen J. (2008) Transcription factor control of growth rate dependent genes in *Saccharomyces cerevisiae*: a three factor design. *BMC Genomics,* 18(9):341.
7. **Usaite R**, Ochmann D, Grotkjær T, Boles E, Regenberg B (submitted) Invasive and pseudohyphal growth evolve under N-limitation via *DIP5*, *GNP1* and the cAMP-PKA pathway
8. Olivares R, **Usaite R**, Patil KR, Nielsen J (in preparation) Translation efficiency in yeast *S. cerevisiae*
9. Development of MicroArray Database (MAD) database (http://www.fbd.dtu.dk/fbd/)
10. Participating in EU AMPKIN project (http://www.gmm.gu.se/AMPKIN/)

Contents

Summary

In yeast, *Saccharomyces cerevisiae*, the Snf1 protein kinase is primarily known as a key component of the glucose repression regulatory cascade. The Snf1 kinase is highly conserved among eukaryotes and its mammalian homolog AMPK is responsible for energy homeostasis in cells, organs and whole bodies. Failure in the AMPK regulatory cascade leads to metabolic disorders, such as obesity or type 2 diabetes. The knowledge about the Snf1 protein kinase remains to be of much interest in studying yeast carbon metabolism and human biology.

To investigate the effect of Snf1 kinase and its regulatory subunit Snf4 on the regulation of glucose and galactose metabolism, I physiologically characterized *Δsnf1*, *Δsnf4,* and *Δsnf1Δsnf4* CEN.PK background yeast strains in glucose and glucose-galactose mixture batch cultivations (chapter 2). The results of this study showed that delayed induction of galactose catabolism was *SNF1* or *SNF4* gene deletion specific. In comparison to the reference strain, growth delay on galactose was found to last 2.4 times (7 hours) longer for the *Δsnf4*, 3.1 times (10.5 hours) longer for the *Δsnf1*, and 9.6 times (43 hours) longer for the *Δsnf1Δsnf4* strains. The maximum specific growth rates on galactose were found to be two to three times lower for the recombinant strains compared to the reference strain (0.13 h^{-1}) and were found to be 0.07 h^{-1} for the *Δsnf1*, 0.08 h^{-1} for the *Δsnf4* and 0.04 h^{-1} for the *Δsnf1Δsnf4* strain. In contrast to what is generally believed, the study showed that the Snf1 kinase was not solely responsible for the derepression of galactose metabolism.

To investigate the regulatory role of Snf1 kinase on a global scale, the global scale mRNA, large-scale yeast quantitative proteome and metabolome datasets were generated. One of the largest yeast global quantitative proteome datasets (2388 proteins) to date was generated using Multidimensional Protein Identification Technology followed by quantitation using stable isotope labeling approach (chapter 3). The stable isotope labeling was compared to the spectral counting quantitative approach and the study showed that the stable isotope labeling approach is highly reproducible among biological replicates when complex protein mixtures containing small expression changes were analyzed. Where poor correlation between stable isotope labeling and spectral counting was found, the major reason behind the discrepancy was the lack of reproducible sampling for proteins with low spectral counts.

To reconstruct a regulatory map of the yeast Snf1 protein kinase, I used the abundances of 5716 mRNAs, 2388 proteins, and 44 metabolites measured for the wild-type, *Δsnf1*, *Δsnf4,* and *Δsnf1Δsnf4* strains. By integrating these measurements with global protein-protein-interactions, protein-DNA-interactions and a genome-scale metabolic model, I mapped the complete network of interactions around the protein kinase Snf1 (chapters 4, 5). Through these interactions, I identified how the Snf1 protein kinase regulated cellular metabolism on gene or protein level. The study revealed that the Snf1 protein kinase played a far more extensive role in controlling both carbon and energy metabolism than previously anticipated. Similar to the function of AMPK in humans, my findings showed that Snf1 was

a low energy checkpoint. Our results suggested that it was possible to use yeast more extensively as a model system for studying the molecular mechanisms underlying the global regulation of AMPK in mammals.

Dansk sammenfatning

I gær, *Saccharomyces cerevisiae,* er Snf1 protein kinase primært kendt som en nøglekomponent i regulerring der er ansvarlig for glukoserepressionen. Proteinet Snf1 kinase er konserveret mellem eukaryoter og analogen i mammalieceller er AMPK, som er ansvarlig for at kontrollere energistofskiftet. Hvis de reguleringer som som AMPK styrer ikke fungerer leder det til udvikling af metabolske sygdomme, såsom fedme eller type 2 diabetes. Det er derfor af stort interesse at studere Snf1 protein kinase nærmere, da den har stor vægt i gærs kulstofmetabolisme og i humanbiologi.

For at undersøge effekten af Snf1 kinase og dens regulatoriske komponent Snf4 på reguleringen af glukose og galaktosemetabolismen i gær, blev der foretaget fysiologiske studier af *Δsn1, Δsnf4* og *Δsnf1Δsnf4* deletionsstammer som blev sammenlignet med en referencestammen. Stammerne blev dyrket i batchkultiveringer på glukose eller på en blanding af glukose og galaktose (kapitel 2). Denne undersøgelse viste at der er en forsinket induktion af galaktosemetabolismen som var *SNF1* og *SNF4* specifik. I sammenligning med referencestammen, var lagfasen mellem glukoseomsætningen og galaktoseomsætningen, henholdsvis 2.4 gange længere (7 timer) for *Δsnf4* stammen, 3.1 gange længere (10.5 timer) for *Δsnf1* stammen og 9.6 gange længere (43 timer) for *Δsnf1Δsnf4* stammen. Den maximale specifikke væksthastighed på galaktose var to til tre gange lavere for de rekombinante stammer, sammenlignet med referencestammen, som havde en maksimal specifik væksthastighed på 0.13 h^{-1}, hvorimod denne hastighed blev bestemt til 0.07 h^{-1} for *Δsnf1* stammen, 0.08 h^{-1} for *Δsnf4* og 0.04 h^{-1} for *Δsnf1Δsnf4* stammen. Det kunne konkluderes at snf1 kinase ikke alene er ansvarlig for derepressionen af galaktosemetabolismen.

For at nærmere undersøge den regulatoriske rolle af snf1 kinase på en global niveau blev der udført en kvantitativ proteome analyse. En af de mest omfattende kvantitative proteomdatasæt blev indsamlet (omfattede 2388 proteiner) og genereret ved brug af multidimensionel proteinidentifikationsteknologi med efterfølgende kvantificering ved hjælp af en isotopmærkningsmetode. Brugen af isotopmærkning for kvantificering blev sammenlignet med den spektrale optælling for kvantificering og studiet viste at isotopmærkningsmetoden var meget reproducerbar mellem de biologiske replika når komplekse proteinblandinger med lav variation i udtryksniveau blev sammenlignet (kapitel 3). Når de to kvantificeringsmetoder blev sammenlignet, kunne man observere en dårlig sammenhæng mellem de to metoder, hvilket primært skyldes begrænsninger i den spektrale metode ved lave spektrale optællinger.

For at rekonstruere det regulatoriske netværk der omfatter Snf1 protein kinase i gær blev 5716 mRNA-niveauer, 2388 proteinniveauer og 44 metabolitniveauer målt. Ved at integrere disse målinger med globale protein-protein interaktioner, protein-DNA interaktioner og en omfattende metabolsk model, var det muligt at kortlægge det komplette netværk af interaktioner omkring Snf1 protein kinase (kapitel 4 og 5). Gennem definitionen af disse interaktioner, var det muligt at identificere hvordan snf1 kinase regulerer den cellulære metabolisme på gen eller proteinniveau. Studiet afslørede at Snf1 protein

kinase spiller en meget mere omfattende rolle i kontrol af både kulstof og energimetabolismen end man tidligere troede. Parallelt med hvad man kender til om AMPKs humane funktion, så viser resultaterne at Snf1 er et kontrolpunkt ved lavt energiniveau. Disse resultater peger på muligheden at mere omfattende bruge gær som modelsystem for at studere de molekylære mekanismer som er ansvarlige for den globale regulering af AMPK i mammalieceller.

Chapter 1: Introduction

1.1 Objective

1.2 Yeast *Saccharomyces cerevisiae*

1.3 Yeast Snf1 protein kinase

1.4 Top-down Systems Biology approach

1.1 Objective of the thesis

I dedicated my Ph.D. study to the Snf1 protein kinase in yeast *Saccharomyces cerevisiae*. I focused on expanding the knowledge about Snf1 kinase's role regulating yeast metabolism. The Snf1 regulation of carbon metabolism results generated in my thesis may be used in metabolic engineering and yeast industrial applications. The Snf1 regulation of yeast metabolism results generated in my thesis may contribute to the reconstruction of the global yeast regulatory network, and through yeast-mammalian homology study, it may improve our understanding of human metabolism.

First, I asked if, by deleting kinase encoding gene *SNF1*, the kinase's regulatory subunit encoding gene *SNF4*, or both of these genes, we could expand our knowledge about the Snf1 kinase's and its regulatory subunit's role regulating carbon metabolism. I aimed to do so by using industrially relevant dynamic glucose and glucose-galactose-mixture batch cultivations. This experimental set-up assisted me in investigating how the Snf1 kinase and its regulatory subunit Snf4 functioned within the glucose repression regulatory cascade and how the Snf1 kinase interlocked glucose repression, glucose induction and galactose induction regulatory cascades.

Second, I aimed to investigate the yeast Snf1 kinase's role on a global scale. To do so, I focused on the reconstruction of the Snf1 kinase global regulatory network. Using top-down Systems Biology approach, I integrated large-scale mRNA, protein and metabolite expression data that I generated in my thesis with publicly available protein-protein-interaction, protein-DNA-interaction and protein-metabolite interaction data. This approach allowed me to identify metabolic, signaling or regulatory proteins, transcription factors and metabolic hot-spots through which the Snf1 protein kinase regulated yeast metabolism *per se*. Through systems approach and yeast-mammalian cell homology study I also aimed to prove that yeast Snf1 kinase, as well as its homologous mammalian AMPK, is a low energy checkpoint.

I believe that the data generated and yeast Snf1 regulatory network reconstructed (in my thesis) may be further transferred to mammalian systems to expand knowledge of the AMPK signaling cascade, which is linked to various metabolic disorders. I will be proud if results generated in my Ph.D. study are used to ultimately understand human biology and address metabolic disease.

1.2 Yeast

Yeasts are eukaryotic micro organisms classified in the kingdom of Fungi, making up more than 1000 species.[1] Among these species there are yeasts that are used for brewing, baking or as food additives, and yeasts which are human pathogens or which may spoil foods. Yeasts have been used in brewing and baking processes for thousands of years. Only in the last 150 years, since the experiments of Louis Pasteur, scientists have begun to explore how yeast works.

The most well known yeast is *Saccharomyces cerevisiae*[a]. *S. cerevisiae* is a unicellular, 5-10 μm in diameter budding yeast that is believed to be originally isolated from the skins of grapes.[2] *S. cerevisiae* has a respiro-fermentative metabolism, i.e. in the presence of oxygen and at high glucose concentration, *S. cerevisiae* produces ethanol in addition to biomass. Biomass production is the usually the dominating metabolic process occurring aerobically in most other yeast.[3, 4] The oxidative metabolism is repressed in the presence of high glucose in *S. cerevisiae* and this makes this yeast produce ethanol (a Crabtree effect).[5] In addition, *S. cerevisiae* has a glucose repression effect, in which yeast first grow on preferable carbon source - glucose, then, after glucose is exhausted, yeast may consume less preferable and non-fermentative carbon sources. The consequences of the Crabtree and glucose repression effects are of great importance for industry as manipulation of growth conditions can either increase fermentation byproducts of interest or increase biomass (for example for protein production). *S. cerevisiae* is the micro organism behind the most common types of fermentation. It is also the most intensively studied eukaryotic model organism in molecular and cell biology.

1.2.1 *Saccharomyces cerevisiae* in industrial applications

For centuries *S. cerevisiae* have been used in the production of food and beverages, and today this organism is also used in a number of different processes within the pharmaceutical industry. *S. cerevisiae* is a very attractive organism to work with since it is nonpathogenic, and due to its long history of application in the production of consumable products such as ethanol and baker's yeast, it is classified as a GRAS organism (Generally Regarded As Safe). Also, the well-established fermentation and process technology for large-scale production with *S. cerevisiae* make this organism attractive for several biotechnological purposes. Another important reason for the applicability of *S. cerevisiae* within the field of biotechnology is its susceptibility to genetic modifications by recombinant DNA technology.

[a] *S. cerevisiae* is also called 'yeast' in further text of my thesis.

Traditionally, strain improvement of baker's and brewer's yeast has relied on random mutagenesis and classical genetic breeding approaches. Progress in the field of recombinant DNA technology has enabled us to manipulate a pathway of interest (e.g. by modifying promoter strength, deleting genes, or introducing new genes or whole pathways) and, therefore, to directly improve desired cell characteristics. This directed improvement of the cellular properties achieved from the interplay of theoretical analysis, relying on biochemical information, and the application of genetic engineering is referred to as Metabolic Engineering.[6] Through Metabolic Engineering, we can improve industrial processes by improving productivity and yield of a compound of interest, by eliminating by-products, by improving process performance, by improving cellular properties, or by extending substrate or product range.[7] For example, through Metabolic Engineering and recombination approaches, *S. cerevisiae* has successfully been used for lactic acid and xylitol production, which are attractive compounds in the chemical and food industries.[8, 9] The first recombinant protein, interferon, and the first genetically engineered vaccine, hepatitis B surface antigen, have been produced by *S.cerevisiae*.[10, 11] In recent times, through new genomic-driven solutions, *S. cerevisiae* is used to address global energy and environmental challenges, e.g. it is used to produce biofuels.[12]

1.2.2 *Saccharomyces cerevisiae* as a model organism to study human metabolism

Yeast, as well as e.g. fly and worm, is a model organism that provides insights in many aspects of cell biology, biochemistry and physiology,[13] and provides grounds for developing new technologies, analysis methods and experience. Yeast is a touchstone model organism in post-genomic research and there is no other eukaryote that has been as manipulated, both genetically and physiologically, as *S. cerevisiae*.[14, 15] The genomic sequences of humans and model organism demonstrate a basic unity in the strategy of life. For example, 30% of human genes involved in the development of diseases have functional homologues in yeast and those can be placed in the context of the informational pathways.[16] Biological processes such as misfolding or aggregation of proteins that are implicated in the process of neurodegeneration (e.g. Alzheimer's), have been recapitulated in yeast.[13] Easy to handle, fast-growing, well-studied *S. cerevisiae* remains to be an attractive and broadly used model organism to study human biology, and ultimately to understand and address human disease and aging.

1.3 Snf1 protein kinase and its regulatory subunit Snf4 in *S. cerevisiae*

Originally, the Snf (Sucrose Non-Fermenting) proteins has been identified in screens for mutants with growth defects on sucrose and raffinose.[17, 18] In 1986, the yeast Snf1[b] protein kinase has been determined to play a central role in carbon catabolite repression (also called glucose repression) in the yeast *S. cerevisiae* (Figure 1).[19] It has been determined that the Snf1 protein kinase is required for expression of glucose repressible genes and also suggested that the Snf4 may regulate Snf1 activity.[19] For the last 20 years, the Snf1 protein kinase has been an object of interest in studying carbon metabolism and fermentation. Understanding of the Snf1 kinase role in the regulation of carbon metabolism is crucial for metabolic engineering and design of industrial fermentation processes.

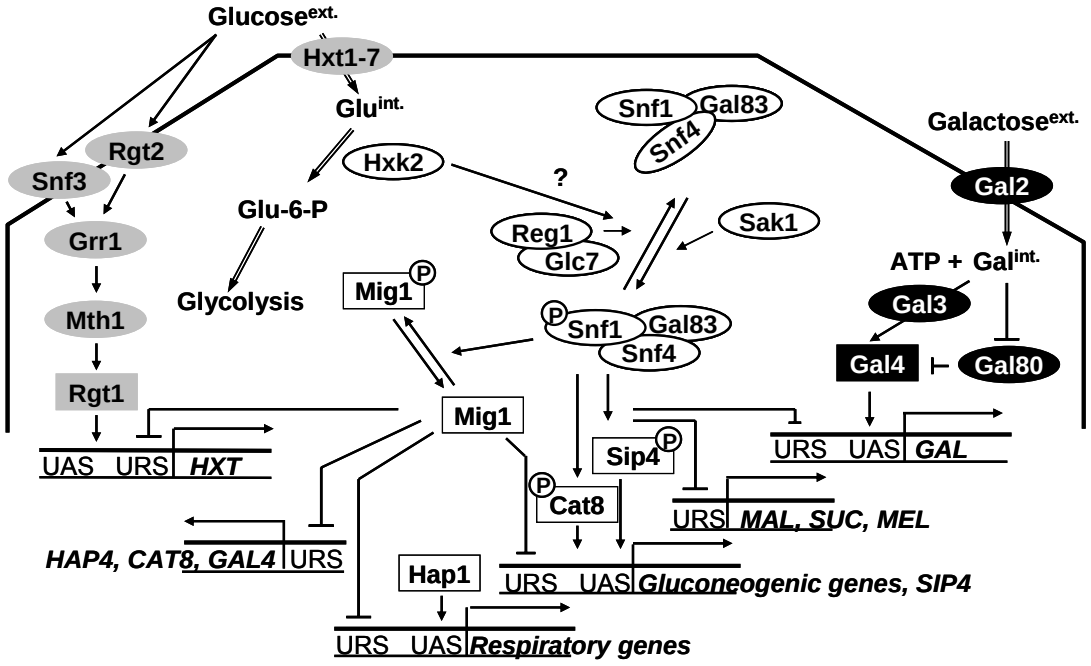


Figure 1. Glucose repression regulatory cascade and its link to glucose induction and galactose induction regulatory cascades. White nodes mark members of the glucose repression regulatory cascade. Grey nodes distinguish members of glucose induction regulatory cascade. Black nodes indicate members of galactose induction regulatory cascade. Transcription factors are marked in squares. Double lines indicate transport and metabolic reactions. The letter-P-in-a-circle symbols mark phosphorylation. For more detailed information refer to the chapter 'The role of Snf1 protein kinase'.

[b] In this thesis, yeast genes and proteins are named according to the nomenclature of SGDatabase (http://www.yeastgenome.org/). Genes are marked using capital italic characters (e.g. *SNF1*). Yeast proteins are marked using the first letter capitalized (e.g. Snf1). Mammalian proteins are marked using all capital letters (e.g. AMPK).

Homology studies indicate that the yeast Snf1 protein kinase is a part of an evolutionary conserved AMPK (AMP-activated Kinase) regulatory cascade. Recent reviews imply that AMP-activated kinase regulatory cascade regulates energy homeostasis in all eukaryotic cells (Figure 2).[20, 21] All energy-consuming reactions (e.g. protein synthesis, cell growth) are powered by the high concentration of ATP. When cellular stress response to certain factors causes a drop in the intracellular energy, the activated AMPK acts as a 'master switch' inducing ATP-generating catabolic pathways and repressing ATP-consuming processes. Failure in mammalian AMPK regulatory cascade may cause metabolic disorders such as dementia, type 2 diabetes and obesity. Today, mammalian AMPK is a major target for two existing classes of drugs used to treat type 2 diabetes. AMPK activators also have potential as anticancer drugs.[22] The knowledge about the yeast Snf1 protein kinase regulatory cascade is successfully transferred to mammalian systems and the Snf1 kinase serves a good model system to study human AMPK.

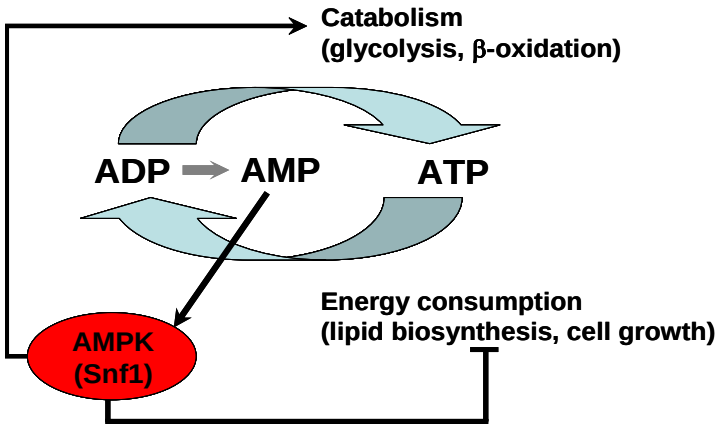


Figure 2. The yeast Snf1 protein kinase, as well as mammalian AMPK, is a low energy checkpoint. Grey and blue arrows symbolize energy homeostasis in a cell. When activated in the low energy growth conditions, the AMP-activated (black arrow) kinases induce (black arrow) energy generating (catabolic) processes, such as glycolysis and β-oxidation, and repress (black 'T-shaped' arrow) energy consuming (anabolic) processes, such as lipid biosynthesis and cell growth. The Snf1 protein kinase's involvement in the regulation of catabolic and anabolic processes is discussed in later chapters in more detail.

This chapter will briefly cover the structural organization, regulators and targets of the Snf1 protein kinase complex and demonstrate the homology between yeast Snf1 protein kinase and mammalian AMP-activated kinase. Prior knowledge about Snf1 protein kinase will be summarized including contributions from my Ph.D. study towards a complete understanding of the global role of the yeast Snf1 protein kinase.

1.3.1 The structure and localization of the Snf1 protein kinase complex

The yeast Snf1 protein kinase functions in a heterotrimeric protein complex that consists of α, β and γ subunits (Figure 3).[23] The *S. cerevisiae* yeast genome codes for one α-subunit Snf1, one γ-subunit Snf4, and three β-subunits Sip1, Sip2 and Gal83. Consequently, the Snf1 protein kinase functions in three complexes Snf1-Snf4-Sip1, Snf1-Snf4-Sip2 and Snf1-Snf4-Gal83. The α–subunit is a serine-threonine protein kinase, which has a catalytic kinase domain that contains a conserved threonine residue (Thr210), and a regulatory domain that contains an AID (Auto-Inhibitory Domain). Based on the kinase domain phylogeny, the catalytic α-subunit is most closely related to the CaMK protein kinases.[24] The γ-subunit regulates the activity of the α-subunit. Through the γ-subunit's interaction with the α-subunit's regulatory domain, the α-subunit's AID is inactivated, its conformation changes, and its kinase catalytic domain is activated.[25, 26] The γ-subunit is composed of four consecutive CBS (Cystathionine β-Synthase) domains that form binding sites for AMP.[27] To date, the evidence shows that the yeast γ-subunit does not bind AMP.[28] The β-subunits act as a scaffold and play a central role in Snf1 protein kinase heterotrimer formation, localization and function. Each of the β-subunits interact with α- and γ-subunits through conserved motifs KIS (Kinase-Interacting Sequence) and ASC (Association with Snf1 Complex), respectively.[29-32] Various N-terminal sequences of β-subunits define subcellular localization and substrate specificity for the Snf1 kinase complex in yeast.[33-35]

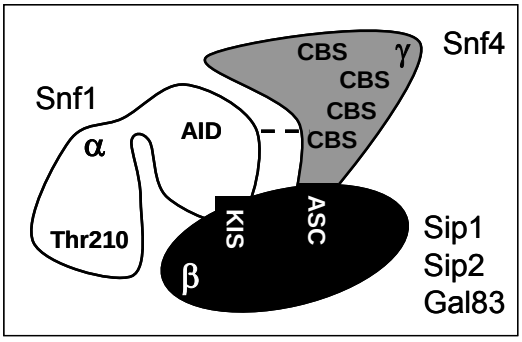


Figure 3. The structure of the yeast Snf1 protein kinase complex. The figure shows interaction between three Snf1 kinase complex subunits α, β and γ, conserved domains and conserved threonine 210 in the catalytic Snf1 kinase domain. The dashed line indicates possible interaction between α- and γ-subunits. For more detailed information, please, see chapter 'The structure and localization of the Snf1 protein kinase complex'.

Based on green fluorescent protein fusion studies[c], the three Snf1 protein kinase complexes (Snf1-Snf4-Sip1, Snf1-Snf4-Sip2 and Snf1-Snf4-Gal83) have been found to localize in the cytoplasm when cells are grown on glucose. When glucose is exhausted, or cells grow in a non-fermentable carbon source, the β-subunits target Snf1 protein kinase complexes to different organelles. The Gal83 targets Snf1-Snf4-Gal83 to the nucleus, the Sip1 targets Snf1-Snf4-Sip1 to the vacuolar membrane, and the Snf1-Snf4-Sip2 complex remains in the cytoplasm (Figure 4).[33, 35] The dynamic relocalization of the three Snf1 protein kinase complexes is likely linked to different Snf1 kinase functions, which will briefly be described in later chapters.

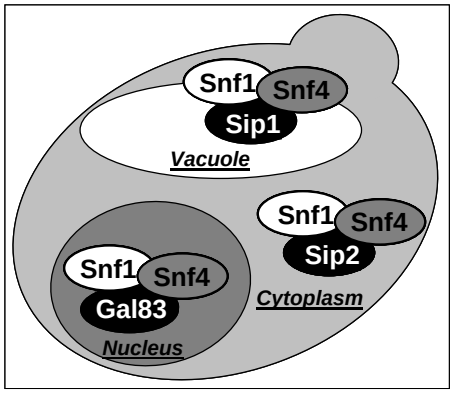


Figure 4. **The localization of the three Snf1 protein kinase complexes under alternative carbon sources or low glucose.** Under glucose growth conditions, the three Snf1 protein kinase complexes are located in the cytoplasm. When glucose is exhausted, the Snf1-Snf4-Gal83 complex is targeted to the nucleus, the Snf1-Snf4-Sip1 complex is targeted to the vacuole, or to the vacuolar membrane and Snf1-Snf4-Sip2 complex remains in the cytoplasm and may be attached to the plasma membrane.

A high degree of homology is identified between the yeast Snf1 protein kinase and its homolog mammalian AMPK.[36] The conservation of these protein kinases extends to the non-catalytic β and γ-subunits of the heterotrimeric complexes (Table 1). In mammals the two α-subunits, two β-subunits and three γ-subunits form 12 heterotrimeric αβγ complexes that are distributed and active in different tissues and organs.[37-39] Homologous to the yeast Snf1 protein kinase, the catalytic domain of the mammalian α–subunit contains conserved threonine (Thr172),[40] phosphorylation of which activates AMPK. The regulatory γ–subunit contains four CBS sequences that form two AMP-binding sites, and the β–subunits act as scaffolding. Differently from the yeast cells, the AMP allosteric binding to the

[c] The green fluorescent protein (GFP) is originally isolated from the jellyfish *Aequorea victoria* and fluoresces green when exposed to blue light. The GFP may be incorporated into the genome of yeast and maintained there through breeding. The recombinant yeast strain is able to express its own protein that is fused to GFP. The illuminated GFP is almost not harmful to living cells and the expression or localization of the recombinant GFP fused protein may be studied in highly automated live cell fluorescence microscopy systems.

γ–subunit is necessary for a stable interaction between α and γ–subunits, which is required for an activation of AMPK.[41]

Table 1. Comparison of the AMP-activated kinases in yeast and mammalian cells.

		Snf1 (yeast)	AMPK (mammalian)
Complex structure	α-subunit	Snf1	AMPK α1, α2
	β-subunit	Sip1, Sip2, Gal83	AMPK β1, β2
	γ-subunit	Snf4	AMPK γ1, γ2, γ3
Physical Regulators	AMP	Indirect	Direct
	Upstream kinases	Tos3, Elm1, Sak1	LKB1, TAK1, CaMKK2
	Phosphoatases	Reg1/Glc7	PP2Cα
	Other interacting proteins	Std1, Hxk2	
Downstream targets	Transcription factors	Mig1, Cat8, Sip4	HNF4α, ChREBP
	Metabolic enzymes	Acc1	ACC, HMGR
	Regulatory proteins	Tor1, Pho85, Gcn5, Histone H3	Heat shock protein, mTOR

1.3.2 The regulation of Snf1 protein kinase

Several proteins and regulatory events participate in the Snf1 kinase activation-deactivation process (Figure 5, Table 1). The Snf1 kinase activity is linked to glycolytic pathway activity.[42, 43] In the presence of glucose (i.e. high glycolytic rate) activated hexokinases (e.g. Hxk2) repress Snf1 protein kinase activity.[42, 43] The exact repression mechanism is not yet understood, but Snf1-Hxk2 constitutive interaction has been identified.[44] In addition, the phosphatase Glc7-Reg1 dephosphorylates and inhibits Snf1 kinase activity. [45, 46] Under low glucose growth conditions (i.e. when intracellular AMP:ATP ratio increases) the Snf1 protein kinase is active. The presence of three α, β and γ-subunits in one complex, the γ-subunit interaction with α-subunit, and the phosphorylation of Thr210 in the catalytic domain are required for Snf1 kinase to be fully active.[25, 26, 32] Three redundant upstream kinases Tos3, Sak1 and Elm1 can phosphorylate the Snf1 kinase's Thr210. The Snf1 kinase phosphorylates Reg1 and releases Glc7-Reg1 phosphatase from the Snf1 kinase complex.[46] The allosteric AMP binding to Snf4 does not activate Snf1 kinase directly, but indirect Snf1 regulation by AMP cannot be excluded.[47]

Research is ongoing to understand which exact signals activate upstream kinases and the phosphatase, and what the key factor regulating Snf1 kinase activity is. It is in debate whether the phosphorylation by the upstream kinases, or dephosphorylation by the phosphatase plays the dominant role regulating Snf1 protein kinase activity.

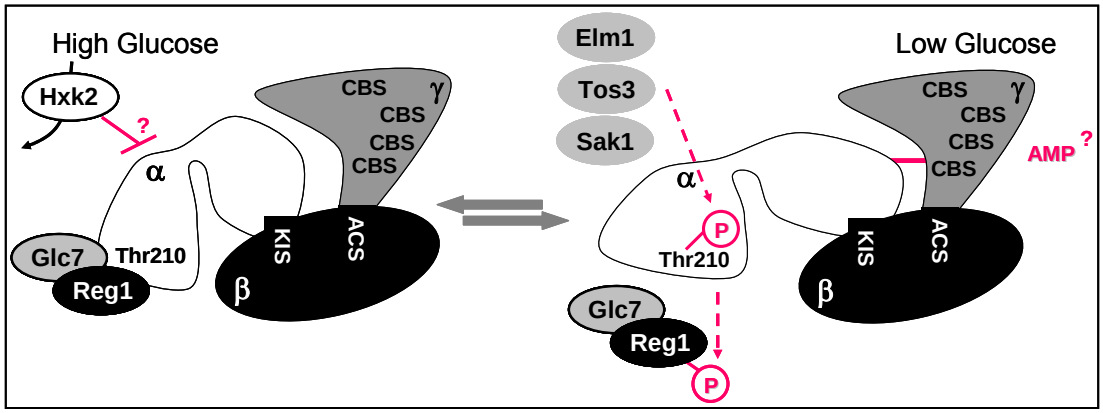


Figure 5. The activation and deactivation of the Snf1 protein kinase complex. At growth conditions with high glucose, the Snf1 protein kinase activity is repressed. At growth conditions with low glucose, the Snf1 protein kinase activity is induced. The Hxk2 hexokinase represses Snf1 protein kinase activity with an unknown mechanism. The upstream kinases (Elm1, Tos3, Sak1) phosphorylate and activate Snf1 kinase. The interaction between α–subunit and γ–subunit (pink line) activates Snf1 kinase. The binding of AMP to γ–subunit may activate the Snf1 kinase complex. The Glc7-Reg1 phosphatase interacts with and inhibits Snf1 kinase at growth conditions with high glucose concentration. Reg1 is phosphorylated by the Snf1 protein kinase and Glc7-Reg1 is released from the active Snf1 kinase complex at growth conditions with low glucose concentration.

The Snf1 protein kinase also responds to nitrogen limitation, salt, ion, alkaline pH, heat shock, and oxidative stresses.[48, 49] These distinct stresses regulate Snf1 kinase's activity and its subcellular localization, dependent or independent of the upstream (Tos3, Sak1 or Elm1) kinase.[49] In addition, other Snf1 regulating proteins, such as Std1,[50] may affect specific Snf1 kinase functions. Responding to specific stresses, the Snf1 protein kinase interacts with specific downstream targets[25] and regulates distinct biological processes.

The conservation among species is not limited to the AMPK complex, but extends through the regulatory cascade.[51, 52] The mammalian tumor suppressor LKB1 and two Ca^{2+}/calmodulin-dependent kinase kinases (CaMKKs), all of which are homologous to the yeast Tos3, Elm1 and Sak1, are upstream kinases of AMPK (Table 1).[53-56] The yeast Snf1 kinase and mammalian AMPK also share homologous downstream targets (e.g. Acc1). The AMP, glucose and oxidative stress regulates activity of the yeast Snf1 and mammalian AMPK.[20] In addition, the AMPK is activated by other physiological and pathological stresses that generally deplete cellular ATP, i.e. change the AMP:ATP ratio. These stresses include hypoxia,[57] ischemia, uncoupled oxidative phosphorylation, exercise and muscle contraction.[58] Other stresses such as hyperosmotic stress,[59] Ca^{2+},[55] and hormones (e.g. leptin and interleukin-6) also regulate AMPK complexes.[60] Recent scientific evidence shows that the upstream kinase LKB1 is constitutively active, and the AMPK protein phosphatase 2Cα (PP2Cα) is regulated by AMP levels indicating that the phosphatase may play a dominant role regulating AMPK activity.[61]

1.3.3 Snf1 kinase interacts with hundreds of proteins

The yeast Snf1 protein kinase interacts[d] with more than 200 proteins (based on the information posted on protein-protein interaction database BioGRID[62] in October, 2007). A single phosphorylome study[e] showed that Snf1 phosphorylates 80 yeast proteins,[63] indicating that all these proteins might be regulated by the Snf1 protein kinase. In the list of proteins, interacting with or being phosphorylated by Snf1, there are transcription factors (TFs) (e.g. Mig1, Sip4, Cat8), signal transducing proteins (e.g. Pho85, Gcn5), enzymes and structural proteins (e.g. Acc1, histone H3). The hundreds of proteins interacting with Snf1 kinase are involved in biological processes, such as organelle organization and biogenesis, transcription, transport, stress response, and metabolism of carbon, lipid, nucleic and amino acids. The abundance and functional variety of identified Snf1-interacting proteins also pinpoints that the yeast Snf1 protein kinase regulates different metabolic processes in the cell.

1.3.4 The role of Snf1 protein kinase

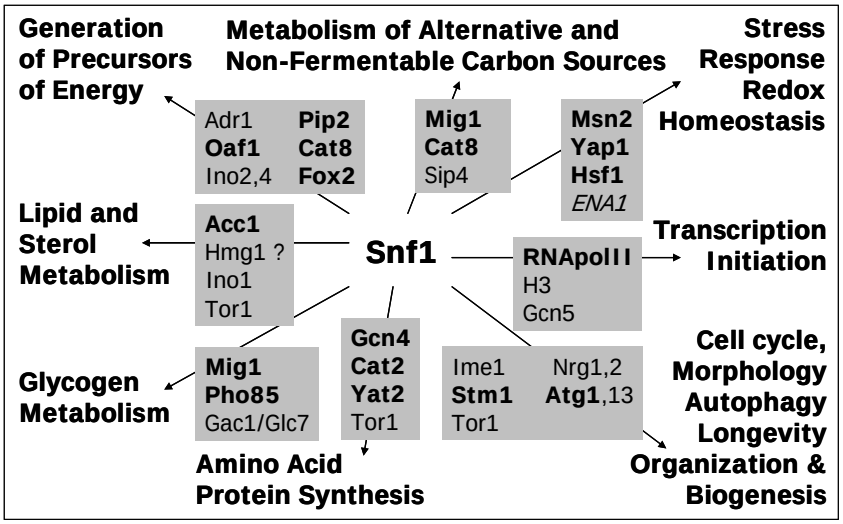


Figure 6. The targets and biological processes regulated by the Snf1 protein kinase. The figure summarizes prior knowledge and knowledge obtained during my thesis. Transcription factors, signal transducing proteins, regulatory proteins and enzymes (all marked in grey boxes) are directly or indirectly affected by the Snf1 kinase. The Snf1 targets marked in bold were identified in my thesis work and were described in chapter 4. More detailed information about most of these Snf1 kinase targets and processes regulated by the Snf1 kinase can be found in the chapter 'The role of Snf1 protein kinase in the cell'.

---

[d] The list of protein-protein interactions, which are posted on the BioGRID (http://www.thebiogrid.org) database, are derived using variety of techniques, highlighting yeast two-hybrid system, mass-spectrometry based approaches, genetic interactions, co-expression, synthetic growth defect and phenotypic enhancement.

[e] Phosphorylome study identifies global protein phosphorylation in yeast. In this particular study, the proteome chip technology was used and a subset of 80 proteins, which were phosphorylated by the protein kinase Snf1, was identified.

The yeast Snf1 protein kinase has primarily been described as a regulator of carbon metabolism and as a key node in the glucose repression regulatory cascade.[64, 65] The Snf1 kinase is also a key regulator of carbon and energy metabolism during diauxic shift.[66] It is likely that on a more global scale, Snf1 contributes to the regulation of transcription initiation.[67-69] The Snf1 kinase has also been shown to regulate lipid[70, 71] and glycogen metabolism,[72] cell cycle,[73] morphogenesis,[74, 75] redox homeostasis[49, 76] and aging.[34, 77] Altogether, Snf1 appears to be a global regulator (Figure 6) of metabolism and the homology studies with mammalian AMPK suggest that Snf1 is a low energy checkpoint (Figure 2).[20, 21, 78]

1.3.4.1 Glucose repression regulatory cascade

In response to a glucose signal, glucose repression regulatory cascade represses or activates metabolism of alternative carbon sources (Figure 1).[64, 65] When glucose is present, Snf1 kinase is not active (Figure 5) and the expression of the genes, that are regulated by glucose repression regulatory cascade, is repressed. When the glucose is exhausted, or in the presence of alternative carbon sources, the Snf1 kinase is active and the expression of these genes is derepressed or activated (Figure 1). [65, 79, 80] The upstream kinases may phosphorylate and thus activate Snf1 kinase. Specifically, the Sak1 upstream kinase has been identified to promote Snf1-Snf4-Gal83 nuclear localization under low glucose.[81] The Snf1-Snf4-Gal83 presence in the nucleus[33] correlates with Gal83's role to mediate Snf1 interaction with TF Sip4[82] and with other TFs of glucose repression regulatory cascade. In the nucleus, the activated Snf1 kinase phosphorylates and deactivates TF Mig1,[83] and phosphorylates and activates TFs Sip4 and Cat8.[84, 85] In addition, the downstream Snf1 targets are interconnected by more complex regulation, such as deactivated Mig1 releases expression of *CAT8*[86] and activated Cat8 induces expression of *SIP4*.[87] Through this complex transcription regulatory network, Snf1 acts as a key node in glucose repression regulatory cascade and regulates gluconeogenesis, respiration and catabolism of alternative carbon sources.[65, 80]

1.3.4.2 Response to various carbon signals

The Snf1 kinase integrates cellular responses to different carbon signals. For example, the glucose repression regulatory cascade is interlocked with glucose induction and galactose induction regulatory cascades (Figure 1). Through the regulation of TF Mig1, the Snf1 kinase regulates the expression of genes coding for members of the glucose induction pathway (sensor Snf3 and signal transducer Mth1) and for glucose transporters Hxt2 and Hxt4.[88] Genes coding for enzymes and regulatory proteins within galactose metabolism are under dual regulation by galactose induction regulatory cascade and

by glucose repression regulatory cascade.[89] In the presence of glucose, Mig1 represses *GAL* expression and in the presence of galactose, *GAL* genes are activated by the TF Gal4 (Figure 1).

During my thesis work the physiology of the *Δsnf1*[f], *Δsnf4* and *Δsnf1Δsnf4* strains compared to the wild-type strain was studied aiming to identify the roles of the Snf1 kinase and its regulatory subunit Snf4 in the regulation of galactose metabolism under dynamic growth conditions (chapter 2). For this study, glucose and glucose-galactose batch cultivation conditions were used (Figure 8). These experimental conditions created a base for tracking Snf1 kinase's and its regulatory subunit's Snf4 role in glucose rich, glucose low and in galactose rich dynamic growth conditions. The study indicated that Snf1 and Snf4 had independent roles regulating galactose catabolism. The study also showed that the Snf1 kinase complex regulated galactose metabolism not solely in response to available glucose or galactose, but also in response to available intracellular energy (ATP) that was required for the induction of galactose catabolism (Figure 1).

1.3.4.3 Diauxic shift and generation of energy

When glucose is exhausted, cells are subjected to 'diauxic shift', during which the cell produces the enzymes needed to metabolize the other available carbon source, and switches carbon metabolism from fermentation to respiration. The expression of more than 2000 genes is changing through the diauxic shift and more than 400 of these genes are regulated by the Snf1 protein kinase.[66] It is not yet known how the Snf1 protein kinase regulates expression of these 400 genes, but some mechanisms are identified (Figure 6). Besides its role in glucose repression regulatory cascade, the Snf1 protein kinase phosphorylates TF Adr1 and induces Adr1 binding to chromatin.[90] Consequently, Adr1, together with different TFs, regulates biochemical pathways that generate precursors of energy, acetyl-CoA and NADH, from non-fermentable carbon substrates.[66] For example, Adr1 together with Cat8 regulate expression of genes within ethanol catabolism, Adr1 together with Pip2 and Oaf1 regulate fatty acid β-oxidation,[91] and Adr1 together with Ino2,4 regulate expression of *INO1* and *GUT1* that are also involved in lipid metabolism.[92]

In addition, Snf1 contributes to transcription initiation on a more global scale.[67, 68] The Snf1 and the acetyltransferase Gcn5 function in an obligated sequence, Snf1 phosphorylates histone H3 Ser10 and Gcn5 deacetylates histone H3 Lys14, thereby enhancing transcription of e.g. *INO1* and *GAL1* contributing to the regulation of lipid and carbon metabolism.[68, 69] It has also been identified that the

[f] The *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* mark recombinant strains, in which the *SNF1*, *SNF4* or both genes are deleted (i.e. knocked-out) from the yeast genomic DNA, respectively.

Snf1 protein kinase through the interaction with RNA polymerase II holoenzyme[g] may regulate expression of hundreds of genes[67] and so regulate metabolic changes during the diauxic shift.[66] These findings indicate that the Snf1 kinase through the regulation of TFs, being part of transcription initiation machinery and being a regulator of histones, may regulate the expression of hundreds of genes and be a global regulator of metabolism in yeast.

1.3.4.4 Response to stress

The Snf1 protein kinase regulates protective mechanisms against various stresses including those that arise simultaneously with glucose depletion.[48, 49] When cells are starved for glucose, besides deactivating the glucose repression regulatory cascade, the Snf1 also promotes expression of salt-stress-responding *ENA1*,[93] activates heat shock TF Hsf1,[94] and deactivates the stress-responsive TF Msn2.[95] In the reconstructed regulatory network of the Snf1 protein kinase (Figure 1 in chapter 4), I identified that the Snf1 kinase regulates protective mechanisms against various stresses including oxidative stress. Recently, Wiatrowski & Carlson have identified the protein-protein-interaction between Sip2 and Yap1[h].[76] The presence of this interaction suggests that the Snf1 kinase may regulate carbon and oxidative metabolism in a coordinated manner.[76, 96] Based on results obtained in my thesis work and on previous knowledge, I suggest that the Snf1 protein kinase in different heterotrimeric complexes regulate various stress responses: e.g. the Snf1 protein kinase in complex Snf1-Snf4-Gal83 regulates carbon metabolism and the Snf1 protein kinase in complex Snf1-Snf4-Sip2 is involved in the regulation of redox balancing (chapter 4).

1.3.4.5 Response to nitrogen and coordinated Snf1-Tor1 regulation

It is generally accepted that the Snf1 protein kinase is active at low glucose concentrations. In addition, scientific evidence presented below shows that the Snf1 protein kinase has basal activity when cells are grown in excess of glucose. Unphosphorylated Snf1 that has basal kinase activity participates in nitrogen signaling,[97] provides resistance to toxic hypoxyurea and toxic cations[98, 99] and regulates diploid pseudohyphal differentiation[75] and filamentous growth.[100] The Tor (Target Of Rapamycin) kinase is a well studied regulator of cell growth and, as the Snf1 protein kinase, is described to regulate nitrogen signaling, filamentous growth and response to stress.[101] Also, both Snf1

---

[g] The RNA polymerase II holoenzyme is a large protein complex that consists of the RNA polymerase core complex and a variety of other proteins including transcription factor complexes. The RNA polymerase II holoenzyme is recruited to the promoters of protein-coding genes and initiates transcription of DNA to synthesize precursors of mRNA.

[h] The Yap1 is a basic leucine zipper transcription factor that is required for oxidative stress tolerance. Yap1 activates the transcription of anti-oxidant genes in response to oxidative stress.

and Tor1 regulate expression of *INO1,*[102] phosphorylate and activate TF Gln3[103] and regulate the subcellular localization and activity of TF Msn2 in response to carbon or nitrogen limitation, respectively.[95] These evidences indicate that the functions of the Snf1 and Tor1 kinases are cross-linked.

In my work, reconstructed Snf1 kinase regulatory network (chapter 4) identified that the loss of the Snf1 kinase impacted general and redox stress response as well as nitrogen and lipid metabolism, all of which were also regulated by the Tor1 protein kinase (Figure 1 in chapter 4). My results supported the prior knowledge and, furthermore, suggested that Snf1 and Tor1 both played a role in integrating information on the nutritional state and performed coordinated regulation of energy, redox metabolism, and, consequently, aging (chapter 4).

In homology to yeast, mammalian AMPK and mTOR[i] may have reciprocal and overlapping functions. In hypothalamus, AMPK and mTOR act as fuel sensors and antagonistically regulate energy intake.[104, 105] The mTOR is also regulated by AMPK. Activated AMPK inhibits mTOR suppressing gluconeogenesis and protein synthesis in the peripheral tissues, thereby conserving low energy levels during low energy states.[58, 106]

1.3.4.6 Lipid metabolism

The Snf1 kinase regulates energy generating fatty acid β-oxidation and energy consuming biosynthesis of fatty acid, sterol and lipid. The yeast Snf1, as its mammalian homolog AMPK, directly phosphorylates and inactivates Acc1 (Acetyl-CoA Carboxylase) and therefore inhibits fatty acid and sterol biosynthesis.[70, 71] Also, the Snf1 regulates the expression of *INO1* and *GUT1,2* that are involved in the biosynthesis of phospholipids.[70, 102] It is also believed (but not yet proven) that, as mammalian and plant AMP-activated kinase,[107, 108] the yeast Snf1 protein kinase phosphorylates and inactivates Hmg1 (3-Hydroxy-3-Metylglutaryl-CoA reductase), which catalyzes a rate limiting step in sterol biosynthesis.

On a global scale and for the first time my work showed that the Snf1 kinase regulated fatty acid β-oxidation, fatty acid synthesis and biosynthesis of phospholipid precursors (Figures 1, 2 in chapter 4; SI Figure 5 in chapter 5). In homology to mammalian system, I showed that the yeast Snf1 protein kinase controlled carnitine metabolism and consequently co-regulated the fatty acid β-oxidation and the synthesis of fatty acids and sterols (Figure 2 in chapter 4).

[i] The mTOR is a mammalian homolog of the yeast Tor proteins that respond to nutrient limitation and control cell growth.

1.3.4.7 Specific roles of the three distinct Snf1 kinase complexes

The Snf1 protein kinase in the Snf1-Snf4-Gal83 complex participates in the glucose repression regulatory cascade (Figure 1). The role of the Snf1-Snf4-Sip1 complex is not yet determined. The cyclic AMP-dependent Pka (Protein Kinase A) maintains the Snf1-Snf4-Sip1 cytoplasmic localization in glucose-excess-grown cells.[35] The Snf1-Snf4-Sip2 complex has a role in controlling cell longevity and response to the oxidative stress.[76, 77] The *N*-myristoylated Sip2 keeps the Snf1 kinase complex by the plasma membrane. In aging cells, when the myristoyl moiety is lost, the Snf1's β-subunit Sip2 relocates from the plasma membrane to the cytoplasm, the Snf1 localizes to the nucleus and regulates aging by affecting chromatin structure and genomic stability.[34]

In my thesis work, the Snf1 regulatory network (Figure 1 in chapter 4) was reconstructed, and it was identified that, in response to the loss of *SNF1*, carbon metabolism and redox balancing changed. The analysis of the observed global metabolic changes indicated that the loss of *SNF1* affected cell longevity as well. Overall, I suggested that the Snf1 kinase mediated energy and redox balancing for optimal yeast growth: (i) the Snf1 kinase in complex with Snf4 and Gal83 regulated carbon and energy metabolism,[64, 79] (ii) the Snf1 kinase in complex with Snf4 and Sip2 was involved in the regulation of redox balancing and longevity[76, 77] (chapter 4).

1.3.4.8 The yeast Snf1, as its mammalian homolog AMPK, is a low energy checkpoint

The AMPK is described as a low energy checkpoint.[21, 60, 78] All energy-consuming reactions are powered by the high concentration of ATP. When cellular stress causes a drop in the intracellular energy, the activated AMPK acts as a 'master switch' inducing ATP-generating catabolic pathways and repressing ATP-consuming processes (Figure 2). The AMPK performs this regulation in different tissues, in response to various additional signals, mainly by either direct phosphorylation of target enzymes, or by regulating gene expression. For example, during exercise, activated AMPK increases glucose uptake in the muscle cells.[37, 78, 109] In white-fat tissues, AMPK balances body lipid storage.[110, 111] In response to low glucose level in blood, the AMPK participates in the inhibition of insulin production and secretion by islet β cells.[112, 113] Food intake is repressed when AMPK in the hypothalamus is inhibited by glucose, leptin and insulin. Food intake is stimulated when AMPK in the hypothalamus is activated by the hormones adiponectin and gherin.[114] Therefore AMPK regulates body weight.[58, 104] Overall, the pivotal role of AMPK places it in an ideal position for regulating whole-body energy metabolism, and thus AMPK could play a part in protecting the body from metabolic diseases such as type 2 diabetes[113] obesity,[60] cardiac hypertrophy and arrhythmia,[115, 116] and cancer.[115, 117] Although the Snf1 protein kinase complex has been more extensively studied in the context of glucose repression, its role in energy homeostasis, in homology to mammalian AMPK, is becoming increasingly evident (Figure 2).[20, 21] The studies discussed above shows that the Snf1

protein kinase is involved in the regulation of energy generation, energy consumption and energy storage.[70, 79, 118]

My Ph.D. work was the first global scale study that demonstrated that the Snf1 protein kinase had coordinated control over energy producing and energy consuming metabolic pathways (chapter 4). Overall, my study detected more similarities between the functions of the yeast Snf1 and the mammalian AMPK than previously published. This opened the door for broader use of yeast as a model for studying the role of AMPK in regulating different parts of human metabolism and, ultimately, for better understanding and addressing metabolic disorders.

1.4 Top-down Systems Biology approach

The completion and publication of the genome sequences of *Haemophilus influenzae* in 1995[119] and the yeast *Saccharomyces cerevisiae* in 1996[120] has been a milestone in the history of biological research. Now, a traditional data-poor discipline has been transformed to data-rich discipline and this has pushed scientists towards a new view of biology, namely Systems Biology, where data-driven discovery meets hypothesis-driven science.[121] Systems Biology does not emphasize the investigation of the individual genes or proteins independently, as has been the highly successful trend for the past 30 years[j]. Rather, it investigates the behavior and relationships of all of the elements in a particular biological system.[16, 122] An important challenge that is faced by investigators today is interpreting new large-scale data sets (e.g. global-scale gene expression dataset, large-scale protein expression or interaction dataset) and thereby deriving fundamental and applied biological information about whole biological systems.[121, 123] To accomplish that, the following are needed:

(i) system-wide component identification,

(ii) identification of macromolecule interactions within the cell,

(iii) computational inference of biological system and

(iv) rigorous integration of heterogeneous data (Figure 7).[123]

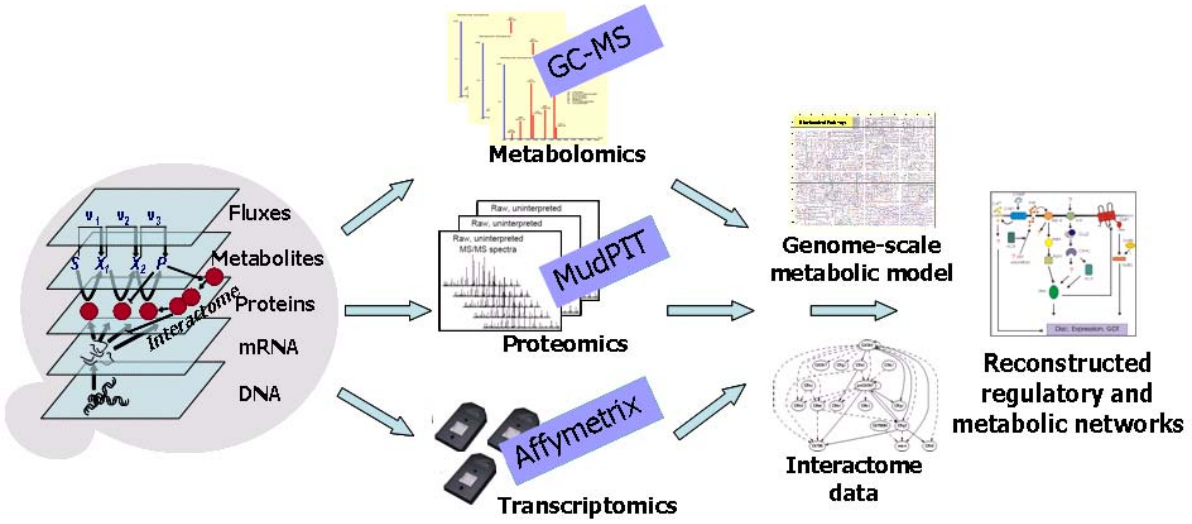


Figure 7. The roadmap of Systems Biology Approach taken in my Ph.D. work. The mRNAs, proteins and metabolites were extracted from chemostat grown wild-type, *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains and analyzed using Affymetrix DNA microarray system, MudPIT (Multidimensional Protein Identification Technology) followed by stable-isotope-labelling-based quantification, and GCMS (Gas Chromatography-Mass Spectrometry), respectively. Quantified 5544 mRNAs, 2388 proteins and 44 metabolites were integrated with available protein-protein ((BIOGRID-Saccharomyces_cerevisiae v.2.0.25), protein-DNA[124, 125] and protein-metabolite interaction[126] datasets with the aim to reconstruct the Snf1 kinase regulatory network. The strategy and results of my study are presented in chapters 4, 5.

[j] The single gene/protein based approach is not obsolete or bad. In fact, a lot of systems biology is possible only because single gene approaches generated enough information so that now one can start looking at a systems rather than individual components. Systems Biology is a natural 'evolution' from previous approach, which was, is and will be equally important.

This chapter will briefly walk through these four strategic steps of the Systems Biology Approach and summarize the work that has been done, including contributions from my Ph.D. work, toward a truly holistic understanding of cellular behavior.

1.4.1 System-wide component identification

One of the major objectives of Systems Biology is to reconstruct biological networks and perform a quantitative description on how these operate under conditions of interest. In top-down Systems Biology[k], consistent and comprehensive quantitative data sets of mRNA, proteins and metabolites are unprecedented starting points to understand systems and provide a solid basis for data integration and further quantitative modeling (Figure 7).[122, 127] It is still not possible to study biology solely *in silico* and it is thus crucial to use an experimental set-up, in which generated datasets can be analyzed together to understand biological systems. To accomplish this, the biological samples of interest have to be cultivated in well-suited growth conditions (e.g. chemostat cultivations), and analyzed by using compatible data analysis platforms (e.g. either cDNA microarrays[128] or oligonucleotide arrays).[129-131] Many of the available Systems Biology studies require high-throughput analyses that are expensive, and it is therefore often limiting factor in performing all needed experiments in one laboratory. Thus, the use of previously published experimental data available from various repositories, such as computer databases and supplementary material of published articles, is commonly used and highly desirable. The available data from various repositories also facilitates the development of computer algorithms and other theoretical studies that requite huge data sets. This creates a definite need for collecting and sharing large datasets, however, inconsistencies among cross-platform and cross-laboratory studies remain an important issue that must be carefully evaluated before these resources are used in the field.[121, 127]

1.4.1.1 Chemostat cultivation

Chemostat cultivations are well-controlled growth conditions.[132, 133] Parameters that affect cell growth, such as nutrient availability, aeration, stirring and temperature, are monitored and kept constant throughout the chemostat cultivation. Nutrients are continuously fed to the culture and the culture fluid is removed (with the same rate) from the cultivation chamber keeping the culture volume constant (Figure 8). This generates steady-state growth conditions, where the specific growth rate, biomass, cell metabolic and regulatory processes are constant. Based on the above characteristics, chemostats

[k] The top-down Systems Biology is a systemic-data driven approach. It starts from measuring genome-wide experimental data and it is followed by data analysis and data integration to discover or refine pre-existing models that ultimately describe the measured data. In this way, previously unidentified interactions, mechanisms and molecules can be identified, and hypotheses concerning co- and inter-regulation of groups of those molecules can be put forward.

are highly reproducible and are well suited to study cell response to a single specific environmental or genetic parameter, while minimizing the number of confounding parameters found in batch cultivations. Ultimately in Systems Biology, data generated from the chemostat cultivations are well suited for the design of genome-scale models.

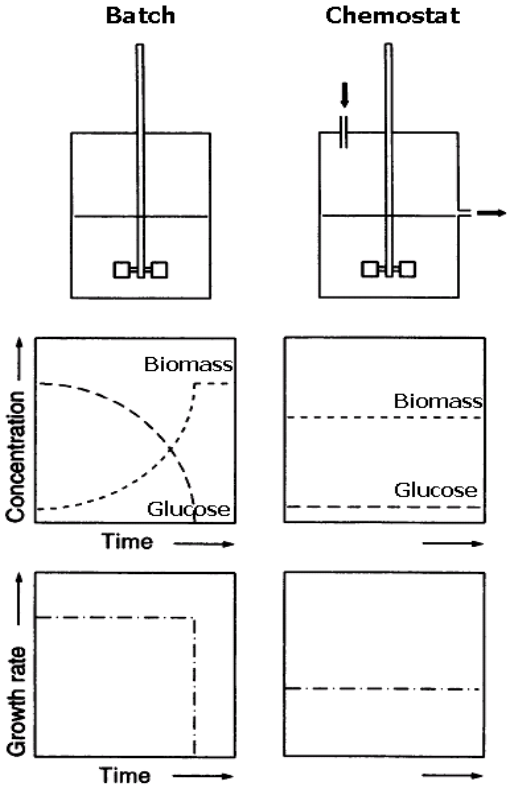


Figure 8. Comparison of batch and chemostat cultivation. During batch cultivations, glucose is consumed and yeast cells grow exponentially. After depletion of glucose, the specific growth rate drops to zero. Differently, in chemostat cultivations, all relevant growth parameters, including glucose concentration in the culture, are constant in time and this creates steady-state growth conditions, were biomass and growth rate are constant. To study the impact of deletion of *SNF1* and *SNF4* genes on the regulation of carbon metabolism, dynamic glucose and glucose-galactose-mixture batch cultivations were used and the results were presented in chapter 2. To generate high-quality large-scale transcriptome, proteome and metabolome datasets, the chemostats cultivations were used. In chemostat cultivations only the deletion of *SNF1* or *SNF4* affected omics dataset differences between the strains of comparison. Extracted proteome samples were used in a large-scale quantitative proteome study (chapter 4). Extracted proteome, transcriptome and metabolome samples were used in an integrative top-down Systems Biology study (chapter 4). This figure is borrowed from the review, written by Weusthuis.[133]

In my thesis work, wild-type, *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains were grown in carbon-limited chemostat cultivations to ensure that metabolic and regulatory changes in these strains were specific to disruptions of the Snf1 complex, and not complicated by external effects resulting from the specific mutant physiology (e.g. different specific growth rates). From these cultivations extracted

transcriptome, proteome and metabolome datasets were successfully used in large-scale comparative and Systems Biology studies (chapters 3, 4).

1.4.1.2 Transcriptomics

Transcriptomics is the study of the complete set of all messenger RNA molecules i.e. "transcripts" that are produced by the genome discretely in one cell or in a population of cells. The field of transcriptomics provides information about both the presence and the relative abundance of transcripts thereby indicating the active components within the cell. Today DNA microarrays represent a robust and powerful discovery tool for expression profiling of thousands of genes simultaneously. Two types of DNA microarrays, cDNA microarrays,[128] developed at Stanford University, and oligonucleotide arrays,[129-131] developed at Affymetrix, Inc., are widely used. Although array technology is exactly the kind of comprehensive high-throughput approach that Systems Biology demands, the various levels of post-transcriptional control might challenge its importance, because proteins, but not mRNAs, are the functional products[I] of the coding sequences within the genome. Several studies have been dedicated to evaluate whether measured mRNAs can be used to extrapolate protein abundances in yeast.[134-136] A recent thorough study has concluded that the correlation between mRNA and protein levels is weakly (70%) positive and this is insufficient to predict protein expression levels from the quantitative mRNA data.[136]

In this thesis, I generated global-scale gene expression data for the *S. cerevisiae* wild-type, *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains using Affymetrix GeneChip® Yeast Genome 2.0 Arrays. The generated gene expression data was analyzed together with proteome, metabolome and various interaction datasets to reconstruct Snf1 kinase regulatory network (chapter 4). In collaboration with Roberto Olivares (CMB) and Kiran R. Patil (CMB), the generated transcriptome and proteome datasets were used to study translation efficiency in yeast. Searching for parameters that influence translation (e.g. codon bias, availability of tRNA, protein production cost, protein length), we aimed to mathematically define translational efficiency, and determine if measured mRNAs could be used to predict protein expression levels (ongoing project, CMB).

[I] Proteins are called functional products of the coding sequences, because proteins, but not mRNA, perform functions in the cell, e.g. catalyze metabolic reaction, transduce signal, or provide structure for the cell.

1.4.1.3 Proteomics

Ultimately proteomics aims to identify and quantify cellular levels of all proteins that are encoded by a genome. The techniques used to perform complete proteome analysis are still limited, mainly due to the high dynamic range (10^1 to 10^6 molecules per cell) of the proteome.[137] The protein modifications, interactions, localization in various cellular compartments[138] and different protein turn-over times[139] also contribute to the complexity in the development of the field. Various tools combining mass spectrometry and liquid chromatography or gel-based protein separation, have been designed and used to improve proteome identification and quantification.[140] Several technical advances include: (i) multidimensional protein identification technology (MudPIT), developed in Prof. Yates' lab, that separates and identifies proteins from complex mixtures;[141, 142] (ii) *in vitro* protein labeling, developed in Prof. Aebersold's lab, that uses isotope coded affinity tags (ICAT);[143] (iii) promising non-labeled quantitative proteomics approach,[144] that is still under development in Prof. Smith's lab, and (iv) tools that identify protein modifications.[145] Other proteome identification and quantification tools based on

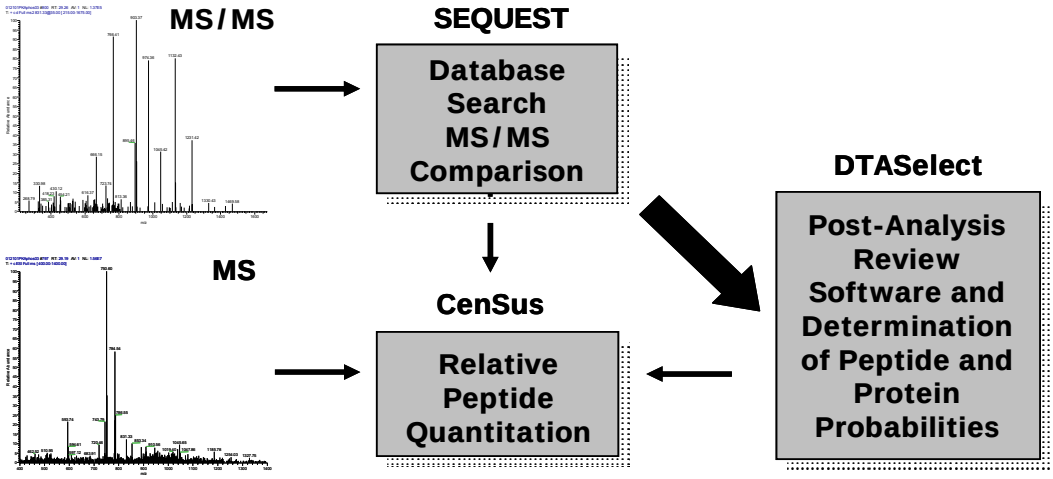


Figure 9. The schematic representation of performed quantitative proteome analysis. The collected ^{14}N-labeled global protein samples from the wild-type, and ^{15}N-labeled global protein samples from the *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains were analyzed using micro-liquid chromatography coupled mass spectrometry, otherwise called MudPIT.[146] Generated both MS and MS/MS spectra[m] were subjected to the computational analysis to produce quantitative proteome dataset. MS/MS were searched with SEQUEST algorithm[147] against a database of the translated *Saccharomyces cerevisiae* ORFs, and so the identification (sequencing) of the detected proteins was performed. SEQUEST results were assembled, filtered and transformed to user-friendly .txt files using the DTASelect2.0 algorithm.[148] SEQUEST and DTASelect2.0 output files were submitted to CenSus software[149] to calculate the relative protein abundance differences based on reconstructed ion chromatograms (i.e.,

[m] The MS spectrum is a full scan of specific mass-to-charge (m/z) peptide ions, which, in our case, was generated in a mass spectrometer and mass analyzer Orbitrap. This MS spectrum was used for relative quantification of the proteome. The MS/MS spectrum is a tandem mass spectrum, that is generated by additional fragmentation of peptide ions to sequence (i.e. identify) proteins. In our case, a tandem-linked LTQ mass spectrometer was used to generate MS/MS spectrum.

$^{14}N/^{15}N$ CenSus Ratio). This generated quantitative proteome dataset was analyzed in the large-scale proteome study, described in chapter 3, and in the Systems Biology study, described in chapter 4.

protein chip technology or other mass spectrometry-based strategies have found their niches in specific proteome areas as reviewed in Patterson & Aebersold.[150] For example, the protein microarray containing a whole yeast proteome on a chip can be used for proteome-wide studies identifying protein interactions, post-translation modifications, antibody specificity and protein activity.[63, 151, 152]

In this thesis, I generated the largest, to date, yeast quantitative proteome dataset in Dr. Yates' *Proteomics Mass Spectrometry* lab. The proteome of the wild-type yeast strain was metabolically labeled in chemostat cultivations using >99% ammonium-^{15}N sulfate, as the sole nitrogen source. The three mutant strains were metabolically labeled using sole >99% ammonium-^{14}N sulfate. The total protein was extracted from the cells, separated using MudPIT and quantified using stable isotope labeling approach (Figure 9) or spectral counting. The comparison and characteristics of these two quantitative datasets was described in chapter 3. The generated high-quality quantitative proteome dataset was used in the Systems Biology Study (chapter 4), where the Snf1 kinase regulatory network was reconstructed.

1.4.1.4 Metabolomics

Metabolomics (also referred to as metabolite profiling) aims at identifying and quantifying the complete set of metabolites in the cell. Metabolites are the mediators and products of metabolism, and thus, they reflect a cell's phenotype the best. In terms of Systems Biology, metabolome datasets are not easy to analyze, since the relationship between the metabolome and the genome is indirect (Figure 7).[153] Metabolite levels are the final result of a complex integration of gene expression, RNA translation, post-translation modification, enzyme activity, and pathway regulation.[154] The high-throughput methods for analyzing as many metabolites as possible in a single experiment are still being developed and typically rely on mass spectrometry and NMR spectroscopy.[155] No method is yet available to analyze a complete metabolite set.

Contributing to this thesis work, post doc Michael C. Jewett (CMB) generated the metabolome dataset of 44 metabolites from chemostat grown wild-type, *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains (chapter 5). I used this metabolome dataset as a back-up experiment that substantially contributed to analyzing transcriptome and proteome datasets, and evaluating the reconstructed regulatory network of the yeast protein kinase Snf1 (chapter 4).

1.4.1.5 Data repositories

To share results of performed transcriptomics, proteomics and metabolomics analyses, members of the scientific community deposit omics data in computer databases, or publish it as Supplementary Online material. To date, hundreds of transcriptome (e.g. http://www.ebi.ac.uk/microarray-as/aer/?#ae-main[0]), proteome (e.g. http://bioinformatics.icmb.utexas.edu/OPD/) and metabolome (e.g. http://www-en.mpimp-golm.mpg.de/03-research/index.html) databases are publicly available free of charge. Most databases require that the submitted experimental data is accompanied with additional information that is needed to unambiguously interpret the results of the submitted experiment and to potentially reproduce the experiment. For example, MIAME (Minimum Information About a Microarray Experiment) standard has been created to document microarray experiments. The proteome databases (e.g. http://info.med.yale.edu/proteome/) have started to use MIAME standard for reporting proteome datasets as well.

The omics data generated in my Ph.D. work is publicly available. The microarray data has been submitted to the ArrayExpress (http://www.ebi.ac.uk/microarray-as/aer/) and is available using the accession number E-MEXP-1407. Proteome dataset has been submitted as a supplementary material for Usaite et al (2008), *J Prot Res* (chapter 3) and is available at http://pubs.acs.org. During my Ph.D. work I have also contributed to the development of the MIAME compliant MicroArray Database (http://www.fbd.dtu.dk/fbd/), in which all CMB generated DNA microarray data is going to be collected.

1.4.2 Identification of macromolecule interactions within the cell

In the overview written in 1998, Alberts has stated that 'the entire cell is as a factory that contains an elaborate network of interlocking assembly lines, each of which is composed of a set of large protein machines'.[156] Before we can understand and describe the structure of the cell, it is necessary to uncover these 'assembly lines'. During the last decade, comprehensive and systematic identification of protein-protein, protein-DNA interactions and protein complexes has accelerated unraveling the functional organization of the yeast proteome, contributing to the understanding of the cell as a whole.[157, 158]

1.4.2.1 Protein-protein interaction

Through signaling cascades and enzyme complex formations, protein-protein interactions dictate many cellular processes. Pioneering Uetz et al[159] and Ito et al[160] have used a technique called yeast two-hybrid system to identify yeast protein-protein interactions on a global scale. The yeast two-hybrid system is a preferable method, since it validates protein interactions *in vivo*. However, only two

proteins are tested at a time and the method predicts only 'possible' interactions, since the hybridization takes place in the nucleus that is not a native compartment for many proteins.[161] The more recently developed technique, affinity purification mass spectrometry of protein complexes, on the other hand, is a promising high-throughput approach. It detects real complexes in physiological settings, and it misses only those complexes that are not present under conditions studied.[161] Pioneering Ho et al,[158] Gavin et al[157] and Krogan et al.[162] have performed affinity purification mass spectrometry studies. Today the knowledge about measured and predicted protein-protein interactions are collected in computer databases that vary by their purpose, size and the sources of protein-protein interactions posted.[62, 125, 163] In my top-down Systems Biology study (chapter 4) I used protein-protein interaction data from the BioGRID database (http://www.thebiogrid.org) that holds information on 57,680 unique protein-protein interactions from 3,868 unique proteins in yeast (BIOGRID-Saccharomyces_cerevisiae v.2.0.25). A variety of techniques, highlighting yeast two-hybrid system, mass-spectrometry based approaches, genetic interactions, co-expression, synthetic growth defect and phenotypic enhancement, contribute to the uncovering of complete yeast interactome.[62]

1.4.2.2 Protein-DNA interaction

Interactions between proteins and DNA mediate transcription, DNA replication, recombination, and DNA repair. Using chromatin immunoprecipitation in conjunction with microarrays, what is called ChIP-chip technology,[164] it is now possible to measure the *in vivo* binding of many transcription factors to the promoters of most genes. Using ChIP-chip technology, Harbison et al has reconstructed an initial map of yeast transcriptional regulatory code by identifying the DNA sequence elements that are bound by 204 regulators under various conditions and that are conserved among *Saccharomyces* species.[124] Just in 3 years, the Harbison et al study has been cited by more than 300 other publications and the dataset of 10,884 high-confidence ($P < 0.001$) protein-DNA interactions has been used as a reference in various studies and algorithms to elucidate the underlying structure of the transcriptional regulatory network.[165-167]

In my Ph.D. work, the protein-DNA interaction dataset was used in Reporter Effector analysis[168] and in high-scoring subnetwork analysis (chapters 4, 5). The outputs of these two analyses contributed to the reconstruction of the Snf1 protein kinase regulatory network (chapter 4).

1.4.3 Computational inference of structure

Most transcriptome, proteome, metabolome, protein-protein interaction, and protein-DNA interaction analyses provide us with tens, hundreds, or even thousands of data nodes, often expanding the known number of expression changes or interactions by order of magnitude. The immediate task is to

develop and use computational strategies to verify and analyze global-scale data with the intention of generating hypotheses and making biological discoveries.

1.4.3.1 Verification of omics data

Replication, normalization and statistical evaluation are required to create and determine the significance of any biological experiment. Normalization has to be done within each and between all datasets of comparison. A minimum of three biological replicates are required to be able to perform valid statistical analysis. Student's t-Test or ANOVA statistical tests are popular in the analysis of transcriptome, proteome or metabolome datasets.

In my Ph.D. work, all experiments were performed in three biological replicates and all generated datasets were statistically validated using Student's t-Test analysis. The DNA microarray data was normalized using dChip software,[169] proteome and metabolome data were normalized based on linear distribution.

1.4.3.2 Cross-platform variation

A cross-platform variation remains to be an issue among metabolome, proteome or transcriptome datasets. The consistency among the data must be carefully verified before a comparison of two datasets is valid. For example, a detailed study, which has compared by various methods identified protein-protein interactions, shows that among 80,000 available protein-protein interactions, only 2,400 (3%) are identified by two or more techniques.[161] This mismatch may be explained by not optimally working methods for any type of protein-protein interactions and by false positives that are generated using any of the methods.[161] Different confidence assignment schemes are developed to combat false positive protein-protein interactions.[170] For example, one of these algorithms is implemented to the DIP (protein-protein interaction) database (http://dip.doe-mbi.ucla.edu/Services.cgi) and this algorithm estimates that only ~50% of the protein-protein interactions that are collected in the DIP database are reliable.[171] In the other example, Michael C. Jewett (CMB) has experimentally demonstrated that different metabolite quenching and extraction methods that are used on the same biological samples result in different quantitative and even qualitative metabolome datasets (unpublished results). In Systems Biology, where several large-scale datasets are frequently analyzed together to understand a biological structure, it is important to make sure that the datasets used in the analysis are compatible.

During my Ph.D. work, I contributed to an ongoing CMB project, where the same biological samples were analyzed using two types of GeneChip® Yeast Genome Arrays (YG_98 and Yeast_2) that were designed by Affymetrix. I found that the gene expression results generated using these two platforms were not equivalent for every gene on the chip. The expression change of some genes measured on one platform deviated from the expression change measured on the other platform. The only

difference causing this inconsistency was that the oligo-probes placed on the YG_98 and Yeast_2 arrays had different chemical properties. Computational approaches and prediction tools are created that ultimately improve comparison of the data generated using different analysis platforms.[172-174] To date, there are no reliable means to correlate complete gene expression profiles among different DNA microarray systems.

1.4.3.3 Data mining

Methods to analyze gene expression datasets have received the most attention to date, because the technology to generate gene expression data had spread widely much before technology to generate quantitative global proteome or metabolome datasets has been available. Clustering, for example, groups genes by their co-expression patterns[175, 176] and is frequently used as the first gene expression data analysis step. Results of clustering suggest the presence of regulatory circuits within groups of co-expressed genes. The clustered gene expression dataset is frequently further analyzed in the context of the biological processes, molecular functions, or genomic sequences to pinpoint metabolic or regulatory changes that are present in the biological system of interest. For example, the promoter analysis that groups genes with shared regulatory sequences in their promoters[177, 178] is commonly used on the clustered gene expression data. The promoter analysis determines the probability with which a certain transcription factor may be responsible for gene expression change within a group of co-expressed genes. Different data analysis algorithms and analysis tools are still being created. The tools used in gene expression analysis methods, such as clustering, are transformed to analyze proteome and metabolome datasets.

1.4.3.4 Data integration to reconstruct complete cellular networks

Data integration is a promising global-scale data analysis strategy that ultimately yields molecular interaction networks that more closely reflect real cellular structures. Through the integration approach, the high-error rate within global-scale datasets is likely mitigated, and thus, the quality of the biological information obtained is improved. For example, through the joined analysis of genetic, biochemical, and ChIP-chip experimental data, global transcription regulation networks have been reconstructed for *S. cerevisiae* and *E. coli*.[165, 179] These networks can help to predict novel protein-DNA interactions and identify transcription regulatory circuits. The other example illustrates that the comparison of protein-protein interaction networks, which are generated for worm, fly and yeast, shows that, although any single network contains false-positive interactions (discussed above), embedded beneath this noise there is a repertoire of protein interaction complexes and pathways that are conserved across all three species.[180]

1.4.3.5 Data integration to identify biologically active subnetworks

The data integration approaches also aim to pinpoint the portions of the networks that are the most active under a studied condition. In one of the first data integration studies, a protein-protein interaction dataset has been analyzed with gene expression data to identify the most responsive protein-protein interaction subnetworks.[181] The identified the most responsive subnetworks describe highly active

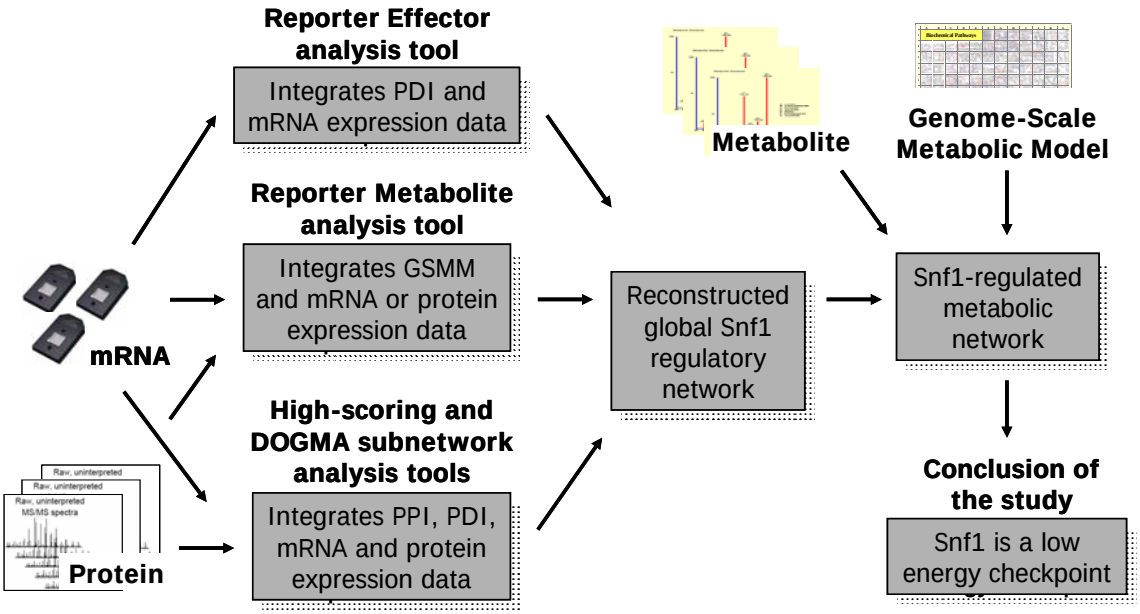


Figure 10. The strategy taken in this Ph.D. thesis to reconstruct and validate Snf1 kinase regulatory network. In my thesis generated global-scale mRNA and protein datasets were analyzed using Reporter Effector,[168] Reporter Metabolite[182] and subnetwork analysis tools. These approaches were in detail described in chapter 5. The outputs of these analysis tools were used to reconstruct global Snf1 kinase regulatory network that was depicted in Figure 1 in chapter 4, SI Figures 3, 4 in chapter 5. The reconstructed regulatory network was analyzed using genome-scale metabolic network and using metabolome dataset generated in this study until biological conclusion was reached. PPI = protein-protein interaction, PDI = protein-DNA interaction, GSMM = genome-scale metabolic model.

regulatory modules of connected proteins and regulated genes that significantly change in response to a perturbation.[181] In the other study, protein-protein interaction dataset has also been integrated with gene expression data and the dynamic protein complex formation during the yeast cell cycle has been reconstructed.[183] By integrating transcriptome, proteome and metabolome datasets, Ishii et al has determined that the *E. coli* metabolic network employs fundamentally different regulatory strategies in response to either environmental or genetic perturbations.[184]

The main goal of my Ph.D. work was to reconstruct the global Snf1 protein kinase regulatory network (Figure 1 in chapter 4). To achieve this, generated transcriptome, proteome and metabolome datasets

were integrated with available protein-protein, protein-DNA and protein-metabolite interaction datasets using Reporter Metabolite,[182] Reporter Effector[168] and high-scoring subnetwork[185] analysis tools (Figure 10). In this data integration study, the novel biological discoveries were strengthened through elimination of false positive data within the large-scale datasets used, and through the confirmation of novel findings on the different cellular levels (chapters 4, 5).

1.4.4 Rigorous integration of heterogeneous data

Many properties of complex biological systems cannot be understood from monitoring the components, unless such information is connected and integrated through mathematical models. The reason is that static component concentrations, albeit extremely informative, do not contain functional informational *per se*. The functional behavior of a network emerges through the nonlinear gene, protein, and metabolite interactions across multiple metabolic and regulatory layers. To really achieve a holistic, quantitative, and predictive understanding of whole systems, we have to develop cellular or systems mathematical models that enable an iterative cycle between prediction and experiments.[121, 123] This ultimately brings Systems Biology up to a level at which it is applicable in industry and medicine.

1.4.4.1 Cellular modeling and analysis

Generated omics data provide measurements that strive not only to integrate data, but also to model the entire cell. These global approaches are the most powerful because they define all possible components of the system, do not require any prior assumptions, and do not involve hypothetical interactions. More than a dozen genome-scale metabolic networks for various organisms are reconstructed,[126, 186] and those are arguably the best characterized complex biological networks, since they demonstrate network behaviors that are not apparent from examination of a few isolated interactions alone. The COBRA (Constraints-Based Reconstruction and Analysis) technique[187] emerged as a successful approach to modeling systems on a genome scale and it seeks to clearly distinguish biologically feasible from biologically unfeasible network states, rather than exactly predicting network behavior. The knowledge of created metabolic networks when integrated e.g. with gene expression data can be used to uncover the structure of transcription regulation.[182, 188] Recently, a reconstructed yeast regulatory network has been coupled with a genome-scale metabolic network.[167] This improved study proves how a systematic approach can be used to fill in missing information and to search for novel regulatory mechanisms. These are just initial steps and much work remains to be done to capture the true internal state of the cell through whole-cell modeling.

1.4.4.2 Systems Biology as an applied science

The integration of omics data in the context of Systems Biology has primarily affected basic biological research. Nevertheless, Systems Biology has already proved to be essential in industry and medical applications. There are examples showing that through the use of Systems Biology, industrial biotechnological processes are improved and, therefore, the market price for commercial products drops.[189] Many studies of human systems are adopting omics data-integration strategies to identify biomarkers that are associated with a disease state. For example, through the data integration approach, transcriptome and metabolome biomarkers can be identified to detect specific metabolic changes and to define prognosis characteristics in human neuroendocrine cancers.[190] Proteomics tools also find applications in clinical diagnostics.[191, 192] Modeling of biological networks is believed to facilitate drug discovery, i.e. model simulations may lead to identification of optimal drug targets and avoidance of side effects.[193]

I will be proud if the transcriptome, proteome and metabolome datasets that are generated in my thesis contribute to the EU AMPKIN project, in which I am participating. This project, based on reconstructed yeast Snf1 protein kinase regulatory network, and based on close homology between yeast Snf1 and mammalian AMPK signaling cascade, aims to identify drug targets against type 2 diabetes. Ultimately, the Systems Biology community, through modeling of complete living cells, aims to develop personalized and preventative therapies.

1.5 Conclusions and future perspectives

My thesis work expands the knowledge of the Snf1 protein kinase's role in yeast. I used physiology and Systems Biology studies to generate novel biological findings about the regulation of yeast carbon and energy metabolism.

First, through the physiological study, I characterized *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains compared to wild-type strain and suggested roles of Snf1 and Snf4 in the regulation of galactose metabolism in dynamic growth conditions (chapter 2). I showed that the Snf1 protein kinase was not a sole regulator of galactose metabolism and that the Snf1 protein kinase and its regulatory subunit Snf4 likely had independent roles in the regulation of galactose metabolism. This study indicated that (i) the continuing Mig1-based repression of *GAL* genes (Figure 1 in chapter 1), (ii) the changes in cellular energetics (i.e. availability of ATP), and (iii) the accumulation of toxic galactose derivatives inhibited galactose growth initiation and maximum specific growth rate on galactose in the recombinant yeast strains compared to the wild-type strain during growth on glucose–galactose-mixtures. Further studies are needed to prove these hypotheses.

Second, using top-down Systems Biology approach, I reconstructed the global-scale Snf1 protein kinase regulatory network (Figure 10 in chapter 1, Figure 1 in chapter 4). Using global-scale mRNA, (largest to date) quantitative protein (chapter 3) and large-scale metabolite data generated in this study, previously generated and publicly available biological data, and computational methods, I identified proteins interacting with the Snf1 kinase, transcription factors and metabolic hotspots that responded to the disruption of the Snf1 kinase complex. My result both confirmed already known information on the regulatory network, proving the validity of my approach as well as novel knowledge were obtained. These Snf1 kinase targets represented key components of the global Snf1 protein kinase regulatory network. Using the information from the reconstructed regulatory network, I proved on a global scale that the yeast Snf1, as its mammalian homolog AMPK, was a low energy check-point (Figure 2 in chapter 1), which had not previously been shown. Through the integrated study I showed that the Snf1 protein kinase regulated fatty acid oxidation and fatty acid biosynthesis (Figure 2 in chapter 4). Furthermore, I showed that the Snf1 protein kinase regulated carnitine metabolism that, in homology to mammalian systems, played a role in fatty acid metabolism, which had not previously been demonstrated in yeast. Changes identified within redox metabolism and energy storage metabolism also contributed to proving the hypothesis that the yeast Snf1 protein kinase is a low energy checkpoint. My study also suggested that the Snf1 kinase mediated energy and redox balancing for optimal yeast growth: (i) the Snf1 kinase in complex with Snf4 and Gal83 regulated carbon and energy metabolism,[64, 79] (ii) the Snf1 kinase in complex with Snf4 and Sip2 was involved in the regulation of redox balancing and longevity.

To my knowledge, this study was the first where three types of ome-measurements were integrated to reconstruct a complex regulatory network. My work demonstrated the promise of integrating measurements from multiple cellular levels to reconstruct regulatory networks.

If the reconstructed Snf1 kinase regulatory network was further computationally analyzed within the yeast genome-scale mathematical models, signal transduction links between the Snf1 protein kinase and transcription factors regulated by Snf1 (Figure 1 in chapter 4) or other nutrient sensing regulating proteins (e.g. Tor1) could be identified. This would lead to a more thorough understanding of the Snf1 kinase's function within a complete yeast regulatory network. Through homology study between the yeast Snf1 kinase and mammalian AMPK signaling cascades and using mathematical modeling, the reconstructed regulatory network of the Snf1 protein kinase could be transferred onto the human cellular or even systems mathematical models and used e.g. for identifying drug targets against metabolic disorders such as diabetes and obesity, where the mammalian AMPK plays a key regulatory role.

1.6 References

1. Boekhout, T. Biodiversity: gut feeling for yeasts. *Nature* **434**, 449-451 (2005).
2. Walker, G.M. Yeast physiology and biotechnology. (John Wiley and Sons, 1998).
3. Lagunas, R. Misconceptions about the energy metabolism of *Saccharomyces cerevisiae*. *Yeast* **2**, 221-228 (1986).
4. Lagunas, R. Energetic irrelevance of aerobiosis for *S. cerevisiae* growing on sugars. *Mol.Cell Biol.* **27**, 139-146 (1979).
5. Van Dijken, J.P. & Scheffers, W.A. Redox balances in the metabolism of sugars by yeasts *FEMS Microbiology Letters* **32**, 199-224 (1986).
6. Bailey, J.E. Toward a science of metabolic engineering. *Science* **252**, 1668-1675 (1991).
7. Ostergaard, S., Olsson, L. & Nielsen, J. Metabolic engineering of *Saccharomyces cerevisiae*. *Microbiol.Mol.Biol.Rev.* **64**, 34-50 (2000).
8. Porro, D., Brambilla, L., Ranzi, B.M., Martegani, E. & Alberghina, L. Development of metabolically engineered *Saccharomyces cerevisiae* cells for the production of lactic acid. *Biotechnol Prog.* **11**, 294-298 (1995).
9. Hallborn, J. et al. Xylitol production by recombinant *Saccharomyces cerevisiae*. *Biotechnology (N Y)* **9**, 1090-1095 (1991).
10. Hitzeman, R.A. et al. Expression of a human gene for interferon in yeast. *Nature* **293**, 717-722 (1981).
11. Valenzuela, P., Medina, A., Rutter, W.J., Ammerer, G. & Hall, B.D. Synthesis and assembly of hepatitis B virus surface antigen particles in yeast. *Nature* **298**, 347-350 (1982).
12. Otero, J.M., Panagiotou, G. & Olsson, L. Fueling industrial biotechnology growth with bioethanol. *Adv Biochem Eng Biotechnol.* **108**, 1-40 (2007).
13. Fields, S. & Johnston, M. Whither Model Organism Research? *Science* **307**, 1885-1886 (2005).
14. Oliver, S.G. Functional genomics: lessons from yeast. *Philos Trans R Soc Lond B Biol Sci.* **357**, 17-23 (2002).
15. Castrillo, J.I. & Oliver, S.G. Yeast as a touchstone in post-genomic research: strategies for integrative analysis in functional genomics. *J Biochem Mol Biol.* **37**, 93-106 (2004).
16. Ideker, T., Galitski, T. & Hood, L. A New Approach Decoding Life: Systems Biology. *Annu.Rev.Genomics Hum.Genet.* **2**, 343-372 (2001).
17. Neigeborn, L. & Carlson, M. Genes affecting the regulation of *SUC2* gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* **108**, 845-858 (1984).
18. Carlson, M., Osmond, B.C. & Botstein, D. Mutants of yeast defective in sucrose utilization. *Genetics* **98**, 25-40 (1981).
19. Celenza, J.L. & Carlson, M. A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* **233**, 1175-1180 (1986).
20. Polge, C. & Thomas, M. SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? *Trends in Plant Science* **12**, 20-28 (2007).
21. Hardie, D.G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nat Rev Mol Cell Biol.* **8**, 774-785 (2007).
22. Hardie, D.G. AMP-activated protein kinase as a drug target *Annual Review of Pharmacology and Toxicology* **47**, 185-210 (2007).
23. Amodeo, G.A., Rudolph, M.J. & Tong, L. Crystal structure of the heterotrimer core of *Saccharomyces cerevisiae* AMPK homologue SNF1. *Nature* **449**, 492-495 (2007).
24. Hanks, S.K. & Hunter, T. Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification. *FASEB J.* **9**, 576-596 (1995).
25. McCartney, R.R. & Schmidt, M.C. Regulation of Snf1 kinase. Activation requires phosphorylation of threonine 210 by an upstream kinase as well as a distinct step mediated by the Snf4 subunit. *J.Biol.Chem.* **276**, 36460-36466 (2001).
26. Jiang, R. & Carlson, M. Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Dev.* **10**, 3105-3115 (1996).
27. Scott, J.W. et al. CBS domains form energy-sensing modules whose binding of adenosine ligands is disrupted by disease mutations. *J Clin Invest* **113**, 274-284 (2004).
28. Rudolph, M.J. et al. Structure of the Bateman2 Domain of Yeast Snf4: Dimeric Association and Relevance for AMP Binding. *Structure* **15**, 65-74 (2007).
29. Jiang, R. & Carlson, M. The Snf1 protein kinase and its activating subunit, Snf4, interact with distinct domains of the Sip1/Sip2/Gal83 component in the kinase complex. *Mol.Cell Biol.* **17**, 2099-2106 (1997).
30. Woods, A. et al. Characterization of AMP-activated protein kinase beta and gamma subunits. Assembly of the heterotrimeric complex *in vitro*. *J.Biol.Chem.* **271**, 10282-10290 (1996).

31. Iseli, T.J. et al. AMP-activated protein kinase beta subunit tethers alpha and gamma subunits via its C-terminal sequence (186-270). *J.Biol.Chem.* **280**, 13395-13400 (2005).
32. Elbing, K., Rubenstein, E.M., McCartney, R.R. & Schmidt, M.C. Subunits of the Snf1 kinase heterotrimer show interdependence for association and activity. *J.Biol.Chem.* **281**, 26170-26180 (2006).
33. Vincent, O., Townley, R., Kuchin, S. & Carlson, M. Subcellular localization of the Snf1 kinase is regulated by specific beta subunits and a novel glucose signaling mechanism. *Genes Dev.* **15**, 1104-1114 (2001).
34. Lin, S.S., Manchester, J.K. & Gordon, J.I. Sip2, an N-myristoylated beta subunit of Snf1 kinase, regulates aging in *Saccharomyces cerevisiae* by affecting cellular histone kinase activity, recombination at rDNA loci, and silencing. *J.Biol.Chem.* **278**, 13390-13397 (2003).
35. Hedbacker, K., Townley, R. & Carlson, M. Cyclic AMP-dependent protein kinase regulates the subcellular localization of Snf1-Sip1 protein kinase. *Mol.Cell Biol.* **24**, 1836-1843 (2004).
36. Carling, D. et al. Mammalian AMP-activated protein kinase is homologous to yeast and plant protein kinases involved in the regulation of carbon metabolism. *J.Biol.Chem.* **269**, 11442-11448 (1994).
37. Birk, J.B. & Wojtaszewski, J.F. Predominant alpha2/beta2/gamma3 AMPK activation during exercise in human skeletal muscle. *J Physiol.* **577**, 1021-1032 (2006).
38. Hardie, D.G. Minireview: the AMP-activated protein kinase cascade: the key sensor of cellular energy status. *Endocrinology* **144**, 5179-5183 (2003).
39. Kemp, B.E. et al. AMP-activated protein kinase, super metabolic regulator. *Biochem Soc Trans.* **31**, 162-168 (2003).
40. Hawley, S.A. et al. Characterization of the AMP-activated protein kinase kinase from rat liver and identification of threonine 172 as the major site at which it phosphorylates AMP-activated protein kinase. *J.Biol.Chem.* **271**, 27879-27887 (1996).
41. Carling, D., Clarke, P.R., Zammit, V.A. & Hardie, D.G. Purification and characterization of the AMP-activated protein kinase. Copurification of acetyl-CoA carboxylase kinase and 3-hydroxy-3-methylglutaryl-CoA reductase kinase activities. *Eur J Biochem.* **186**, 129-136 (1989).
42. Hohmann, S. et al. Novel alleles of yeast hexokinase PII with distinct effects on catalytic activity and catabolite repression of *SUC2*. *Microbiology* **145**, 703-714 (1999).
43. Otterstedt, K. et al. Switching the mode of metabolism in the yeast *Saccharomyces cerevisiae*. *EMBO J.* **5**, 532-537 (2004).
44. Ahuatzi, D., Riera, A., Peláez, R., Herrero, P. & Moreno, F. Hxk2 regulates the phosphorylation state of Mig1 and therefore its nucleocytoplasmic distribution. *J.Biol.Chem.* **282**, 4485-4493 (2007).
45. Ludin, K., Jiang, R. & Carlson, M. Glucose-regulated interaction of a regulatory subunit of protein phosphatase 1 with the Snf1 protein kinase in *Saccharomyces cerevisiae*. *Proc.Natl.Acad.Sci.U.S.A* **95**, 6245-6250 (1998).
46. Sanz, P., Alms, G.R., Haystead, T.A. & Carlson, M. Regulatory interactions between the Reg1-Glc7 protein phosphatase and the Snf1 protein kinase. *Mol.Cell Biol.* **20**, 1321-1328 (2000).
47. Wilson, W.A., Hawley, S.A. & Hardie, D.G. Glucose repression/derepression in budding yeast: SNF1 protein kinase is activated by phosphorylation under derepressing conditions, and this correlates with a high AMP:ATP ratio. *Curr.Biol.* **6**, 1426-1434 (1996).
48. Sanz, P. Snf1 protein kinase: a key player in the response to cellular stress in yeast. *Biochem Soc Trans.* **31**, 178-181 (2003).
49. Hong, S.P. & Carlson, M. Regulation of snf1 protein kinase in response to environmental stress. *J.Biol.Chem.* **282**, 16838-16845 (2007).
50. Kuchin, S., Vyas, V.K., Kanter, E., Hong, S.P. & Carlson, M. Std1p (Msn3p) positively regulates the Snf1 kinase in *Saccharomyces cerevisiae*. *Genetics* **163**, 507-514 (2003).
51. Hong, S.P., Leiper, F.C., Woods, A., Carling, D. & Carlson, M. Activation of yeast Snf1 and mammalian AMP-activated protein kinase by upstream kinases. *Proc.Natl.Acad.Sci.U.S.A* **100**, 8839-8843 (2003).
52. Hong, S.P., Momcilovic, M. & Carlson, M. Function of mammalian LKB1 and Ca2+/calmodulin-dependent protein kinase kinase alpha as Snf1-activating kinases in yeast. *J.Biol.Chem.* **280**, 21804-21809 (2005).
53. Woods, A. et al. LKB1 is the upstream kinase in the AMP-activated protein kinase cascade. *Curr.Biol.* **13**, 2004-2008 (2003).
54. Hawley, S.A. et al. Complexes between the LKB1 tumor suppressor, STRAD alpha/beta and MO25 alpha/beta are upstream kinases in the AMP-activated protein kinase cascade. *J Biol.* **2**, 28 (2003).
55. Hawley, S.A. et al. Calmodulin-dependent protein kinase kinase-beta is an alternative upstream kinase for AMP-activated protein kinase. *Cell Metabolism* **2**, 9-19 (2005).
56. Hurley, R.L. et al. The Ca2+/calmodulin-dependent protein kinase kinases are AMP-activated protein kinase kinases. *J.Biol.Chem.* **280**, 29060-29066 (2005).
57. Choi, S.L. et al. The regulation of AMP-activated protein kinase by H(2)O(2). *Biochem Biophys Res Commun.* **287**, 92-97 (2001).
58. Xue, B. & Kahn, B.B. AMPK integrates nutrient and hormonal signals to regulate food intake and energy balance through effects in the hypothalamus and peripheral tissues. *J Physiol.* **574**, 73-83 (2006).

59. Smith, J.L., Patil, P.B. & Fisher, J.S. AICAR and hyperosmotic stress increase insulin-stimulated glucose transport. *J Appl Physiol* **99**, 877-883 (2005).
60. Kahn, B.B., Alquier, T., Carling, D. & Hardie, D.G. AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism *Cell Metabolism* **1**, 15-25 (2005).
61. Sanders, M.J., Grondin, P.O., Hegarty, B.D., Snowden, M.A. & Carling, D. Investigating the mechanism for AMP activation of the AMP-activated protein kinase cascade. *Biochem.J.* **403**, 139-148 (2007).
62. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **1**, D535-D539 (2006).
63. Ptacek, J. et al. Global analysis of protein phosphorylation in yeast. *Nature* **438**, 679-684 (2005).
64. Gancedo, J.M. Yeast carbon catabolite repression. *Microbiol.Mol.Biol.Rev.* **62**, 334-361 (1998).
65. Carlson, M. Glucose repression in yeast. *Curr.Opin.Microbiol.* **2**, 202-207 (1999).
66. Young, E.T., Dombek, K.M., Tachibana, C. & Ideker, T. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J.Biol.Chem.* **278**, 26146-26158 (2003).
67. Kuchin, S., Treich, I. & Carlson, M. A regulatory shortcut between the Snf1 protein kinase and RNA polymerase II holoenzyme. *Proc.Natl.Acad.Sci.U.S.A* **97**, 7916-7920 (2000).
68. Lo, W.S. et al. Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science* **293**, 1142-1146 (2001).
69. Lo, W.S. et al. Histone H3 phosphorylation can promote TBP recruitment through distinct promoter-specific mechanisms. *EMBO J.* **24**, 997-1008 (2005).
70. Shirra, M.K. et al. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the *INO1* gene in a *snf1* mutant strain of *Saccharomyces cerevisiae*. *Mol.Cell Biol.* **21**, 5710-5722 (2001).
71. Woods, A. et al. Yeast SNF1 is functionally related to mammalian AMP-activated protein kinase and regulates acetyl-CoA carboxylase *in vivo*. *J.Biol.Chem.* **269**, 19509-19515 (1994).
72. Hardy, T.A., Huang, D. & Roach, P.J. Interactions between cAMP-dependent and SNF1 protein kinases in the control of glycogen accumulation in *Saccharomyces cerevisiae*. *J.Biol.Chem.* **269**, 27907-27913 (1994).
73. Aon, M.A. & Cortassa, S. Catabolite repression mutants of *Saccharomyces cerevisiae* show altered fermentative metabolism as well as cell cycle behavior in glucose-limited chemostat cultures. *Biotechnol.Bioeng.* **59**, 203-213 (1998).
74. Palecek, S.P., Parikh, A.S., Huh, J.H. & Kron, S.J. Depression of *Saccharomyces cerevisiae* invasive growth on non-glucose carbon sources requires the Snf1 kinase. *Mol.Microbiol.* **45**, 453-469 (2002).
75. Kuchin, S., Vyas, V.K. & Carlson, M. Snf1 protein kinase and the repressors Nrg1 and Nrg2 regulate *FLO11*, haploid invasive growth, and diploid pseudohyphal differentiation. *Mol.Cell Biol.* **22**, 3994-4000 (2002).
76. Wiatrowski, H.A. & Carlson, M. Yap1 Accumulates in the Nucleus in Response to Carbon Stress in *Saccharomyces cerevisiae*. *Eukaryot.Cell* **2**, 19-26 (2003).
77. Ashrafi, K., Lin, S.S., Manchester, J.K. & Gordon, J.I. Sip2p and its partner snf1p kinase affect aging in *S. cerevisiae*. *Genes Dev.* **14**, 1872-1885 (2000).
78. Hardie, D.G. & Carling, D. The AMP-activated protein kinase--fuel gauge of the mammalian cell? *Eur J Biochem.* **246**, 259-273 (1997).
79. Schuller, H.J. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. *Curr.Genet.* **43**, 139-160 (2003).
80. Rolland, F., Winderickx, J. & Thevelein, J.M. Glucose-sensing and -signalling mechanisms in yeast. *FEMS Yeast Res.* **2**, 183-201 (2002).
81. Hedbacker, K., Hong, S.P. & Carlson, M. Pak1 protein kinase regulates activation and nuclear localization of Snf1-Gal83 protein kinase. *Mol.Cell Biol.* **24**, 8255-8263 (2004).
82. Vincent, O. & Carlson, M. Sip4, a Snf1 kinase-dependent transcriptional activator, binds to the carbon source-responsive element of gluconeogenic genes. *EMBO J.* **17**, 7002-7008 (1998).
83. Papamichos-Chronakis, M., Gligoris, T. & Tzamarias, D. The Snf1 kinase controls glucose repression in yeast by modulating interactions between the Mig1 repressor and the Cyc8-Tup1 co-repressor. *EMBO Rep.* **5**, 368-372 (2004).
84. Charbon, G., Breunig, K.D., Wattiez, R., Vandenhaute, J. & Noël-Georis, I. Key role of Ser562/661 in Snf1-dependent regulation of Cat8p in *Saccharomyces cerevisiae* and *Kluyveromyces lactis*. *Mol.Cell Biol.* **24**, 4083-4091 (2004).
85. Schmidt, M.C. & McCartney, R.R. beta-subunits of Snf1 kinase are required for kinase function and substrate definition. *EMBO J.* **19**, 4936-4943 (2000).
86. Hedges, D., Proft, M. & Entian, K.D. *CAT8*, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae*. *Mol.Cell Biol.* **15**, 1915-1922 (1995).
87. Roth, S., Kumme, J. & Schüller, H.J. Transcriptional activators Cat8 and Sip4 discriminate between sequence variants of the carbon source-responsive promoter element in the yeast *Saccharomyces cerevisiae*. *Curr.Genet.* **45**, 121-128 (2004).

88. Kaniak, A., Xue, Z., Macool, D., Kim, J.H. & Johnston, M. Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. *Eukaryot.Cell* **3**, 221-231 (2004).
89. Rubio-Texeira, M. A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. *FEMS Yeast Res.* **5**, 1115-1128 (2005).
90. Young, E.T., Kacherovsky, N. & Van, R.K. Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation. *J.Biol.Chem.* **277**, 38095-38103 (2002).
91. Rottensteiner, H. et al. *Saccharomyces cerevisiae PIP2* mediating oleic acid induction and peroxisome proliferation is regulated by Adr1p and Pip2p-Oaf1p. *J.Biol.Chem.* **278**, 27605-27611 (2003).
92. Grauslund, M., Lopes, J.M. & Ronnow, B. Expression of *GUT1*, which encodes glycerol kinase in *Saccharomyces cerevisiae*, is controlled by the positive regulators Adr1p, Ino2p and Ino4p and the negative regulator Opi1p in a carbon source-dependent fashion. *Nucleic Acids Res.* **27**, 4391-4398 (1999).
93. Alepuz, P.M., Matheos, D., Cunningham, K.W. & Estruch, F. The *Saccharomyces cerevisiae* RanGTP-binding protein msn5p is involved in different signal transduction pathways. *Genetics* **153**, 1219-1231 (1999).
94. Hahn, J.S. & Thiele, D.J. Activation of the *Saccharomyces cerevisiae* Heat Shock Transcription Factor Under Glucose Starvation Conditions by Snf1 Protein Kinase. *J.Biol.Chem.* **279**, 5169–5176 (2004).
95. Mayordomo, I., Estruch, F. & Sanz, P. Convergence of the target of rapamycin and the Snf1 protein kinase pathways in the regulation of the subcellular localization of Msn2, a transcriptional activator of STRE (Stress Response Element)-regulated genes. *J.Biol.Chem.* **277**, 35650-35656 (2002).
96. Jamieson, D.J. Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. *Yeast* **14**, 1511-1527 (1998).
97. Orlova, M., Kanter, E., Krakovich, D. & Kuchin, S. Nitrogen availability and TOR regulate the Snf1 protein kinase in *Saccharomyces cerevisiae*. *Eukaryot.Cell* **5**, 1831-1837 (2006).
98. Dubacq, C., Chevalier, A. & Mann, C. The protein kinase Snf1 is required for tolerance to the ribonucleotide reductase inhibitor hydroxyurea. *Mol.Cell Biol.* **24**, 2560-2572 (2004).
99. Portillo, F., Mulet, J.M. & Serrano, R. A role for the non-phosphorylated form of yeast Snf1: tolerance to toxic cations and activation of potassium transport. *FEBS Letters* **579**, 512-516 (2005).
100. Gimeno, C.J., Ljungdahl, P.O., Styles, C.A. & Fink, G.R. Unipolar cell divisions in the yeast *S. cerevisiae* lead to filamentous growth: regulation by starvation and RAS. *Cell* **68**, 1077-1090 (1992).
101. Schmelzle, T. & Hall, M.N. TOR, a central controller of cell growth. *Cell* **103**, 253-262 (2000).
102. Shirra, M.K., Rogers, S.E., Alexander, D.E. & Arndt, K.M. The Snf1 protein kinase and Sit4 protein phosphatase have opposing functions in regulating TATA-binding protein association with the *Saccharomyces cerevisiae INO1* promoter. *Genetics* **169**, 1957-1972 (2005).
103. Bertram, P.G. et al. Convergence of TOR-nitrogen and Snf1-glucose signaling pathways onto Gln3. *Mol.Cell Biol.* **22**, 1246-1252 (2002).
104. Minokoshi, Y. et al. AMP-kinase regulates food intake by responding to hormonal and nutrient signals in the hypothalamus. *Nature* **428**, 569-574 (2004).
105. Cota, D. et al. Hypothalamic mTOR signaling regulates food intake. *Science* **312**, 927-930 (2006).
106. Inoki, K., Zhu, T. & Guan, K.L. TSC2 mediates cellular energy response to control cell growth and survival. *Cell* **115**, 577-590 (2003).
107. Halford, N.G. et al. Metabolic signalling and carbon partitioning: role of Snf1-related (SnRK1) protein kinase. *J Exp Bot.* **54**, 467-475 (2003).
108. Carling, D., Zammit, V.A. & Hardie, D.G. A common bicyclic protein kinase cascade inactivates the regulatory enzymes of fatty acid and cholesterol biosynthesis. *FEBS Letters* **223**, 217-222 (1987).
109. Fryer, L.G. et al. Characterization of the role of the AMP-activated protein kinase in the stimulation of glucose transport in skeletal muscle cells. *Biochem.J.* **363**, 167-174 (2002).
110. Yamauchi, T. et al. Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase. *Nat Med.* **8**, 1288-1295 (2002).
111. Minokoshi, Y. et al. Leptin stimulates fatty-acid oxidation by activating AMP-activated protein kinase. *Nature* **415**, 339-343 (2002).
112. da Silva Xavier, G. et al. Role for AMP-activated protein kinase in glucose-stimulated insulin secretion and preproinsulin gene expression. *Biochem.J.* **371**, 761-774 (2003).
113. Towler, M.C. & Hardie, D.G. AMP-activated protein kinase in metabolic control and insulin signaling. *Circ Res* **100**, 328-341 (2007).
114. Kubota, N. et al. Adiponectin stimulates AMP-activated protein kinase in the hypothalamus and increases food intake. *Cell Metabolism* **6**, 55-68 (2007).
115. Carling, D. The AMP-activated protein kinase cascade--a unifying system for energy control. *Trends Biochem Sci* **29**, 18-24 (2004).
116. Arad, M., Seidman, C.E. & Seidman, J.G. AMP-activated protein kinase in the heart: role during health and disease. *Circ Res* **100**, 474-488 (2007).

117. Shaw, R.J. Glucose metabolism and cancer. *Curr Opin Cell Biol.* **18**, 598-608 (2006).
118. François, J. & Parrou, J.L. Reserve carbohydrates metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol.Rev.* **25**, 125-145 (2001).
119. Fleischmann, R.D. et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae Rd*. *Science* **269**, 496-512 (1995).
120. Goffeau, A. et al. Life with 6000 Genes. *Science* **274**, 546, 563-567 (1996).
121. Joyce, A.R. & Palsson, B.O. The model organism as a system: integrating 'omics' data sets. *Nat Rev Mol Cell Biol.* **7**, 198-210 (2006).
122. Bruggeman, F.J. & Westerhoff, H.V. The nature of systems biology. *Trends Microbiol.* **15**, 45-50 (2007).
123. Sauer, U., Heinemann, M. & Zamboni, N. Getting Closer to the Whole Picture. *Science* **316**, 550-551 (2007).
124. Harbison, C.T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).
125. Hodges, P.E., McKee, A.H., Davis, B.P., Payne, W.E. & Garrels, J.I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* **27**, 69-73 (1999).
126. Forster, J., Famili, I., Fu, P., Palsson, B.O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244-253 (2003).
127. Kitano, H. International alliances for quantitative modeling in systems biology. *Mol.Syst.Biol.* **1**, 1-2 (2005).
128. Schena, M., Shalon, D., Davis, R.W. & Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470 (1995).
129. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. & Lockhart, D.J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat.Biotechnol.* **15**, 1359-1367 (1997).
130. Fodor, S.P. et al. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767-773 (1991).
131. Hughes, T.R. et al. Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol.* **19**, 342-347 (2001).
132. Novick, A. & Szilard, L. Description of the chemostat. *Science* **112**, 715-716 (1950).
133. Weusthuis, R.A., Pronk, J.T., van den Broek, P.J. & van Dijken, J.P. Chemostat cultivation as a tool for studies on sugar transport in yeasts. *Microbiol.Rev.* **58**, 616-630 (1994).
134. Washburn, M.P. et al. Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc.Natl.Acad.Sci.U.S.A* **100**, 3107–3112 (2003).
135. Gygi, S.P., Rochon, Y., Robert Franza, B. & Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol.Cell Biol.* **19**, 1720-1730 (1999).
136. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol.* **25**, 17-124 (2007).
137. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. *Nature* **425**, 737-741 (2003).
138. Huh, W.K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-691 (2003).
139. Belle, A., Tanay, A., Bitincka, L., Shamir, R. & O'Shea, E.K. Quantification of protein half-lives in the budding yeast proteome. *Proc.Natl.Acad.Sci.U.S.A* **103**, 13004-13009 (2006).
140. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
141. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* **19**, 242-247 (2001).
142. Delahunty, C. & Yates, J.R., 3rd. Protein identifcation using 2D-LC-MS/MS. *Methods* **35**, 248–255 (2005).
143. Gygi, S.P. et al. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat.Biotechnol.* **17**, 994-999 (1999).
144. Fang, R. et al. Differential label-free quantitative proteomic analysis of *Shewanella oneidensis* cultured under aerobic and suboxic conditions by accurate mass and time tag approach. *Mol.Cell Proteomics* **5**, 714-725 (2006).
145. Cantin, G.T., Venable, J.D., Cociorva, D. & Yates, J.R.r. Quantitative phosphoproteomic analysis of the tumor necrosis factor pathway. *J Proteome Res* **5**, 127-134 (2006).
146. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat.Biotechnol.* **19**, 242-247 (2001).
147. Eng, J., McCormack, A. & Yates, J.R., 3rd. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J Am Soc Mass Spectrom* **5**, 976-989 (1994).
148. Tabb, D.L., McDonald, W.H. & Yates, J.R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J Proteome Res* **1**, 21-26 (2002).

149. Venable, J.D., Wohlschlegel, J., McClatchy, D.B., Park, S.K. & Yates, J.R., 3rd. Relative quantification of stable isotope labeled peptides using a linear ion trap-orbitrap hybrid mass spectrometer. *Anal.Chem.* **79**, 3056-3064 (2007).
150. Patterson, S.D. & Aebersold, R.H. Proteomics: the first decade and beyond. *Nat.Genetics* **33**, 311-323 (2003).
151. Zhu, H. et al. Global analysis of protein activities using proteome chips. *Science* **293**, 2101-2105 (2001).
152. Michaud, G.A. et al. Analyzing antibody specificity with whole proteome microarrays. *Nat Biotechnol.* **21**, 1509-1512 (2003).
153. Nielsen, J. & Oliver, S. The next wave in metabolome analysis. *Trends Biotechnol.* **23**, 544-546 (2005).
154. Smedsgaard, J. & Nielsen, J. Metabolite profiling of fungi and yeast: from phenotype to metabolome by MS and informatics. *J Exp Bot.* **56**, 273-286 (2005).
155. Dunn, W.B., Bailey, N.J. & Johnson, H.E. Measuring the metabolome: current analytical technologies. *Analyst* **130**, 606-625 (2005).
156. Alberts, B. The Cell as a Collection of Protein Machines: Preparing the Next Generation of Molecular Biologists. *Cell* **92**, 291-294 (1998).
157. Gavin, A.C. et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141-147 (2002).
158. Ho, Y. et al. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180-183 (2002).
159. Uetz, P. et al. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623-627 (2000).
160. Ito, T. et al. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc.Natl.Acad.Sci.U.S.A* **98**, 4569-4574 (2001).
161. von Mering, C. et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399-403 (2002).
162. Krogan, N.J. et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637-643 (2006).
163. von Mering, C. et al. STRING: known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **33**, D433-D437 (2005).
164. Buck, M.J. & Lieb;J.D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349-360 (2004).
165. Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc.Natl.Acad.Sci.U.S.A* **103**, 14724-14731 (2006).
166. Cokus, S., Rose, S., Haynor, D., Gronbech-Jensen, N. & Pellegrini, M. Modelling the network of cell cycle transcription factors in the yeast *Saccharomyces cerevisiae*. *BMC Bioinformatics* **7**, 381 (2006).
167. Herrgård, M.J., Lee, B., Portnoy, V. & Palsson, B.O. Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res.* **16**, 627-635 (2006).
168. Oliveira, A.P., Patil, K.R. & Nielsen, J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2**, 17 (2008).
169. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc.Natl.Acad.Sci.U.S.A* **98**, 31-36 (2001).
170. Suthram, S., Shlomi, T., Ruppin, E., Sharan, R. & Ideker, T. A direct comparison of protein interaction confidence assignment schemes. *BMC Bioinformatics* **7** (2006).
171. Deane, C.M., Salwinski, L., Xenarios, I. & Eisenberg, D. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol.Cell Proteomics* **1**, 349-356 (2002).
172. Cheadle, C. et al. A rapid method for microarray cross platform comparisons using gene expression signatures. *Mol Cell Probes* **21**, 35-46 (2007).
173. Severgnini, M. et al. Strategies for comparing gene expression profiles from different microarray platforms: application to a case-control experiment. *Anal.Biochem.* **353**, 43-56 (2006).
174. Stec, J. et al. Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. *J Mol Diagn.* **7**, 357-367 (2005).
175. Alon, U. et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays *Proc.Natl.Acad.Sci.U.S.A* **96**, 6745-6750 (1999).
176. Grotkjaer, T., Winther, O., Regenberg, B., Nielsen, J. & Hansen, L.K. Robust multi-scale clustering of large DNA microarray datasets with the consensus algorithm. *Bioinformatics.* **22**, 58-67 (2006).
177. van Helden, J., Andre, B. & Collado-Vides, J. A web site for the computational analysis of yeast regulatory sequences. *Yeast* **16**, 177-187 (2000).
178. Roth, F.P., Hughes, J.D., Estep, P.W. & Church, G.M. Finding DNA regulatory motifs within unaligned noncoding sequences clustered by whole-genome mRNA quantitation. *Nat Biotechnol.* **16**, 939-945 (1998).

179. Beyer, A. et al. Integrated assessment and prediction of transcription factor binding. *PLoS.Comput Biol* **2**, 615-626 (2006).
180. Sharan, R. et al. Conserved patterns of protein interaction in multiple species. *Proc.Natl.Acad.Sci.U.S.A* **102**, 1974-1979 (2005).
181. Ideker, T. et al. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929-934 (2001).
182. Patil, K.R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc.Natl.Acad.Sci.U.S.A* **102**, 2685-2689 (2005).
183. de Lichtenberg, U., Jensen, L.J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724-727 (2005).
184. Ishii, N. et al. Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations. *Science* **316**, 593-597 (2007).
185. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**, S233-S240 (2002).
186. Reed, J.L., Famili, I., Thiele, I. & Palsson, B.O. Towards multidimensional genome annotation. *Nat.Rev.Genet.* **7**, 130-141 (2006).
187. Price, N.D., Reed, J.L. & Palsson, B.Ø. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat.Rev.Microbiol.* **2**, 886-897 (2004).
188. Usaite, R., Patil, K.R., Grotkjær, T., Nielsen, J. & Regenberg, B. Global Transcriptional and Physiological Responses of *Saccharomyces cerevisiae* to Ammonium, L-Alanine, or L-Glutamine Limitation. *Appl.Microbiol.Biotechnol.* **72**, 6194–6203 (2006).
189. Mustacchi, R., Hohmann, S. & Nielsen, J. Yeast systems biology to unravel the network of life. *Yeast* **23**, 227-238 (2006).
190. Ippolito, J.E. et al. An integrated functional genomics and metabolomics approach for defining poor prognosis in human neuroendocrine cancers. *Proc.Natl.Acad.Sci.U.S.A* **102**, 9901-9906 (2005).
191. Anderson, N.L. & Anderson, N.G. The human plasma proteome: history, character, and diagnostic prospects. *Mol.Cell Proteomics* **1**, 845-867 (2002).
192. Villanueva, J. et al. Differential exoprotease activities confer tumor-specific serum peptidome patterns. *J Clin Invest* **116**, 271-284 (2006).
193. Hood, L. & Perlmutter, R.M. The impact of systems approaches on biological problems in drug discovery. *Nat Biotechnol.* **22**, 1215-1217 (2004).

Chapter 2:

Physiological characterization of glucose repression in the strains with *SNF1* and *SNF4* genes deleted

Renata Usaite, Jens Nielsen and Lisbeth Olsson

J Biotechnol. 2008 Jan; 133(1):73-81

Available online at www.sciencedirect.com
ScienceDirect
Journal of Biotechnology 133 (2008) 73–81
Journal of BIOTECHNOLOGY
www.elsevier.com/locate/jbiotec

Physiological characterization of glucose repression in the strains with *SNF1* and *SNF4* genes deleted

Renata Usaite, Jens Nielsen, Lisbeth Olsson *

Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Building 223, DK-2800 Kgs. Lyngby, Denmark

Received 14 March 2007; received in revised form 28 August 2007; accepted 10 September 2007

Abstract

We investigated the effect of Snf1 kinase and its regulatory subunit Snf4 on the regulation of glucose and galactose metabolism in the yeast *Saccharomyces cerevisiae* by physiologically characterizing Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* in CEN.PK background in glucose and glucose–galactose-mixture batch cultivations. The main result of this study showed that delayed induction of galactose catabolism was *SNF1* or *SNF4* gene deletion specific. In comparison to the reference strain, growth delay on galactose was found to last 2.4 times (7 h), 3.1 times (10.5 h) and 9.6 times (43 h) longer for the Δ*snf4*, Δ*snf1* and Δ*snf1*Δ*snf4* strains, respectively. The maximum specific growth rates on galactose were determined to be two to three times lower for the recombinant strains compared to the reference strain (0.13 h^{-1}) and were found to be 0.07, 0.08 and 0.04 h^{-1} for the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains, respectively. The study showed that Snf1 kinase was not solely responsible for the derepression of galactose metabolism.

Keywords: Budding yeast; Glucose repression; Galactose induction

1. Introduction

Glucose (or catabolite) repression has been identified in many microorganisms, and it has been very well documented and investigated in *Saccharomyces cerevisiae* (Gancedo, 1998; Rolland et al., 2002). Today this phenomenon, as well as co-consumption of different carbon sources remains to be of interest among a lot of research groups working in fundamental science and for industrial applications of yeast. For example, the industrially used carbon sources often consist of sugar mixtures, and due to glucose repression these sugars are utilized sequentially, resulting in prolonged production time (Olsson and Nielsen, 2000). Industrially relevant lactose, molasses and lignocelluloses contain galactose and the induction of galactose catabolism through metabolic engineering strategies and construction of glucose-derepressed strains may improve cell growth and production of desired compounds (Kim et al., 2004; Ostergaard et al., 2000, 2001).

In the presence of a rapidly fermentable carbon source, such as glucose, the transcription of genes, whose products are essential for catabolism of slowly fermentable or completely non-fermentable carbon sources, is repressed (Gancedo, 1998; Rolland et al., 2002). In the absence of glucose, *S. cerevisiae* metabolizes alternative carbon sources. Typically, the transcription of genes that encode enzymes required for metabolizing such alternative carbon sources are induced by those specific carbon sources. For example, the *GAL* genes that encode the enzymes of the Leloir pathway are repressed by glucose and induced by galactose (Nehlin et al., 1991; Rubio-Texeira, 2005).

Glucose repression involves several different signal transduction pathways. The main glucose repression pathway involves the Snf1 kinase complex (Carlson et al., 1981), which under glucose limitation inactivates the transcription repressor Mig1 and hereby prevents Mig1 from interacting with the transcriptional co-repressors Cyc8–Tup1. This leads to derepression of genes involved in the metabolism of alternative carbon sources (Treitel et al., 1998). Snf1 is a serine–threonine protein kinase that, in the glucose repression pathway, is active in a complex containing the regulatory subunit Snf4, together with the Gal83 protein. Gal83 is required for Snf1 nucleus localization, where

* Corresponding author. Tel.: +45 45252677; fax: +45 45884148.
E-mail address: lo@biocentrum.dtu.dk (L. Olsson).

0168-1656/$ – see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jbiotec.2007.09.001

Snf1 may inactivate Mig1 or stimulate transcription activators such as Cat8 (Rahner et al., 1999; Treitel et al., 1998).

The glucose induction pathway (through sensors Snf3 and Rgt2 and Grr1) regulates transcription factor Rgt1, which represses or activates *HXT* expression in response to the availability of extracellular glucose (Ozcan and Johnston, 1999). Kaniak et al. (2004) have shown that there is a complex cross-communication between the glucose repression and glucose induction pathways. Mig1 (and Mig2) represses metabolic genes (like *HXT2*, *HXT4* and *SUC2*) and regulatory genes (such as *MIG1* and *SNF3*), intertwining these two glucose signal transduction pathways.

Galactose utilization and regulation of *GAL* genes have been extensively reviewed (Bhat and Murthy, 2001; Rubio-Texeira, 2005). The induction of the *GAL* genes is mediated by the interplay between three regulatory proteins: transcriptional activator Gal4; transcriptional repressor Gal80; and substrate (galactose) sensor/inducer Gal3 (Rubio-Texeira, 2005). It has recently been shown that in glucose non-repressive growth conditions, Gal3 is active only when it binds both of its allosteric effectors galactose and ATP (Bhat and Murthy, 2001). As *GAL* regulon induction continues, bifunctional Gal1 can replace Gal3 and maintain the induced state of *GAL* genes in the presence of galactose (Thoden et al., 2005).

When glucose is depleted, regulation of carbon metabolism changes and growth on other available carbon sources is induced. The glucose repression signaling pathway takes a role in this shift. It has been shown previously that Δ*snf1* and Δ*snf4* strains are unable to grow on non-fermentable carbon sources such as ethanol, glycerol and acetate due to the repression of gluconeogenesis (Schuller and Entian, 1987). The expression of galactose metabolizing genes are under the competitive regulation of the transcription repressor Mig1 and the transcription activator Gal4 (Nehlin et al., 1991). Previous molecular biology studies described that the Δ*snf1* and Δ*snf4* strains had poor or no growth on galactose (Carlson et al., 1981; Kuchin et al., 2003; Neigeborn and Carlson, 1984; Palecek et al., 2002; Schmidt and McCartney, 2000; Van Driessche et al., 2005; Zhou and Winston, 2001). It has also been found that not all yeast species are able to grow on galactose indicating that the regulation of galactose metabolism might be yeast-strain dependent.

The overall objective of this study was to perform a physiological profiling of the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains in glucose and glucose–galactose batch cultivations and investigate how, in comparison to the reference strain, the disruptions of the Snf1 kinase complex affected the regulation of carbon metabolism and, in particular, the galactose metabolism under dynamic growth conditions. The *S. cerevisiae* CEN.PK genotypic background reference and three recombinant Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains were used in this study. The physiological response of the recombinant strains to glucose and galactose substrate availability was systematically evaluated by: (i) performing glucose and glucose–galactose-mixture batch cultivations, (ii) calculating specific substrate uptake, product production rates and yields of extracellular metabolites and (iii) comparing the results to a physiological profile of the reference strain.

2. Materials and methods

2.1. Yeast strains

The *S. cerevisiae* strains used in this study were CEN.PK 113-7D (*MAT***a** *MAL2-8c SUC2*), a prototrophic strain from P. Kötter (Frankfurt, Germany) (Van Dijken et al., 2000) and its derivatives. The only genotypic difference among strains used is summarized in Table 1.

2.2. Construction of the yeast strain IBT100072

The strain IBT100072 with double *SNF1* and *SNF4* gene deletion was created using a two-step gene deletion strategy, sporulation and dissection of tetrads. The deletion of *SNF4* (*snf4* (4-966)::loxP-Kan-loxP) was introduced into IBT100072 from the CEN.PK 507-5B strain (Table 1). To have two distinct gene-deletion verification strategies (G148 plating assay for the deletion of *SNF4* and PCR for the deletion of *SNF1*) in the IBT10072 strain, *SNF1* was deleted without using loxP-Kan-loxP insertion. First, four PCR fragments were created to delete *SNF1* by using the bipartite gene-targeting technique (Reid et al., 2002). The upstream and downstream regions of *SNF1* were amplified using the genomic DNA of CEN.PK 113-5D strain as a template and two primer pairs: SNF1_A (5′-GCT ATC AAA TGC TGA ACC TTC C-3′) and SNF1_B (5′-GCA GGG ATG CGG CCG CTG ACA GGG AGT GTA GCA AAA CTT GTT AC-3′); SNF1_C (5′-CCG CTG CTA GGC GCG CCG TGG GTG GAA CGT AAA AGA ATG ATA TGG-3′) and SNF1_D (5′-TGT TCT GGC AGC ATG ATT TG-3′). *Kluyveromyces lactis URA3* was amplified as two overlapping fragments (referred to as upstream and downstream) using the plasmid pWJ1042 (Reid et al., 2002) as a template and the primer pairs: dKL5′ (5′-GTC AGC GGC CGC ATC CCT GCT TCG GCT TCA TGG CAA TTC CCG-3′) and 3′-int (5′-GAG CAA TGA ACC CAA TAA CGA AAT C-3′); 5′-int (5′-CTT GAC GTT CGT TCG ACT GAT GAG C-3′) and cKL3′ (5′-CAC GGC GCG CCT AGC AGC GGT AAC GCC AGG GTT TTC CCA GTC AC-3′). The amplified *SNF1* upstream region fragment was fused to the upstream *K. lactis URA3* fragment, and the *SNF1* downstream region fragment was fused to the downstream *K. lactis URA3* fragment. These two fusion fragments were used as transformation material for the CEN.PK 113-5D strain. The transformation was performed as previously described by Gietz and Woods (Gietz and Woods, 2002). A total of 10 Ura^+ transformants (grown on synthetic complete (SC) Ura^- medium) were streak-purified and re-streaked on SC media plates containing 5-fluoroorotic acid (5-FOA) (ZymoResearch) to selectively loop out the *K. lactis URA3* gene (Boeke et al., 1984). The 5-FOA resistant colonies were picked out and checked for loss of *K. lactis URA3* by replica plating on SC Ura^- medium. Finally, the deletion of *SNF1* was confirmed by PCR using SNF1_A and SNF1_D primers.

In order to create a MATα His^- *snf4* strain, the CEN.PK 507-5B strain was crossed with CEN.PK 110-10C on Esposito supplemented (SPO) medium plates and left to sporulate for 4 days. Each of 10 tetrads was divided into 4 spores under dissection microscope (Nikon eclipse *50i*). The strain was confirmed

Table 1
Yeast strains used in this study

Strain[a]	Name in text	Genotype	Source
CEN.PK 113-7D	Reference strain	*MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8^C*	Provided by P. Kötter[b]
CEN.PK 113-5D		*MATa HIS3 TRP1 LEU2 SUC2 MAL2-8^C ura* (3–52)::loxP-Kan-loxP	Provided by P. Kötter
CEN.PK 110-10C		*MATα URA3 TRP1 LEU2 SUC2 MAL2-8^C his3-delta1*	Provided by P. Kötter
CEN.PK 506-1C	Δ*snf1*	*MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8^C snf1* (4-1899)::loxP-Kan-loxP	Provided by P. Kötter
CEN.PK 507-5B	Δ*snf4*	*MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8^C snf4* (4-966)::loxP-Kan-loxP	Provided by P. Kötter
IBT100072	Δ*snf1*Δ*snf4*	*MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8^C snf1* (1-1903) *snf4* (4-966)::loxP-Kan-loxP[c]	This study

[a] All the strains of this study are derived from the same parental CEN.PK background strain (Van Dijken et al., 2000).
[b] Institut für Mikrobiologie, Frankfurt, Germany.
[c] The deletions of *SNF1* are different in IBT00072 and CEN.PK 506-1C strains (explanation in Section 2).

by selective replica plating on SC His$^-$ medium, and on G418 antibiotic containing YPD medium. Finally, the transformant (MATa Ura$^-$ *snf1*) was crossed with the MATα His$^-$ *snf4* strain and left to sporulate for 4 days on Esposito supplemented (SPO) medium. Each tetrad was divided into 4 spores under dissection microscope (Nikon eclipse *50i*); the IBT100072 strain was characterized by PCR, selective plating on SC Ura$^-$ medium, SC His$^-$ medium, SC Ura$^-$, His$^-$ medium and also on G418 containing YPD medium.

2.3. Inoculum and pre-culture

The strains were stored at −80 °C in Eppendorf tubes containing YPD medium (Rose et al., 1990) with 20% (v/v) glycerol. The yeast suspension was thawed and streaked out onto YPD agar plates. A single colony was picked from these plates and used for inoculation of a shake flask.

2.4. Shake-flask cultivations

For the pre-cultures, a minimal medium based on Verduyn et al. (1992) was used. The composition of contents per liter was: $(NH_4)_2SO_4$, 7.50 g; KH_2PO_4, 14.40 g; $MgSO_4 \times 7H_2O$, 0.50 g; D-glucose, 10.0 g; Antifoam 289 (A-5551, Sigma–Aldrich), 0.050 ml; trace metals, 2.0 ml (composition given below); vitamins, 1.0 ml (composition given below). The trace metal solution contained per liter: EDTA (Titriplex III®), 15.0 g; $ZnSO_4 \times 7H_2O$, 4.5 g; $MnCl_2 \times 2H_2O$, 0.82 g; $CoCl_2 \times 6H_2O$, 0.3 g; $CuSO_4 \times 5H_2O$, 0.3 g; $Na_2MoO_4 \times 2H_2O$, 0.4 g; $CaCl_2 \times 2H_2O$, 4.5 g; $FeSO_4 \times 7H_2O$, 3.0 g; H_3BO_3, 1.0 g; KI, 0.10 g. The vitamin solution contained per liter: biotin, 0.05 g; *p*-benzoic acid, 0.20 g; nicotinic acid, 1.00 g; Ca-pantothenate, 1.00 g; pyridoxin HCl, 1.00 g; thiamin HCl, 1.00 g; *myo*-inositol, 25.00 g. The pH was adjusted to 6.5 via drop-wise addition of 2 M NaOH. The carbon substrate was autoclaved separately. Thereafter the carbon source and sterile filtered vitamins were aseptically mixed into the final formulation. Two-baffled shake flasks (500 ml) with 100 ml of medium were inoculated and incubated for 24 h at 30 °C in an orbital shaker at 150 rpm until the cell mass concentration (dry weight) reached 1–1.5 g l^{-1}. Pre-cultures from shake-flasks with 20 g l^{-1} of D-glucose were used for inoculation of glucose and glucose–galactose-mixture batch cultivations. The bioreactors were inoculated to a final biomass (dry weight) concentration of 1 mg l^{-1}.

2.5. Batch cultivations

Cultivations were carried out in well-controlled, four-baffled, 5 l in-house manufactured bioreactors with a working volume and temperature of 4 l and 30 °C, respectively. The bioreactors were equipped with two disk-turbine impellers rotating at 600 rpm. The final formulated medium contained per liter: $(NH_4)_2SO_4$, 15.0 g; KH_2PO_4, 3.0 g; $MgSO_4 \times 7H_2O$, 1.50 g; Antifoam 289 (A-5551, Sigma–Aldrich), 0.050 ml; trace metals, 3 ml (composition given above); vitamins, 3 ml (composition given above). A total of 30.0 g l^{-1} of D-glucose was included for the glucose batch; 15.0 g l^{-1} of D-glucose and 15.0 g l^{-1} of D-galactose were used for the glucose–galactose-mixture batch cultivations. The pH was maintained at 5.0 by automatic addition of 2 M NaOH. The air flow was controlled at 4 l per min (1 vvm), and the off gas was directed through a condenser chilled to 4 °C to minimize evaporation of ethanol from the bioreactor.

2.6. Tests to detect suppressor mutations arising under selective pressures of galactose

The following experiments were performed to ensure that the observed and described phenotypes were not caused by suppressor mutations generated in this study. Never subjected to galactose selection before, the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains were plated in dilute suspensions on YP medium agar plates (Rose et al., 1990) containing 20 g l^{-1} D-glucose or 20 g l^{-1} D-galactose as the only carbon source. The same number of colonies appearing on both carbon sources indicated that no additional mutations were required for growth on galactose. To determine if any suppression mutations happened throughout extensive glucose–galactose-mixture batch cultivations, samples of the Δ*snf1*, Δ*snf4* or Δ*snf1*Δ*snf4* strains were collected from the end point of these cultivations and inoculated into shake-flasks with 2% of ethanol, 2% of acetate or 2% of galactose as a sole carbon source. Reference CEN.PK 113-7D and never subjected to galactose selection Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains were used in the control experiments. None of the recombinant strains showed any differences in growth phenotypes before and after batch cultivation: they did not grow on non-fermentable carbon sources, but grew poorly on galactose. Reference CEN.PK 113-7D strain grew on all carbon sources tested.

2.7. Biomass determination

The dry weight was determined by using 0.45 μm pore size nitrocellulose filters (Supor®-450 Membrane Filters, PALL Life Sciences, Ann Arbor, MI, USA), as previously described by Ostergaard et al. (2001). The OD was determined at 600 nm by using a Hitachi model U-1100 spectrophotometer. Samples from the cultivation broth were diluted with water to maintain an OD measurement range of 0.1–0.3. The dry weight measurements were used for determination of maximum specific growth rates.

2.8. Analysis of extracellular metabolites

Samples (2 ml) were taken from the cultivation broth, immediately filtered through a 0.45 μm pore size cellulose acetate filter (Osmonics, Westborough, MA, USA) and stored at −20 °C until further analysis. Glucose, galactose, glycerol, ethanol, acetate, pyruvate and succinate were separated and quantified on HPLC using an Aminex HPX-87H column (Bio-Rad) according to Zaldivar et al. (2002).

2.9. Off-gas analysis

Carbon dioxide and oxygen concentrations in the exhaust gas were monitored each 15 s by a Brüel and Kjær 1308 acoustic gas analyzer (Brüel and Kjær, Nærum, Denmark) (Christensen et al., 1995).

2.10. Calculations of specific growth rates and yield coefficients

The maximum specific growth rate was determined from measurement of dry weight of biomass as a function of time. Determination of yield coefficients for extracellular metabolites as well as biomass was based on a linear regression of their concentration as a function of the residual glucose concentration in the exponential growth phase. Specific product formation rates (r_p) and specific substrate uptake rates (r_s) for each strain within a particular growth condition were calculated using graphically determined values of the maximum specific growth rate (μ) and the biomass yield on substrate (Y_{sx}) or the product yield on biomass (Y_{xp}).

$$r_s = \frac{\mu}{Y_{sx}} \quad (1)$$

$$r_p = \mu Y_{xp} \quad (2)$$

3. Results

3.1. SNF1 and SNF4 deletions caused an increased acetate production and decreased glucose uptake during the exponential growth on glucose

The physiological characterization of the glucose repression recombinant Δ*snf1*, Δ*snf4*, Δ*snf1*Δ*snf4* strains in comparison to the reference strain was performed in glucose or glucose–galactose dynamic growth conditions. All the CEN.PK genotypic background *S. cerevisiae* yeast strains were cultivated and studied in aerobic batch cultivations with a minimal medium containing 30 g l^{-1} of D-glucose or a mixture of 15 g l^{-1} D-glucose and 15 g l^{-1} D-galactose as the sole carbon source.

During the exponential growth on glucose, the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains were found to have lower glucose uptake rates and higher acetate production rates compared to the reference strain (Table 2). The Δ*snf1* strain was found to have a 24% lower specific glucose uptake rate, a 40% lower specific ethanol production rate compared to the reference strain. This result correlated well with previous physiological observations and suggested that Snf1 through Mig1 was involved in the regulation of glucose sensing/uptake (Aon and Cortassa, 1998; Kaniak et al., 2004). The acetate yield on glucose was found to be 0.02 g g^{-1} for the Δ*snf4* and the Δ*snf1*Δ*snf4* strains, and 0.03 g g^{-1} for the Δ*snf1* strain and it was more than five-fold higher compared to the 0.004 g g^{-1} determined for the reference strain (Westergaard et al., 2004). Data indicated that the acetate conversion to CO_2 was lower in the recombinant strains compared to the reference strain during an exponential growth on glucose.

Table 2
Physiological parameters from aerobic glucose batch cultivations

Relevant genotype	Maximum specific glucose uptake rate[a] (g g^{-1} h^{-1})	Maximum specific growth rate (h^{-1})	Maximum specific ethanol production rate[a] (g g^{-1} h^{-1})	Biomass yield[b] (g g^{-1})	Ethanol yield[b] (g g^{-1})	Acetate yield[b] (g g^{-1})	Glycerol yield[b] (g g^{-1})
Reference strain[c]	3.1	0.31	1.0	0.10	0.34	0.004	0.07
Δ*snf1*	2.4	0.26	0.6	0.11	0.35	0.03	0.06
Δ*snf4*	2.8	0.27	1.0	0.10	0.37	0.02	0.08
Δ*snf1*Δ*snf4*	2.4	0.33	0.9	0.14	0.37	0.02	0.04

The physiological parameters in this table were calculated as described in Section 2. The average values calculated from two or three biological replica batch experiments are presented. Standard deviations were determined to be lower then 5% of the average value.

[a] Maximum specific substrate uptake rate is expressed as g of glucose consumed per g of dry weight per h; maximum specific ethanol production rate is expressed as g of ethanol produced per g of dry weight per h.

[b] Yields are expressed as g of product (biomass, ethanol, acetate or glycerol) produced per g of glucose consumed.

[c] Data from Westergaard et al. (2004).

3.2. The presence of galactose affects glucose fermentation of the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains, but not of the reference strain

The maximum specific growth rates on glucose did not change for any of the strains of the study whether they were cultivated in glucose or glucose–galactose-mixtures (Tables 2 and 3). The specific glucose uptake rate was found to be 20% lower in glucose–galactose-mixtures compared to the glucose batch cultivations for the recombinant strains and was found to be 2.0 g g^{-1} h^{-1} for the Δ*snf1* and 2.2 g g^{-1} h^{-1} for the Δ*snf4* and Δ*snf1*Δ*snf4* strains. The specific glucose uptake rate for the reference strain remained constant in both sets of cultivations and was equal to 3.1 g g^{-1} h^{-1} (Ostergaard et al., 2001; Westergaard et al., 2004). The Δ*snf1*Δ*snf4* strain was determined to have the highest biomass yield (0.14 g g^{-1}, 0.14 g g^{-1}) and the lowest glycerol yield (0.04 g g^{-1}, 0.03 g g^{-1}) on glucose, in glucose and glucose–galactose-mixtures, respectively (Tables 2 and 3). In glucose–galactose-mixtures, the single deletion Δ*snf1* and Δ*snf4* strains increased biomass yield on glucose by 30%, reduced glycerol yield on glucose by 60% (compared to glucose batches) and appeared to have the same physiological characteristics as those found for the double deletion strain (Table 3). Overall, our results showed that the double deletion of *SNF1* and *SNF4* or the presence of galactose in glucose growth conditions triggered changes in the central carbon metabolism towards biomass production for the recombinant strains of the study. The biomass yield on glucose for the reference strain remained constant in glucose, glucose–galactose-mixtures and was equal to 0.10 g g^{-1}.

3.3. Disruption of Snf1 kinase complex causes a delay of the induction of growth on galactose

It has been shown that Snf1 kinase is required for galactose metabolism only to relieve repression of the *GAL* genes by the Mig1–Cyc8–Tup1 complex (Johnston et al., 1994). In order to evaluate the effect of the disruption of the Snf1 kinase complex on the repression of the *GAL* regulon, the reference, Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains were grown in glucose–galactose-mixture batch cultivations. The lag phase between glucose depletion and initiation of galactose uptake was found to be 5 h for the reference strain, 12 h for the Δ*snf4*, 15.5 h for the Δ*snf1* and 48 h for the Δ*snf1*Δ*snf4* strains (Table 3). During the lag phase, a very slow specific growth rate (0.01 h^{-1}) was calculated for the Δ*snf4* and the Δ*snf1* strains. For the Δ*snf1*Δ*snf4* strain, there was no increase in biomass after glucose depletion for a period of 20 h, thereafter a specific growth rate of 0.01 h^{-1} was calculated for the remaining 28 h of the lag phase. The data showed that deletion of the genes *SNF4* or *SNF1* caused a similar, but reproducibly distinct delay in the induction of the *GAL* regulon and a delayed initiation of exponential growth on galactose. The galactose growth delay for the double deletion strain was found to be more than two times longer compared to the single deletion strains (Fig. 1).

It was not possible to observe any decrease of the extracellular galactose in any of the lag phases after the glucose was depleted

Table 3
Physiological parameters from aerobic glucose–galactose-mixture batch cultivations

Relevant genotype	Maximum specific glucose uptake rate[a] (g g^{-1} h^{-1})	Maximum specific galactose uptake rate[a] (g g^{-1} h^{-1})	Maximum specific growth rate on glucose (h^{-1})	Maximum specific growth rate on galactose (h^{-1})	Length of the lag phase (h)	Biomass yield[b] (g g^{-1})	Ethanol yield[b] (g g^{-1})	Acetate yield[b] (g g^{-1})	Glycerol yield[b] (g g^{-1})
Reference strain[c]	3.1	0.54	0.31	0.13	5	0.10	0.3	–	–
Δ*snf1*	2.0	0.24	0.26	0.07	12	0.13	0.36	0.02	0.02
Δ*snf4*	2.2	0.27	0.27	0.08	15.5	0.14	0.35	0.01	0.03
Δ*snf1*Δ*snf4*	2.2	0.14	0.30	0.04	48	0.14	0.36	0.02	0.03

The physiological parameters in this table were calculated as described in Section 2. The average values calculated from two biological replica batch experiments are presented. Standard deviations were determined to be lower then 5% of the average value.

[a] Maximum specific substrate uptake rate is expressed as g of glucose (or galactose) per g of dry weight per h.

[b] Yields are expressed as g of product (biomass, ethanol, acetate or glycerol) produced per g of glucose consumed.

[c] Data from Ostergaard et al. (2001).

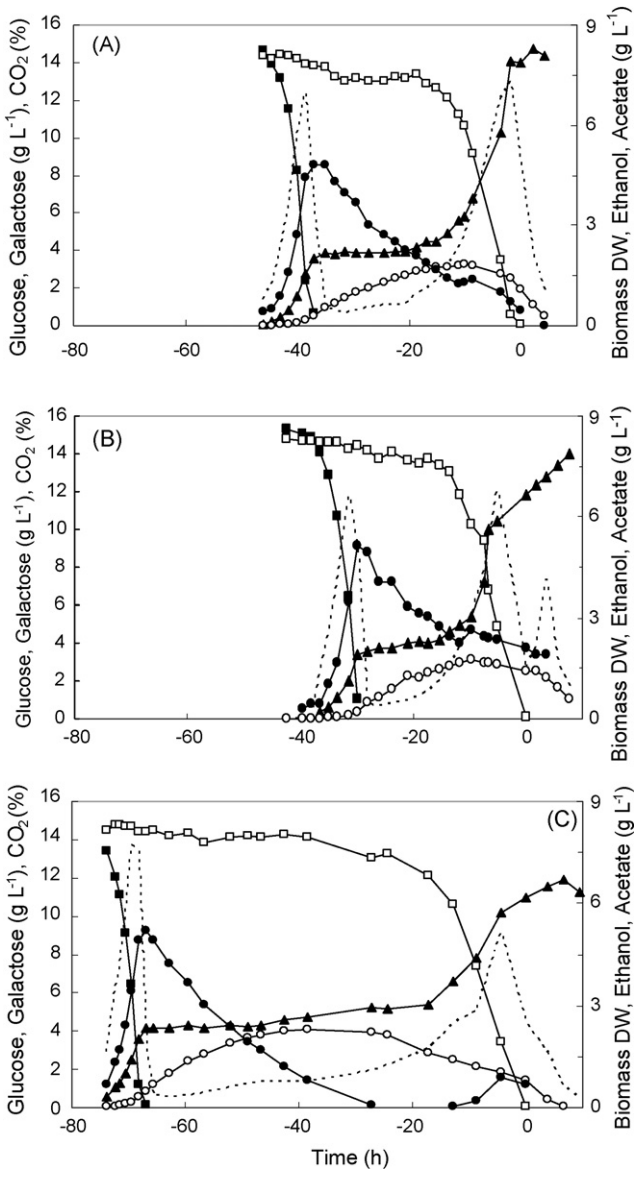


Fig. 1. Representative growth and fermentation profiles of the Δ*snf1* (A), Δ*snf4* (B) and Δ*snf1*Δ*snf4* (C) strains from aerobic glucose–galactose-mixture batch cultivations. Glycerol and pyruvate production profiles are not shown. Time zero is set to be the time of galactose depletion. Symbols: glucose, black square (■); galactose, white square (□); ethanol, black circle (●); acetate, white circle (○); biomass, black triangle (▲); CO_2 measure in outlet gas, % (C-mol l^{-1}), dashed line (- - -).

(Fig. 1). However, the amount of ethanol produced, during the growth phase on glucose, decreased after glucose depletion and throughout the lag phase. We neglected a possibility that cells grew on ethanol based on: (i) the a priori knowledge that Δ*snf1* and Δ*snf4* are not able to grow on ethanol (Schuller and Entian, 1987), (ii) no exponential growth was observed through the lag phase, (iii) the absence of CO_2 production during the decrease in ethanol concentration, which would indicate ethanol oxidation by the culture. Based on carbon balance calculations, we determined that part (∼65%) of the ethanol was oxidized to acetate, while producing NADH for cell maintenance. The other part (∼35%) of the ethanol was lost due to evaporation.

3.4. Investigation of poor galactose growth in Δsnf1, Δsnf4 and Δsnf1Δsnf4 CEN.PK background strains

The maximum specific growth rate on galactose was found to be 0.13 h^{-1} for the reference strain, 0.08 h^{-1} for the Δ*snf4*, 0.07 h^{-1} for the Δ*snf1* and 0.04 h^{-1} for the Δ*snf1*Δ*snf4* in glucose–galactose-mixture batch cultivations (Table 3). It correlated well with galactose growth induction delay and could probably be explained by gradually different Mig1-based repression of *GAL* genes in the recombinant strains. In addition, we observed that the recombinant strains were consuming acetate simultaneously with galactose in the glucose–galactose-mixture batch cultivations. This phenotype was observed especially clearly in the Δ*snf1*Δ*snf4* strain cultivations (Fig. 1). Previous studies in our laboratory showed that the reference strain consumed acetate straight after galactose depletion (data not published). Further studies have to be done to investigate whether acetate co-consumption has any impact on the calculated specific growth rates on galactose for the recombinant strains of this study.

4. Discussion

4.1. Changes within glucose repression regulatory cascade under glucose excess

It has been shown that under glucose excess the Snf1 kinase is inactive (McCartney and Schmidt, 2001), thus non-phosphorylated Mig1 locates in the nucleus and represses genes involved in the catabolism of alternative carbon sources as well as hexose transporters (Kaniak et al., 2004; Papamichos-Chronakis et al., 2004). In this study, increased acetate production and decreased glucose uptake were calculated under glucose excess for the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* compared to the reference strain (Tables 2 and 3). These results suggested that genes involved in the degradation of acetate (e.g., *ACS1*) (Kratzer and Schuller, 1997) and genes coding for hexose transporters (*HXT2*, *HXT4*) (Kaniak et al., 2004) were likely to be less expressed in the deletion strains compared to the reference strain even under glucose excess, where the repression of the genes is normally fairly strong. To explain this phenomenon, we proposed that the disrupted Snf1 kinase complex resulted in a shift in the equilibrium of the kinase Snf1 and phosphatase Glc7 catalyzed reactions towards the non-phosphorylated state of Mig1 (Gancedo, 1998; Schuller, 2003). Consequently, more of Mig1 was available to repress the expression of metabolic genes as well as genes coding for the activators of gluconeogenesis (example Cat8) (Hedges et al., 1995) in the conditions where there is normally fairly strong glucose repression. In addition, the absence of the Snf1 kinase complex excluded the phosphorylation and, thus, the activation of Cat8 (Hedges et al., 1995; Young et al., 2003) in the deletion strains. The sum of the changes within glucose repression cascade that resulted due to the deletions of *SNF1* and *SNF4* likely caused the observed changes in carbon metabolism.

4.2. Snf1 and Snf4 roles in Snf1 kinase complex in glucose repression regulatory cascade and outside of it

Deletion of *SNF1* or *SNF4* disrupts glucose repression regulatory cascade, since neither Δ*snf1* nor Δ*snf4* are able to grow on non-fermentable carbon sources (Schuller and Entian, 1987). Galactose catabolism, in addition to glucose repression regulatory cascade, is under the galactose induction system (Rubio-Texeira, 2005) and our study showed that the deletion strains were able to initiate growth on galactose. In comparison to the reference strain it was expected that the loss of Snf1 kinase (Δ*snf1* and Δ*snf1*Δ*snf4* cases) caused more severe changes in galactose metabolism than the change caused by the loss of its regulatory subunit Snf4 (Δ*snf4* case). The study showed that the delay of the induction of growth on galactose was much longer in the double deletion strain compared to the delays for the Δ*snf1* and Δ*snf4* strains in the glucose–galactose-mixture cultivations (Fig. 1 and Table 3). The results suggested that not only Snf1 kinase complex regulated Mig1 repression of *GAL* genes (Gancedo, 1998), but other functions of Snf1 or Snf4 contribute to the regulation of galactose metabolism.

It has been shown by Cortassa and Aon (1998) that Snf1 kinase is involved not only in the release of glucose repression, but also in the regulation of cellular energetics. It has been supported by the three- to four-fold lower ATP/ADP concentrations measured in cellular extracts from the Δ*snf1* compared to the Δ*snf4* and reference strains. Lin et al. (2001) have shown that the Δ*snf4* strain had four-fold lower intracellular ATP levels compared to the reference strain. In glucose non-repressive growth conditions Gal3 requires the presence of galactose and ATP to initiate galactose induction (Bhat and Murthy, 2001). It has not been proven that Gal1 requires ATP to perform the regulatory function of the expression of *GAL* genes (Bhat and Hopper, 1992). Based on the provided background information and the results from this study we propose that, besides the continuous Mig1-based repression of *GAL* genes in all of the recombinant strains, the lower intracellular ATP in the Δ*snf1* and Δ*snf4* strains enabled the Gal3 protein to interact with Gal80. Consequently, the Gal1 protein might have over-shadowed the Gal3 function causing a delayed activation of the *GAL* regulon (Bhat and Murthy, 2001), however further investigations are needed to prove this.

4.3. CEN.PK background strains with disrupted Snf1 kinase complex are able to poorly grow on galactose

This is the first study describing the Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains' growth characteristics in dynamic glucose–galactose batch cultivations under well-defined cultivation conditions. Previously, Δ*snf1* and Δ*snf4* galactose growth studies were performed using shake-flask cultivations and plate assays. Different galactose growth phenotypes have been observed for the Δ*snf1* and Δ*snf4* strains and the possibility of suppressor mutations causing those phenotypes have been discussed in previous and mainly agar plating studies (Carlson et al., 1984; Palecek et al., 2002; Schuller and Entian, 1991; Thompson-Jaeger et al., 1991; Van Driessche et al., 2005). All necessary experiments were performed to assure that no suppressor mutations have appeared in response to selective pressures of galactose in this study (described in Section 2).

A more than 24 h lag phase was observed for the Δ*snf1* and Δ*snf4* recombinant strains when these strains were inoculated in minimal media with galactose as a sole carbon source in shake-flask or batch cultivations (data not shown). In glucose–galactose-mixtures, after cells were pre-grown on glucose, twice as fast (15.5 h for the Δ*snf1* and 12 h for the Δ*snf4* case) initiation of growth on galactose was observed (Fig. 1 and Table 3) compared to conditions where galactose was a sole carbon source. The results showed that different cultivation conditions influenced observed galactose growth phenotypes. In addition, liquid cultures offer different growth conditions than solid substitute (agar plates) and therefore our results can not be directly compared to earlier studies where ability to grow on galactose was mainly tested on agar plates (Carlson et al., 1984; Schuller and Entian, 1991; Thompson-Jaeger et al., 1991).

It has been previously shown that in different genotypic background strains, Mig1 represses expression of the same genes in a range from a few to a few hundred folds (Lutfiyya and Johnston, 1996; Olsson et al., 1997). It has been shown that higher levels of cAMP can decrease the expression of genes subjected to glucose repression (Zaragoza et al., 1999). The CEN.PK background strains, also used in this study, have been described as having 'a lack of cAMP response' phenotype, due to a point mutation in the adenylate cyclase gene (Van Dijken et al., 2000; Vanhalewyn et al., 1999). We propose that distinct cAMP signaling properties may contribute to the incomplete repression of *GAL* genes in the CEN.PK background Δ*snf1*, Δ*snf4* and Δ*snf1*Δ*snf4* strains. Overall, the genotypic background of the strain and cultivation conditions may impact observed galactose growth phenotypes. The results of this study implied that not only the Snf1 kinase was responsible for the repression/derepression of galactose metabolism, but other signaling mechanism (i.e. cAMP level) had an impact on the observed phenotypes.

4.4. The Mig1 repression and an accumulation of intracellular galactose derivatives affect the maximum specific growth rates on galactose for the Δsnf1, Δsnf4 and Δsnf1Δsnf4 strains

In glucose–galactose-mixture cultivations, the maximum specific growth rates on galactose were found to be lower for the recombinant strains compared to the reference strain and were found to correlate with the delayed growth induction on galactose (Table 3). We propose that the reduced growth of the recombinant strains on galactose is a likely response to a continuous Mig1-based repression of *GAL* genes (Nehlin et al., 1991) leading to repressed galactose uptake and catabolism. In addition, we suggest that due to the partly inactive Leloir pathway (caused by the continuous Mig1 repression of the *GAL* genes) in the recombinant strains, the intracellular galactose may be partly converted into galactitol through a metabolic pathway yet poorly described in *S. cerevisiae*, but clearly unrelated to the Leloir pathway (Petrash et al., 2001). We have previously observed that *S. cerevisiae* accumulated galactitol during its

growth on galactose, and reduced maximum specific growth rate correlated well with the increased galactitol accumulation (data not published). Another toxic compound, galactose-1-phosphate (Bhat, 2003), may also accumulate in the recombinant strains, if Gal1 instead of Gal3 was activated in order to induce the growth initiation on galactose. As Gal1 also has a kinase catalytic activity, it converts galactose to galactose-1-phosphate, meanwhile, the other Leloir pathway enzymes are still not expressed at this early stage of induction of *GAL* regulon leading to possible galactose-1-phosphate accumulation in the cell, as it has been shown previously (Meyer et al., 1991). Toxicity caused by the galactose derivatives should be proved in further studies.

Overall, the results of this study strongly suggest that the continuing Mig1-based repression of *GAL* genes, changes in cellular energetics and possible accumulation of toxic galactose derivatives inhibited galactose growth initiation and maximum specific growth rate on galactose in the recombinant strains compared to the reference strain in the glucose–galactose-mixtures.

Acknowledgements

We thank Jakob R.H. Nielsen for initial batch cultivations of Δ*snf1* and Δ*snf4* strains, Wian de Jongh for sharing unpublished results regarding galactitol. This work was supported by the Danish Research Agency.

References

Aon, M.A., Cortassa, S., 1998. Catabolite repression mutants of *Saccharomyces cerevisiae* show altered fermentative metabolism as well as cell cycle behavior in glucose-limited chemostat cultures. Biotechnol. Bioeng. 59, 203–213.

Bhat, P.J., 2003. Galactose-1-phosphate is a regulator of inositol monophosphatase: a fact or a fiction? Med. Hypotheses 60, 123–128.

Bhat, P.J., Hopper, J.E., 1992. Overproduction of the GAL1 or GAL3 protein causes galactose-independent activation of the GAL4 protein: evidence for a new model of induction for the yeast *GAL/MEL* regulon. Mol. Cell Biol. 12, 2701–2707.

Bhat, P.J., Murthy, T.V., 2001. Transcriptional control of the *GAL/MEL* regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. Mol. Microbiol. 40, 1059–1066.

Boeke, J.D., LaCroute, F., Fink, G.R., 1984. A positive selection for mutants lacking orotidine-5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. 197, 345–346.

Carlson, M., Osmond, B.C., Botstein, D., 1981. Mutants of yeast defective in sucrose utilization. Genetics 98, 25–40.

Carlson, M., Osmond, B.C., Neigeborn, L., Botstein, D., 1984. A supressor of *SNF1* mutations causes constitutive high-level invertase synthesis in yeast. Genetics 107, 19–32.

Christensen, L.J., Schulze, U., Nielsen, J., Villadsen, J., 1995. Acoustic gas analysis for fast and precise monitoring of bioreactors. Chem. Eng. Sci. 50, 2601–2610.

Cortassa, S., Aon, M.A., 1998. The onset of fermentative metabolism in continuous cultures depends on the catabolite repression properties of *Saccharomyces cerevisiae*. Enzyme Microb. Technol. 22, 705–712.

Gancedo, J.M., 1998. Yeast carbon catabolite repression. Microbiol. Mol. Biol. Rev. 62, 334–361.

Gietz, R.D., Woods, R.A., 2002. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. Methods Enzymol. 350, 87–96.

Hedges, D., Proft, M., Entian, K.D., 1995. *CAT8*, a new zinc cluster-encoding gene necessary for derepression of gluconeogenic enzymes in the yeast *Saccharomyces cerevisiae*. Mol. Cell Biol. 15, 1915–1922.

Johnston, M., Flick, J.S., Pexton, T., 1994. Multiple mechanisms provide rapid and stringent glucose repression of *GAL* gene expression in *Saccharomyces cerevisiae*. Mol. Cell Biol. 14, 3834–3841.

Kaniak, A., Xue, Z., Macool, D., Kim, J.H., Johnston, M., 2004. Regulatory network connecting two glucose signal transduction pathways in *Saccharomyces cerevisiae*. Eukaryot. Cell 3, 221–231.

Kim, M.D., Lee, T.H., Lim, H.K., Seo, J.H., 2004. Production of antithrombotic hirudin in *GAL1*-disrupted *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 65, 259–262.

Kratzer, S., Schuller, H.J., 1997. Transcriptional control of the yeast acetyl-CoA synthetase gene, *ACS1*, by the positive regulators CAT8 and ADR1 and the pleiotropic repressor UME6. Mol. Microbiol. 26, 631–641.

Kuchin, S., Vyas, V.K., Kanter, E., Hong, S.P., Carlson, M., 2003. Std1p (Msn3p) positively regulates the Snf1 kinase in *Saccharomyces cerevisiae*. Genetics 163, 507–514.

Lin, S.S., Manchester, J.K., Gordon, J.I., 2001. Enhanced gluconeogenesis and increased energy storage as hallmarks of aging in *Saccharomyces cerevisiae*. J. Biol. Chem. 276, 36000–36007.

Lutfiyya, L.L., Johnston, M., 1996. Two zinc-finger-containing repressors are responsible for glucose repression of *SUC2* expression. Mol. Cell Biol. 16, 4790–4797.

McCartney, R.R., Schmidt, M.C., 2001. Regulation of Snf1 kinase. Activation requires phosphorylation of threonine 210 by an upstream kinase as well as a distinct step mediated by the Snf4 subunit. J. Biol. Chem. 276, 36460–36466.

Meyer, J., Walker-Jonah, A., Hollenberg, C.P., 1991. Galactokinase encoded by *GAL1* is a bifunctional protein required for induction of the *GAL* genes in *Kluyveromyces lactis* and is able to suppress the gal3 phenotype in *Saccharomyces cerevisiae*. Mol. Cell Biol. 11, 5454–5461.

Nehlin, J.O., Carlberg, M., Ronne, H., 1991. Control of yeast *GAL* genes by MIG1 repressor: a transcriptional cascade in the glucose response. EMBO J. 10, 3373–3377.

Neigeborn, L., Carlson, M., 1984. Genes affecting the regulation of *SUC2* gene expression by glucose repression in *Saccharomyces cerevisiae*. Genetics 108, 845–858.

Olsson, L., Nielsen, J., 2000. The role of metabolic engineering in the improvement of *Saccharomyces cerevisiae*: utilization of industrial media. Enzyme Microb. Technol. 26, 785–792.

Olsson, L., Larsen, M.E., Rønnow, B., Mikkelsen, J.D., Nielsen, J., 1997. Silencing *MIG1* in *Saccharomyces cerevisiae*: effects of antisense *MIG1* expression and *MIG1* gene disruption. Appl. Env. Microbiol. 63, 2366–2371.

Ostergaard, S., Roca, C., Rønnow, B., Nielsen, J., Olsson, L., 2000. Physiological studies in aerobic batch cultivations of *Saccharomyces cerevisiae* strains harboring the *MEL1* gene. Biotechnol. Bioeng. 68, 252–259.

Ostergaard, S., Walloe, K.O., Gomes, S.G., Olsson, L., Nielsen, J., 2001. The impact of *GAL6*, *GAL80*, and *MIG1* on glucose control of the *GAL* system in *Saccharomyces cerevisiae*. FEMS Yeast Res. 1, 47–55.

Ozcan, S., Johnston, M., 1999. Function and regulation of yeast hexose transporters. Microbiol. Mol. Biol. Rev. 63, 554–569.

Palecek, S.P., Parikh, A.S., Huh, J.H., Kron, S.J., 2002. Depression of *Saccharomyces cerevisiae* invasive growth on non-glucose carbon sources requires the Snf1 kinase. Mol. Microbiol. 45, 453–469.

Papamichos-Chronakis, M., Gligoris, T., Tzamarias, D., 2004. The Snf1 kinase controls glucose repression in yeast by modulating interactions between the Mig1 repressor and the Cyc8–Tup1 co-repressor. EMBO Rep. 5, 368–372.

Petrash, J.M., Murthy, B.S., Young, M., Morris, K., Rikimaru, L., Griest, T.A., Harter, T., 2001. Functional genomic studies of aldo-keto reductases. Chem. Biol. Interact. 130–132, 673–683.

Rahner, A., Hiesinger, M., Schuller, H.J., 1999. Deregulation of gluconeogenic structural genes by variants of the transcriptional activator Cat8p of the yeast *Saccharomyces cerevisiae*. Mol. Microbiol. 34, 146–156.

Reid, R.J., Sunjevaric, I., Keddache, M., Rothstein, R., 2002. Efficient PCR-based gene disruption in *Saccharomyces* strains using intergenic primers. Yeast 19, 319–328.

Rolland, F., Winderickx, J., Thevelein, J.M., 2002. Glucose-sensing and -signalling mechanisms in yeast. FEMS Yeast Res. 2, 183–201.

Rose, M., Winston, F., Hieter, P., 1990. Methods in Yeast Genetics: A Laboratory Course Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Rubio-Texeira, M., 2005. A comparative analysis of the *GAL* genetic switch between not-so-distant cousins: *Saccharomyces cerevisiae* versus *Kluyveromyces lactis*. FEMS Yeast Res. 5, 1115–1128.

Schmidt, M.C., McCartney, R.R., 2000. Beta-subunits of Snf1 kinase are required for kinase function and substrate definition. EMBO J. 19, 4936–4943.

Schuller, H.J., 2003. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. Curr. Genet. 43, 139–160.

Schuller, H.J., Entian, K.D., 1987. Isolation and expression analysis of two yeast regulatory genes involved in the derepression of glucose-repressible enzymes. Mol. Gen. Genet. 209, 366–373.

Schuller, H.J., Entian, K.D., 1991. Extragenic suppressors of yeast glucose derepression mutants leading to constitutive synthesis of several glucose-repressible enzymes. J. Bacteriol. 173, 2045–2052.

Thoden, J.B., Sellick, C.A., Timson, D.J., Reece, R.J., Holden, H.M., 2005. Molecular structure of *Saccharomyces cerevisiae* Gal1p, a bifunctional galactokinase and transcriptional inducer. J. Biol. Chem. 280, 36905–36911.

Thompson-Jaeger, S., Francois, J., Gaughran, J.P., Tatchell, K., 1991. Deletion of *SNF1* affects the nutrient response of yeast and resembles mutations which activate the adenylate cyclase pathway. Genetics 129, 697–706.

Treitel, M.A., Kuchin, S., Carlson, M., 1998. Snf1 protein kinase regulates phosphorylation of the Mig1 repressor in *Saccharomyces cerevisiae*. Mol. Cell Biol. 18, 6273–6280.

Van Dijken, J.P., Bauer, J., Brambilla, L., Duboc, P., Francois, J.M., Gancedo, C., Giuseppin, M.L., Heijnen, J.J., Hoare, M., Lange, H.C., Madden, E.A., Niederberger, P., Nielsen, J., Parrou, J.L., Petit, T., Porro, D., Reuss, M., van, R.N., Rizzi, M., Steensma, H.Y., Verrips, C.T., Vindelov, J., Pronk, J.T., 2000. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. Enzyme Microb. Technol. 26, 706–714.

Van Driessche, B., Coddens, S., Van Mullem, V., Vandenhaute, J., 2005. Glucose deprivation mediates interaction between CTDK-I and Snf1 in *Saccharomyces cerevisiae*. FEBS Lett. 579, 5318–5324.

Vanhalewyn, M., Dumortier, F., Debast, G., Colombo, S., Ma, P., Winderickx, J., Van Dijck, P., Thevelein, J.M., 1999. A mutation in *Saccharomyces cerevisiae* adenylate cyclase, Cyr1K1876M, specifically affects glucose- and acidification-induced cAMP signaling and not the basal cAMP level. Mol. Microbiol. 32, 363–376.

Verduyn, C., Postma, E., Scheffers, W.A., Van Dijken, J.P., 1992. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. Yeast 8, 501–517.

Westergaard, S.L., Bro, C., Olsson, L., Nielsen, J., 2004. Elucidation of the role of Grr1p in glucose sensing by *Saccharomyces cerevisiae* through genome-wide transcription analysis. FEMS Yeast Res. 5, 193–204.

Young, E.T., Dombek, K.M., Tachibana, C., Ideker, T., 2003. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. J. Biol. Chem. 278, 26146–26158.

Zaldivar, J., Borges, A., Johansson, B., Smits, H.P., Villas-Boas, S.G., Nielsen, J., Olsson, L., 2002. Fermentation performance and intracellular metabolite patterns in laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 59, 436–442.

Zaragoza, O., Lindley, C., Gancedo, J.M., 1999. Cyclic AMP can decrease expression of genes subject to catabolite repression in *Saccharomyces cerevisiae*. J. Bacteriol. 181, 2640–2642.

Zhou, H., Winston, F., 2001. *NRG1* is required for glucose repression of the *SUC2* and *GAL* genes of *Saccharomyces cerevisiae*. BMC Genet. 2, 5.

Chapter 3:

Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods

Renata Usaite, James Wohlschlegel, John D. Venable, Sung K. Park, Jens Nielsen, Lisbeth Olsson and John R. Yates III

J Proteome Res. 2008 Jan; 7(1):266-275

research articles Journal of proteome research

Characterization of Global Yeast Quantitative Proteome Data Generated from the Wild-Type and Glucose Repression *Saccharomyces cerevisiae* Strains: The Comparison of Two Quantitative Methods

Renata Usaite,[†,‡] James Wohlschlegel,[‡] John D. Venable,[‡] Sung K. Park,[‡] Jens Nielsen,[†] Lisbeth Olsson,[†] and John R. Yates III*[,‡]

BioCentrum-DTU, Technical University of Denmark, Kgs. Lyngby, Denmark, and The Scripps Research Institute, La Jolla, California

Received September 06, 2007

The quantitative proteomic analysis of complex protein mixtures is emerging as a technically challenging but viable systems-level approach for studying cellular function. This study presents a large-scale comparative analysis of protein abundances from yeast protein lysates derived from both wild-type yeast and yeast strains lacking key components of the Snf1 kinase complex. Four different strains were grown under well-controlled chemostat conditions. Multidimensional protein identification technology followed by quantitation using either spectral counting or stable isotope labeling approaches was used to identify relative changes in the protein expression levels between the strains. A total of 2388 proteins were relatively quantified, and more than 350 proteins were found to have significantly different expression levels between the two strains of comparison when using the stable isotope labeling strategy. The stable isotope labeling based quantitative approach was found to be highly reproducible among biological replicates when complex protein mixtures containing small expression changes were analyzed. Where poor correlation between stable isotope labeling and spectral counting was found, the major reason behind the discrepancy was the lack of reproducible sampling for proteins with low spectral counts. The functional categorization of the relative protein expression differences that occur in Snf1-deficient strains uncovers a wide range of biological processes regulated by this important cellular kinase.

Keywords: CenSus • MudPIT • Snf4 • AMP-activated kinase • 15N • mass spectrometry • proteomics

Introduction

In the field of systems biology, the integration of different -omics technologies is critical for understanding the cell as a whole. Compared to transcriptomics, proteomics continues to remain elusive in terms of generating and integrating experimental data into interpretive and predictive models. Since proteome analysis provides complementary information about biological systems and pathways, it is essential that new proteomic methodologies are developed to provide a comprehensive look at a biological process.

The proteome complexity is overwhelming, and based on Human Proteome Project findings, there are far fewer protein-coding genes in the human genome than proteins in the human proteome, which is estimated to reach 1 000 000 protein isoforms.[1] Diversity of post-translation modifications, alternative splicing, and dynamics of protein complexes generate this complexity, which holds important biological information. Detection and understanding of the complete proteome will therefore allow uncovering of global signaling and kinetic traits. However, the complexity of the proteome remains one of the major issues preventing us from having analytical tools that allow comprehensive detection and/or quantification of the "complete" proteome.

Various tools combining mass spectrometry and liquid chromatography or gel-based protein separation are being developed and used to improve complex proteome identification.[2] Multidimensional protein identification technology (MudPIT) has emerged as a sensitive tool for the separation and identification of proteins from complex mixtures.[3,4] MudPIT, in combination with the metabolic or in vitro labeling of proteins, generates accurate and abundant quantitative proteome results.[5,6] Other proteome identification and quantification tools based on protein chip technology or other mass spectrometry based strategies find their niche in specific proteome areas as reviewed in Patterson et al.[7]

Accurate quantitative global proteomics data are critical for studying global cellular processes and for integrating the data with other -omics data. Different approaches have been and

* Corresponding author. The Proteomic Mass Spectrometry Lab, The Scripps Research Institute, 10550 North Torrey Pines Rd., SR11, Department of Cell Biology, La Jolla, CA 92037. Tel.: (858)784-8862. Fax: (858)784-8883. E-mail: jyates@scripps.edu.
† Technical University of Denmark.
‡ The Scripps Research Institute.

10.1021/pr700580m CCC: $40.75
Published on Web 01/04/2008

Table 1. Yeast *Saccharomyces cerevisiae* Strains Used in this Study

strain[a]	name in text	genotype	source/ref
CEN.PK 113-7D	wild-type strain	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8^C	provided by P. Kötter[b]
CEN.PK 506-1C	Δsnf1	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2–8^C snf1 (4,1899)::loxP-Kan-loxP	provided by P. Kötter
CEN.PK 507-5B	Δsnf4	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2–8^C snf4 (4,966)::loxP-Kan-loxP	provided by P. Kötter
IBT100072	Δsnf1Δsnf4	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2–8^C snf1 (1,1903) snf4 (4,966)::loxP-Kan-loxP	Usaite et al.[c]

[a] The strains were derived from the parental laboratory CEN.PK background strain.[34] [b] Institut für Mikrobiologie, Frankfurt, Germany. [c] Strain design and characteristics have been published in Usaite et al.[19]

are being developed to quantify protein expression differences using mass spectrometry.[8–10] Relatively few comparative studies of different quantitative proteome approaches have been performed (when real biological samples were used) to evaluate accuracy and reproducibility of quantitative proteome results.[11,12] Zybailov and coauthors have compared quantitative proteome outputs generated by spectral counting and the stable isotope labeling approach, using RelEx.[13] They argued that a strong correlation between two proteome quantification methods can be obtained and that spectral counting was more reproducible and capable of quantifying proteins over a wider dynamic range.

In this study, we used the yeast *Saccharomyces cerevisiae* as a model system. To evaluate the effect of a genetic perturbation, we evaluated three deletion strains, where Snf1 kinase, its regulatory subunit Snf4, or both Snf1 and Snf4 complex subunits were deleted, in comparison to the wild-type strain. The Snf1 protein kinase complex plays a role in the glucose repression signaling cascade and regulates carbon metabolism through phosphorylation of transcription factors such as Mig1 and Cat8.[14] Snf1 has been found to directly regulate the activity of signaling proteins such as Hog85, metabolic proteins such as Acc1, and phosphorylates histone H3.[15,16] Various small-scale studies on the Snf1 protein imply that Snf1, as well as its mammalian homologue AMPK, might be a global regulator and energy balancer in the cell.[17] It is therefore interesting to perform global-scale quantitative proteome studies on the Snf1-deficient strains, as we expect to affect the expression of a large number of proteins when the Snf1 kinase complex is dysfunctional.

In this study, a global quantitative proteome data set was generated and analyzed using two different quantitative strategies, and its quality was evaluated using classical statistics and prior biological knowledge. Four yeast strains (Δsnf1, Δsnf4, Δsnf1Δsnf4, and wild-type strain) were metabolically ^{14}N- and ^{15}N-labeled and pregrown in biological triplicates in well-controlled steady state carbon limited chemostat cultivations, which are highly reproducible and suitable for global-scale comparative studies.[18] On the basis of previous studies,[14,19] it was expected that the Snf1-deficient strains had different transcripts and consequently altered protein expression profiles (compared to the wild-type strain) due to the lasting glucose repression effect under carbon limitation. Global protein extracts were generated from the yeast culture samples and analyzed using online multidimensional fractionation coupled with tandem mass spectrometry using an LTQ-Orbitrap mass spectrometer.[20] Stable isotope labeling and spectral counting approaches were used to calculate protein expression differences, and the two quantitative proteomic outputs were statistically evaluated and compared.

Materials and Methods

Yeast Strains. All *S. cerevisiae* strains used in this study were generated from the CEN.PK 113-7D laboratory strain (Scientific Research & Development GmbH, Oberursel, Germany) (Table 1).

Chemostat Cultivation of Cells. Steady state aerobic chemostat cultures were grown at 30 °C in 2 L bioreactors (Applikon) using a working volume of 0.5 L and a dilution rate of $D = 0.100$ ($\pm$ 0.005) h^{-1}. Cultures were fed with a modified minimal medium containing 75 mM nitrogen and 250 mM carbon (calculated on a per atom basis), which ensured that growth was limited by the carbon source and nitrogen was in excess. The sole difference in growth conditions among the strains used was that the wild-type strain was grown in >99% ammonium−^{15}N sulfate, >99 atom % ^{15}N, manufactured by Spectra Stable Isotopes (Columbia, MD), while the deletion Δsnf1, Δsnf4, and Δsnf1Δsnf4 strains were grown in 99% ammonium−^{14}N sulfate, 99 atom % ^{14}N, manufactured by ISOTEC INC. (Miamisburg, OH). The pH was measured online and kept constant at 5.0 by automatic titration with 4 M KOH using an Applikon (ADI1030) biocontroller. The stirring speed was set to 800 rpm, and the dry airflow rate was 0.5 L/min. The exhaust gas from chemostat cultivation was led through a condenser, and the mole % of carbon dioxide and oxygen was measured online using a PC-controlled acoustic gas analyzer (Brüel & Kjær, Denmark). After batch cultivation, dry weight, metabolite concentrations, and gas profiles were monitored until they reached a steady state (constant over at least five residence times) after which samples for protein extraction were taken. All samples were collected no later than 10–15 residence times from the start of continuous operation: (i) to avoid any strain adaptation that generally occurs over long-term cultivation,[21] and (ii) to ensure that metabolic labeling was enriched to >99%.

Media. The carbon limited minimal medium composition was based on the study described by Verduyn et al.[22] The amounts per liter of the compounds were: $(NH_4)_2SO_4$, 5 g; KH_2PO_4, 3 g; $MgSO_4 \cdot 7H_2O$, 0.5 g; D-glucose, 7.5 g; Antifoam 289 (A-5551, Sigma-Aldrich), 0.05 mL; EDTA (Titriplex III), 15.0 mg; $ZnSO_4 \cdot 7H_2O$, 4.5 mg; $MnCl_2 \cdot 2H_2O$, 0.82 mg; $CoCl_2 \cdot 6H_2O$, 0.3 mg; $CuSO_4 \cdot 5H_2O$, 0.3 mg; $Na_2MoO_4 \cdot 2H_2O$, 0.4 mg; $CaCl_2 \cdot 2H_2O$, 4.5 mg; $FeSO_4 \cdot 7H_2O$, 3.0 mg; H_3BO_3, 1.0 mg; KI, 0.1 mg; biotin, 0.05 mg; *p*-benzoic acid, 0.2 mg; nicotinic acid, 1.0 mg; Ca-pantothenate, 1.0 mg; pyridoxin, HCl, 1.0 mg; thiamin, HCl, 1.0 mg; *m*-inositol, 25.0 mg.

Isolation and Extraction of the Yeast Total Protein Pool. Samples for total protein isolation were taken from the chemostat cultivations by rapid sampling of 2 × 200 mL of culture in a 500 mL centrifuge tube containing 200 mL of crushed ice.

Cells were quickly pelleted (4500 rpm at 0 °C for 2 min), instantly frozen by dropwise addition into liquid nitrogen, and stored at −80 °C until further analysis. Cells were lysed in a buffer containing 7.4 mL/L of β-mercaptoethanol and 7.4 g/L of NaOH while incubated at 4 °C for 20 min. The proteins were precipitated by using 25% TCA (v/v) and ice-cold acetone. The pellet was air-dried, and the protein fraction was obtained by extracting the pellet with a 1 × Invitrosol (Invitrogen) and 8 M urea mixture. The protein concentration was determined using the BCA Protein Assay Kit (PIERCE). Lysates from ^{14}N-labeled and ^{15}N-labeled samples were mixed 1:1 by protein weight and subjected to protein digestion.

Digest of Extracted Protein Pool. An amount of 200 μg of total protein (100 μg of ^{14}N-labeled and 100 μg of ^{15}N-labeled) was reduced by adding Tris(2-carboxyethyl) phosphine (TCEP) to 5 mM and incubating at 20 °C for 30 min. The reduced sample was then carboxyamidomethylated by adding iodoacetamide (IAA) to 10 mM and incubating at 20 °C for 20 min in the dark. Endoproteinase LysC was added at an enzyme/substrate ratio of 1:100, and the samples were incubated for 4 h at 37 °C. The samples were subsequently diluted to 2 M urea with 100 mM Tris-HCl, pH 8.5, brought to 1 mM $CaCl_2$, and further digested by adding trypsin at an enzyme/substrate ratio of 1:20 and incubating overnight at 37 °C. The digestion reaction was quenched by adding formic acid to 5% (v/v) to reduce the pH to 2–3. The peptide mixture was purified by solid-phase extraction using SPEC-PLUS PTC18 cartridges (Ansys Diagnostics, Lake Forest, CA). Samples were freeze-dried and resuspended in 5% formic acid solution. Samples not immediately analyzed were stored at −80 °C.

MudPIT Analysis. The protein pool digest was pressure-loaded onto a 250 μM ID fused silica capillary column with a filtered union (UpChurch Scientific, Oak Harbor, WA) that was previously packed with 3 cm of 5 μm Partisphere strong cation exchanger (Whatman, Clifton, NJ) followed by 3 cm of 5 μm Aqua C18 material (Phenomenex, Ventura, CA). After loading, this trapping column was washed with buffer containing 5% acetonitrile/0.1% formic acid. After desalting, a 100 μm i.d. capillary with a 5 μm pulled tip packed with 18 cm of 3 μm Aqua C18 material (Phenomenex, Ventura, CA) was attached to the filter union, and the entire split-phase column (desalting column–filter union–analytical column) was placed inline with an Eksigent nanoLC-2D HPLC (Dublin, CA) and analyzed using a seven-step separation modified from that described previously.[3,4] A further optimization of MudPIT was performed in this study. It showed that extending the length of the analytical column from 10 to 18 cm and reducing the number of chromatography steps from 12 to 7, while keeping the total analysis time the same (24 h), improved peptide detection and resulted in identification of 30% more protein IDs. The buffers used were 5% acetonitrile/0.1% formic acid (buffer A), 80% acetonitrile/0.1% formic acid (buffer B), and 500 mM ammonium acetate/5% acetonitrile/0.1% formic acid (buffer C). Steps 1 and 7 had a 120 min lasting buffer B gradient profile of 2–90%. Steps 2–6 had the following profile: 10 min of 98% buffer A, 5 min of *X*% salt pulse, a 10 min gradient from 2 to 10% buffer B, a 180 min gradient from 10 to 50% buffer B, and a 15 min gradient from 50 to 90% buffer B. The 5 min salt pulse percentages (*X*) were 0, 10, 20, 30, 50, 70, and 100%, respectively, for the seven-step analysis. As peptides eluted from the microcapillary column, they were electrosprayed directly into a linear ion trap/Orbitrap (LTQ-Orbitrap) hybrid mass spectrometer (Thermo Electron Corp., Bremen, Germany) with the application of a distal electrospray voltage of 2.5 kV versus the inlet of the mass spectrometer. A cycle of one full-scan mass spectrum (150–2000 m/z) collected in the orbitrap mass analyzer followed by four data-dependent MS/MS spectra collected in the LTQ was repeated continuously throughout each step of the multidimensional separation. A more detailed description of LTQ-Orbitrap settings is described by Yates et al.[20] Application of mass spectrometer scan functions and HPLC solvent gradients were controlled by the Xcalibur data system.

Analysis of Tandem Mass Spectra. MS/MS were analyzed using the following software analysis protocol. MS/MS remaining after filtering were searched with the SEQUEST algorithm[23] against a database of *Saccharomyces cerevisiae* ORFs downloaded from the Saccharomyces Genome Database (SGD) on March 3, 2005. This database was concatenated to a decoy database in which the sequence for each entry in the original database was reversed to estimate the false positive rate from the search.[24] No enzyme specificity was considered for any search. Two separate SEQUEST parameter files were prepared, and SEQUEST was run twice on each of the ms2 files to separately sequence peptides that were either ^{14}N- or ^{15}N-labeled. SEQUEST results were assembled and filtered using the DTASelect2.0 program, an improved version of DTASelect[25] that uses a linear discriminant analysis to dynamically set XCorr and DeltaCN thresholds for the entire data set to achieve a user-specified false positive rate (5% in this analysis). The false positive rates are estimated by the program from the number and quality of spectral matches to the decoy database.

Quantification of the Relative Protein Abundances. DTASelect2.0 output files were submitted to CenSus as described in Venable et al.[9] to calculate the relative protein abundance differences based on reconstructed ion chromatograms (i.e., ^{14}N/^{15}N CenSus Ratio). The same DTASelect2.0 output files were also used to calculate normalized spectral counts (NSpC) for each protein k:

$$(\mathrm{NSpC})k = \frac{(\mathrm{SpC}/L)_k}{\sum_{i=1}^{N} (\mathrm{SpC}/L)_i}$$

in which the total number of tandem MS spectra matching peptides from protein k(SpC) was divided by the protein's length (L), then divided by the sum of SpC/L for all N proteins identified. Spectral counts serve as a parameter for estimating protein abundances and have been used for calculating relative protein abundance differences.[8,26] Normally spectral counts are merged among all experiments to average variation and increase the number of significantly detected peptides. In this study, relative protein abundance differences based on normalized ^{14}N- and ^{15}N-spectral counting were calculated in each MudPIT experiment separately and compared to relative quantification of the extracted ion chromatograms from each labeled and unlabeled peptide pair using the Census algorithm. For the purpose of this paper, this is an effective way to compare spectral counting versus stable isotope labeling since both methods are applied to the same data set which eliminates run to run variation and offers insight into the strengths and weaknesses of both quantitation methods.

Normalization and Statistical Data Analysis. Calculated spectral counts in each experiment were normalized based on protein length and a total number of spectra detected per experiment. In addition, linear normalization was used in those cases, when the median of the protein abundance ratios (calculated using normalized spectral counts) was not equal to one. In stable isotope labeling based quantification, nor-

Table 2. Number of Proteins Identified in the Study[a]

threshold	wild-type strain (9)[b]	Δsnf1 (3)	Δsnf4 (3)	Δsnf1Δsnf4 (3)
two-peptide	1853 (2029)[c]	1808 (2111)	1855 (2124)	1477 (1636)
one-peptide	3907 (4100)	3299 (3580)	3359 (3641)	2954 (3113)

[a] A total of 5% false positive spectra threshold (determined by DTASelect2.0) was used; protein was included in this table and in the analysis, if it was identified in 2 out of 3 biological replicates. [b] Number of MudPIT experiments merged. [c] The number of identified proteins based on only unique peptides; the number in parentheses indicates the number of proteins identified based on all (unique and nonunique) detected peptides.

malization is performed using the CenSus algorithm (Venable et al.[9]). A CenSus algorithm was used to correct inaccurate ^{15}N and ^{14}N ratios based on errors in sample mixing. This correction is based on the assumption that the natural log of all ratios will form a Gaussian distribution and the median of all ratios in a properly mixed sample should be equal to zero.

These computational steps assured that quantified relative protein abundance ratios between mass spectrometry runs could be fairly compared. Outputs generated by spectral counting and stable isotope labeling were subjected to comparative analysis. A *t* test using a threshold of $P < 0.05$ was used to describe biological variance. A hypergeometric distribution test was used to define biological processes, in which the proteins, with significantly ($P < 0.05$) changed abundances, were enriched (hypergeometric test: $P < 0.01$) in response to the disruption of the Snf1 kinase complex.

Results

Physiological Profile of Strains Used in this Study. Three *S. cerevisiae* yeast deletion strains, Δsnf1, Δsnf4, and Δsnf1Δsnf4, and the wild-type strain were cultivated in biological triplicates in glucose-limited chemostat cultivations at a dilution rate of 0.1 h^{-1}. The only nitrogen source used in these cultivations was ammonium sulfate: ^{15}N-labeled ammonium sulfate for the wild-type strain and ^{14}N-labeled ammonium sulfate for the deletion strains. At steady state, yeast performed only respiratory metabolism, and the biomass yield on glucose was found to be 0.50 g/g for the wild-type strain and 0.45 g/g for the Δsnf1, Δsnf4, and Δsnf1Δsnf4 strains. It was assumed that total protein was constant in the deletion as well as wild-type strains since no major differences in morphology or distribution throughout cell cycle phases were observed while performing microscopy screening. The Δsnf1, Δsnf4, and Δsnf1Δsnf4 strains were compared to the wild-type strain by analyzing them in 1:1 total protein mixtures derived from one of the 99% ^{14}N-labeled deletion strains and 99% ^{15}N-labeled wild-type strain (details in Materials and Methods).

Characterization of Identified Proteome. One sample of each yeast lysate was generated and analyzed using MudPIT. The total number of proteins identified per each strain was calculated by merging spectral counting data from three biological replicates (Table 2). On average, 1600 proteins based on a two-peptide threshold (and 2300 proteins based on a one-peptide threshold) were identified per one 24 h MudPIT experiment. The number of identified proteins in a given time frame found in this study was compatible with the latest achievements that were generated using other global protein identification technologies like GelC-MS/MS or three-dimensional LC-MS/MS.[27,28] A previous immunoblotting TAP Western study (referred to as the "control study" in this report)[29] has estimated that over 4000 proteins are expressed in asynchronously growing yeast cultures. This suggested that using a two-peptide per protein threshold we identified less than half of the possible proteins present in the global proteome mixture

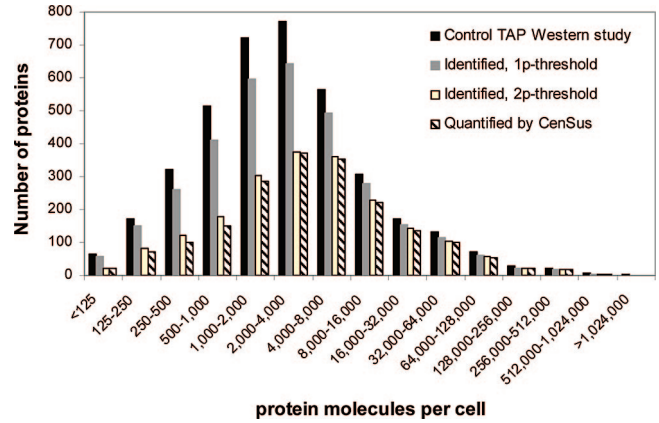


Figure 1. Numbers of identified and quantified proteins based on only unique peptides. The black column represents the control data set from the TAP Western study[29] and includes 3868 proteins. The gray column represents in this study identified the proteome based on a one-peptide threshold and includes 3257 proteins. The white column represents the identified proteome based on a two-peptide threshold and includes 2019 IDs. The column with an angled stripe pattern represents the stable isotope labeling (CenSus) quantified proteome data set and includes 1910 proteins. (Numbers of identified and quantified protein data sets did not match numbers presented in Table 2 and Table 3 because not all proteins identified and quantified in this study were identified and quantified in the TAP Western control study as well).

analyzed. Two factors that may limit our ability to identify all possible proteins in the sample are: (i) our analytical setup, which includes protein extraction, protein digestion, and peptide fractionation methods, failed to identify the subsets of proteins with unusual chemical properties; and (ii) the dynamic range of the lysate is too large to be comprehensively analyzed in one run using our current two-dimensional chromatography strategy.

In addition to SEQUEST and the DTASelect2.0 algorithm used in this study, we also examined the effects of considering only unique peptide identifications. All peptides that were found to match the sequences of two or more proteins were discarded from the analysis and not used to define the list of identified or quantified proteins. This additional filtering step reduced the list of identified proteins by 5–14% per experiment (Table 2) but offered significant benefits in data quality as discussed below.

In prior studies, dynamic range, sensitivity, and sequencing speed were discussed as being the major parameters limiting the complete identification of the yeast proteome.[27,30] In our study, assuming that the abundance of proteins was distributed over the same range in the cells throughout their exponential[29] and steady state (chemostat cultivations) growth phases, we determined the sensitivity and the dynamic range of MudPIT analysis performed in this study (Figure 1). A total of 100 μg of total protein per sample per strain was used in each MudPIT

Table 3. Results of Quantitative Proteome Analysis

method	wild-type strain	Δsnf1	Δsnf4	Δsnf1Δsnf4
total quantified by CenSus[a]	2388	1954 (1816)	2012 (1862)	1697 (1432)
significantly changed by CenSus[b]		381	393	352
total quantified by spectral counting[a]	1798	1535 (1404)	1478 (1365)	1261 (1090)
significantly changed by spectral counting[b]		175	161	109

[a] Calculated at least from one biological replica, if two-unique-peptide pairs (four peptide/protein) were available. A total of one-unique-peptide pair (two peptides/protein) per biological replica was included into quantitative analysis, if this protein was quantitated at least in two biological replicates. A number of proteins, which were quantified at least in two biological replicates, was indicated in parentheses. [b] Significantly changed proteins were those in which protein expression levels deviated less than $P < 0.05$ among biological replicates.

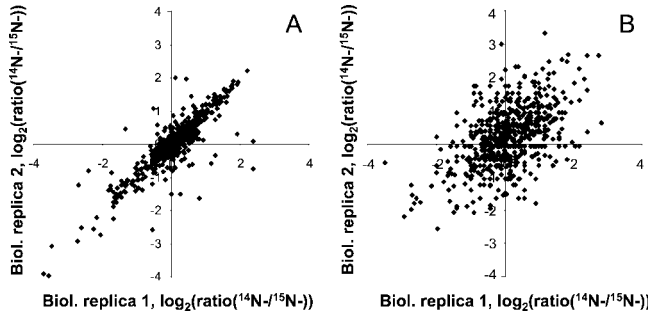


Figure 2. Correlation of relative protein abundance differences calculated for two biological replicates of the Δsnf1Δsnf4 strain versus the wild-type strain. Plot A presented comparison of relative protein abundance differences calculated for two biological replicates by using the stable isotope labeling approach (649 IDs). Plot B presented a comparison of relative protein expression differences calculated for the same two biological replicates by using spectral counting (654 IDs). In this figure, presented spectral counting results are based on 10 spectra (median) and 22 spectra (average) per biological replicate.

run. In agreement with de Godoy calculations,[27] it was assumed that the lowest-abundance proteins were present at a concentration of about 20 fmol in every sample analyzed. Results showed that 40% of the proteins were detected in each range of protein abundances from 40 to 2000 molecules per cell (Figure 1). A total of 77% of proteins were detected in each range spanning protein abundances from 4000 to 512 000 molecules per cell. A default two-peptide minimum per protein was used to derive these numbers of identified proteins. If a one-peptide per protein threshold was used, 84% of proteins were identified in the protein abundance range from the lowest up to 512 000 molecules per cell. A one-peptide threshold dramatically increased the number of detected proteins for the lower range of protein abundances (lowest up to 8000 molecules/cell)—this fraction of the proteome, based on the control study,[29] covers 80% of the total yeast proteome.

The identified proteins were also analyzed to look for trends in other characteristics such as protein pI or protein mass. These parameters did not appear to have a significant impact as the identified proteins had a wide range of masses and pIs (data not shown). An additional pool of 650 proteins, beyond the list of proteins quantified in the control study,[29] was identified indicating that a differently expressed protein pool might have been present when cells were pregrown in the set growth rate chemostat cultivations compared to the exponential growth cultivations. The main limitation for the reliable identification of the complete proteome was found to be the inability to identify more than one peptide from the lowest-abundance proteins. Overall, the result of this study showed that by using the Orbitrap platform for generation of high mass accuracy data and DTASelect2.0 we could begin to sample the lowest-abundance proteins.

Relative Protein Expression Differences Quantified Using Two Algorithms. Two quantitative approaches, spectral counting and stable isotope labeling, were used to assess relative protein expression differences between the Δsnf1, Δsnf4, and Δsnf1Δsnf4 strains and the wild-type strain. While spectral counting uses all identified unique peptide spectra for a given protein to calculate a relative protein abundance, the stable isotope labeling approach uses individual ^{14}N- and ^{15}N-labeled unique peptide pairs and quantitates relative peptide abundances to infer average relative abundance of a corresponding protein (performed by CenSus).[9] These two features constitute the main differences between the CenSus and the spectral counting calculated outputs.

The "total quantified" proteome was determined by merging relative protein abundance differences calculated for each of the three biological replicates using stable isotope labeling or spectral counting. One third of the translated yeast proteome was quantified using the stable isotope labeling approach (Table 3). Using the same four-peptide/protein threshold for both of the quantitative approaches, stable isotope labeling (compared to spectral counting) quantified more proteins in the study. Table 3 shows that twice as many significant changes in protein expression could be found using isotope labeling compared to spectral counting. Additional evidence supporting the idea that stable isotope labeling outperformed spectral counting under these study conditions can be seen in Figure 2. The spectra detected in the same two biological replicate MudPIT experiments were used for a comparative relative protein abundance quantification study using stable isotope and spectral counting approaches. Only proteins, which were quantified based on two-peptide pairs (four-peptide/protein) per MudPIT run, were used in the graphical representation of the comparison (Figure 2). These results indicated that in complex protein mixtures, where protein sequence coverage was frequently low and relatively small differences in protein expression were expected, the calculated relative protein expression differences appeared to be more accurate and reproducible when stable isotope labeling was used compared to spectral counting.

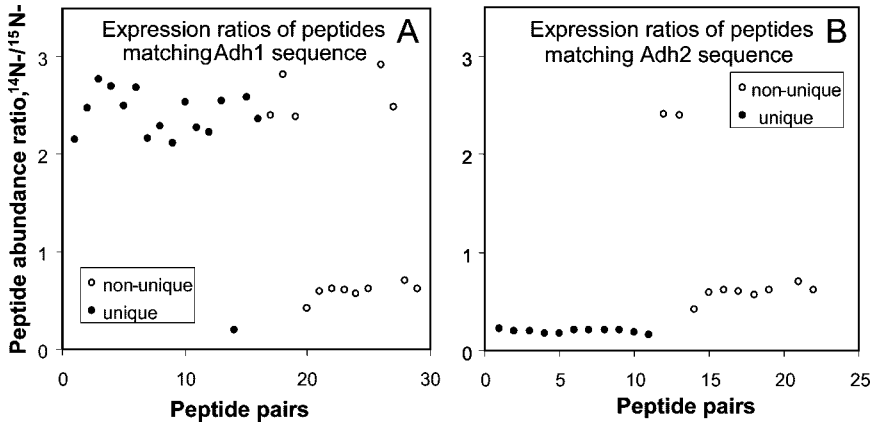


Figure 3. Peptide abundance ratios quantified by CenSus for Adh1 and Adh2 in the Δsnf1 strain versus the wild-type strain. Plots presented relative peptide abundance ratios (differences) quantified for Adh1 (plot A) and Adh2 (plot B) proteins in one of the biological replicates. Values that corresponded to relative peptide abundance differences calculated based on unique (to particular protein) peptides were marked with black circles. Values that were calculated based on nonunique peptides were marked with open circles.

CenSus quantified averaged peptide abundance ratios

protein	peptides	Average	St. dev.
Adh1	all	1.9	0.9
	Only unique	2.3	0.5
Adh2	All	-1.6	0.7
	Only unique	-5.3	0.02

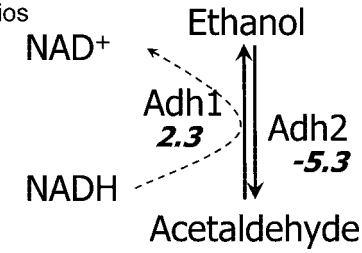


Figure 4. Relative Adh1 and Adh2 averaged protein abundance differences in the Δsnf1 strain versus the wild-type strain when calculated by CenSus and based on all or exclusively on unique peptides only. The relative protein abundance difference was calculated by averaging the determined peptide abundance ratios. The standard deviation was determined based on the variation among calculated peptide abundance ratios. The acetaldehyde−ethanol conversion reaction presented directionality and relative protein abundance levels calculated for Adh1 and Adh2 proteins.

Protein quantification based on spectral counting was found to be largely improved when spectra generated from multiple experiments were averaged and merged so that abundance values could be determined from a larger number of spectra per protein. Larger numbers of proteins were quantified using stable averaged spectral counts (2373 proteins for the wild type, 2057 for the Δsnf1, 2081 for the Δsnf4, and 1763 for the Δsnf1Δsnf4) as compared to the use of spectral counting in this study (Table 3), and averaged spectral counts from biological replicates generated a more accurate measurement of relative protein abundance then those of each individual replicate. Merging of the spectral counting results was not used in this study since it did not allow statistical evaluation of the quantitative data.

A detailed comparison between our quantitative analysis and the absolute protein levels determined in the control study showed that ∼37% of the proteins were quantitated in each range of protein abundances from lowest up to 4000 molecules per cell, and ∼74% of the proteins were quantitated in each range spanning from 4000 up to 512 000 molecules per cell. Similar profiles, ∼25% of low-abundance proteins and ∼70% of high-abundance proteins, were relatively quantified when spectral counting was used.

The result of this study presents the largest number of quantified global yeast proteome by mass spectrometry to date. Stable isotope labeling accurately and reproducibly quantified proteins even when only a small number of spectra per protein (generated from a single mass spectrometry experiment) were identified. The stringent statistical evaluation of biological variance derived among single MudPIT experiments was suitable to use in this study, and it strengthened the presence of true biological changes substantially.

Elimination of Nonunique Spectra Improves the Quantitative Proteome Data Set. Our results showed that by removing nonunique peptides from the experiment the number of identified proteins was reduced by 5–14% per MudPIT run. Importantly, this led to a significant improvement in the ability to identify statistically meaningful differences in protein expression among proteins identified only by unique peptides. The data from the same mass spectrometry experiment were analyzed with and without the inclusion of nonunique peptides using both stable isotope labeling and spectral counting. A total of 3% of proteins quantified by stable isotope labeling and a total of 12% of proteins quantified by spectral counting were found to have different relative protein expression levels depending on whether nonunique spectra were included in the analysis (data not shown).

Significant improvement of quantitative data after nonunique peptides were excluded was illustrated with the following example (Figures 3 and 4). This example focused on the relative quantitation of the two highly homologous proteins, Adh1 and Adh2 (89% sequence similarity). After digestion, Adh1 and Adh2 yielded a combination of unique and shared peptides that were identified and relatively quantified. Figure 3 presents the ratios of quantified Adh1 and Adh2 peptides when the spectra of the Δsnf1 strain were compared to the spectra of the wild-type strain. From this figure, it could be clearly seen that the unique peptides clustered at a specific ratio for each protein, while

the shared, nonunique peptides, representing a mixed population of both Adh1- and Adh2-derived peptides, clustered at a different, inaccurate ratio. The ultimate result of including or excluding these nonunique peptides in the quantitative analysis can be seen in Figure 4. The Adh1 and Adh2 averaged protein abundance ratios were inaccurate (showed a large standard deviation) when nonunique peptides were included and were more accurate (displayed a lower standard deviation) when nonunique peptides were excluded. To interpret these results in a biological context, quantitative data based on only unique peptides clearly showed that acetaldehyde−ethanol conversion was shifted toward ethanol production when the Δsnf1 strain was compared to the wild-type strain. This result correlated with prior data and the fact that the Δsnf1 strain cannot grow on ethanol.[31] In contrast, when nonunique peptide spectra (causing the large standard deviation) were included in the quantitative data sets for the Adh1 and Adn2 proteins, no meaningful biological conclusions could be derived about ethanol oxidation. Overall, when nonunique peptides were discarded from the data sets, the calculation of averaged protein expression differences and their standard deviations were much improved: 3% more proteins (for the CenSus case) were found to have significantly changed ($P < 0.05$) expression in the mutant strains compared to the wild-type strain and could be included in further biological analyses.

Correlation between Stable Isotope Labeling and Spectral Counting Analyses. We next examined how well the proteins quantitated by either spectral counting or stable isotope labeling could be correlated with each other. For this analysis, we focused on the quantitative proteomics analysis of the Δsnf1 strain. A pool of significantly changed proteins ($P < 0.05$) based on spectral counting and stable isotope labeling was used in this comparison (Table 3). A total of 79 proteins were found to have changed expression ($P < 0.05$) levels when quantified by both approaches, and only three of these proteins (Ybr078w, Ydr505c, Ygl103w) correlated poorly. The quantitation of these three proteins proved to be difficult by either method due to the low number (~7) of spectra detected per protein per strain. The poor spectra quality for these three outliers prevented us from concluding whether stable isotope labeling or spectra counting was more successful in these cases.

Among the remaining 302 proteins, for which expression was determined to be significantly changed based on CenSus output, 65 were not quantified by spectral counting due to too few (1 or 2) spectra identified per protein per strain in each experiment. 242 out of the 302 proteins were quantified using less than ten-peptide pairs per protein, further highlighting the ability of stable isotope labeling to quantify proteins based on fewer peptides than spectral counting. Several biologically meaningful and differentially expressed proteins were found in this subset. Some of these proteins, such as the long-chain fatty acyl-CoA synthetase Faa1 (Yor317w), hexokinase Hxk2 (Ygl253w), and glycogen synthase Gsy1 (Yfr014c), were identified as having significantly changed expression (in the Δsnf1 strain compared to the wild-type strain) using the stable isotope labeling approach. However, these proteins were not found to be reproducibly quantified among biological replicates using spectral counting due to the low number of identified spectra per each of these proteins. The Faa1, Hxk2, and Gsy1 were predicted to be affected in the Snf1−kinase complex disrupted strains.[14,31] These findings demonstrate that by using the stable isotope labeling approach the list of quantified proteins was enhanced, and more biologically important findings were revealed, even from low numbers of spectra.

A total of 96 proteins were found to be significantly changed when quantified by spectral counting alone. Seventy-seven (out of 96) proteins were quantified based on less than ten-peptide pairs per protein. Twenty-nine (out of 96) proteins were quantified based on five-peptide pairs per protein, and these were found to have opposite abundance ratios compared to the stable isotope labeling approach. Overall, the results indicated that there is a good correlation between stable isotope labeling and spectral counting, except when spectral counting is based on low numbers of spectra.

Biological Evaluation of the Quantified Proteome. Our results suggested that for the data set used in this study stable isotope labeling was more effective and reproducible than spectral counting for discerning protein expression differences in the mutant yeast strains. On the basis of this observation, we focused on the CenSus-produced quantitative results, emphasizing the subset of proteins that had significant ($P < 0.05$), thus biologically relevant, protein expression changes when the proteome of the three mutants was compared to the proteome of the wild-type strain. A reproducible protein expression change across multiple biological replicates strongly supported the conclusion that a true biological phenomenon was being observed instead of a stochastic, insignificant event.

The CenSus-generated list of 2388 proteins was categorized by GO annotations to determine their connections to different biological processes, molecular functions, and localizations. We found that the subcellular distribution of the quantified (and identified) proteins was not significantly different from the subcellular distribution of the entire yeast proteome. This suggests that the protein extraction method using the urea−invitrosol mixture based protocol was unbiased and contained soluble and membrane-associated proteins. The hypergeometric distribution analysis test was applied to determine whether the enrichment of proteins with significantly changed expression was present among certain GO biological processes and GO molecular function categories. The test was performed in relation to the total pool of 2388 quantified proteins, and the results were summarized in the Table 4.

Enriched protein expression changes (hypergeometric test: $P < 0.01$) within carbon metabolism and respiration GO biological process categories were expected since the list of genes within these processes are known to be transcriptionally regulated by Snf1.[31] Enriched protein expression changes within the nucleic acid metabolic process group were also expected as there is an increasing number of reports describing the role of Snf1 in transcription regulation through histone modification and chromatin remodeling.[16,32] Our global scale proteomics study also indicated potentially new areas of Snf1-mediated regulation. For example, little was known about the Snf1's involvement in the regulation of amino acid metabolism. Table 4 shows that the enriched number of proteins with significant expression changes was found within the amino acid and derivative metabolic process group for all three mutant strains. This indicated that amino acid metabolism was highly linked to the Snf1 kinase's function under these experimental conditions. Further studies are required to determine the links. The oxidoreductase activity GO molecular function category was found to be enriched (hypergeometric distribution test: $P < 0.01$) with significantly changed expression having proteins, in the three mutant strains compared to the wild-type strain. These proteins were distributed among the GO biological

Table 4. List of Significantly Changed Proteins in the Δsnf1, Δsnf4, and Δsnf1Δsnf4 and the GO Process Categories, in which These Proteins Were Found to Be Enriched[a]

	Δsnf1		Δsnf4		Δsnf1Δsnf4	
GO process	*P* value	list of proteins[b]	*P* value	list of the proteins	*P* value	list of proteins
generation of precursor metabolites and energy	2E–09	ACS1, ADH2, PGM2, GND2, CIT3, IDP2, CYB2, TDH1, NDE2, GSY2, PET10	1E–11	MLS1, ACS1, ICL1, ADH2, GND2, PGM2, IDP2, SOL4, CYB2, NDE2, CIT3, TKL2, IDP3, TDH1, GPH1, YJL045W	1E–11	ADH2, ACS1, MLS1, TDH1, PGM2, ICL1, IDP3, CYB2, GLC3, IDP2, TSL1, MDH2, CIT3, TPS2, INO1, GSY2, ETR1, MCR1, NTH1, IDH2, HXK2, CIT2, IDH1, AAT2, ALD5
amino acid and derivative metabolic process	2E–09	YAT1, YAT2, ICL2, GAD1, IDP2, CAT2, STR3	6E–11	YAT2, YAT1, GAD1, IDP2, ICL2, CAT2, STR3, GLY1	1E–11	CAT2, IDP2, YAT1, ICL2, YAT2, GLT1, THR4, SER1, IDH2, CPA2, THR1, HOM3, CIT2, HIS5, BAT1, SER3, IDH1, ARG8, ILV3, ARG4, ARO9, ECM40, PRO2, TRP3, ILV1, HIS1, AAT2, ASN2, ILV2, ARO8, ARO2, HIS7, HOM2, GLY1, HIS4, MET22, ARG1, ORT1, ARO4, ARO3, ASN1
carbohydrate metabolic process	2E–05	PGM2, GND2, CIT3, IDP2, TDH1, GSY2	1E–10	MLS1, ICL1, GND2, PGM2, IDP2, SOL4, CIT3, TKL2, IDP3, TDH1, GPH1	3E–10	MLS1, TDH1, PGM2, ICL1, IDP3, GLC3, IDP2, GUT2, TSL1, MDH2, CIT3, TPS2, INO1, GSY2, AMS1, NTH1, IDH2, HXK2, CIT2, IDH1, AAT2
vitamin metabolic process	3E–05	YAT1, YAT2, ADH2, SNZ1, GND2, CAT2, NDE2	3E–06	YAT2, YAT1, ADH2, GND2, SOL4, CAT2, NDE2, TKL2, IDP3	2E–04	ADH2, IDP3, CAT2, GUT2, YAT1, YAT2
DNA metabolic process	2E–04	ACS1, SWR1, MRC1, MYO4	7E–04	ACS1, MSC1, SWR1	–	–
RNA metabolic process, transcription	5E–04	MRC1, MUD2, TFC3, RRP8	2E–05	MUD2, DCS2, CSR2	2E–04	HXK2, PUS6, CSR2
lipid metabolic process	2E–03	SPS19, GPT2, PLB2, FOX2, POT1, ERG10	4E–05	LEM3, IDP3, POT1, FOX2, PLB2, SPS19, ERG10, ATF2	6E–06	ATF2, POT1, PLB2, IDP3, FOX2, FAS1, POX1, FAS2, FAA2, AYR1, SPS19, ERG11, TES1, GPT2, ACC1, ETR1, MCR1
cell cycle	3E–03	MRC1	3E–03	MSC1	8E–04	MSC1, ADY2, SDS24
cellular respiration	–	–	–	–	2E–03	ALD5, CYB2, MCR1
electron transport	–	–	–	–	4E–03	CIT3, ETR1, IDH2, IDH1

[a] The GO process categories, in which a chance of seeing significant protein expression change ($P < 0.05$) was high (hypergeometric test: $P < 0.01$), were summarized in this table. The test was performed in relation to the distribution of 2388 quantified proteins among GO process categories. [b] Only proteins for which expression was ≥2-fold change were included in the table.

process categories, such as generation of precursor metabolites and energy (Adh2, Idp2, Tdh1), electron transport, amino acid (Glt1), lipid, carbohydrate, and vitamin (Gut2) metabolic processes (Table 4). These findings indicated that Snf1 was clearly playing a role as a global regulator in the central metabolism of yeast.[33]

Overall, the study identified protein expression changes that were expected based on prior literature, confirming the high quality of the generated global quantitative proteome data set. It also indicated Snf1 involvement in the regulation of metabolic processes for which Snf1 has not been explicitly discussed before, indicating its broader regulatory role in yeast.

Discussion

Elimination of Nonunique Peptides Improved the Quality of the Quantitative Proteome Data Set Substantially. A significant improvement in the ability to identify statistically meaningful protein expression differences among proteins identified just by unique peptides was illustrated by Figure 3. One major advantage of using CenSus was the ability to visualize the data. Graphical evaluation of the data (relatively quantified peptide expression differences) offered an opportunity to identify unusual distributions of peptide ratios that might limit the ability to accurately determine the protein abundance ratio. This was clearly demonstrated in Figure 3 where the difficulties caused by the presence of redundant (nonunique) peptides in the analysis were readily identified. Thus, manual evaluation of quantitative data, where possible, could lead to a significant improvement in the quality of the quantitative results. For spectral counting, on the other hand, the high variability in spectral counts for different peptides of a protein limited their utility. Meaningful quantitative information required summing up all peptide spectra corresponding to a protein across an entire experiment to generate one abundance value for the protein. In this case, it was not possible to identify unusual distributions of peptide

ratios to assess their affects on the quality of the overall quantitative measurements. In conclusion, the elimination of nonunique peptides improved the quality of the identified and quantified global proteome data sets. This was particularly crucial when the quantification was performed using spectral counting since no information regarding the distribution of the peptide ratios could be obtained.

Benefits Found when Stable Isotope Labeling or Spectral Counting Was Used. The results of the comparative analysis between stable isotope labeling and spectral counting showed distinct benefits in each method. Stable isotope labeling based quantification by CenSus was found to be more sensitive, was highly reproducible between biological replicates, and generated higher precision data for low-abundance peptides and proteins compared to spectral counting (Figure 2). On the basis of these characteristics and the ability to statistically evaluate the distribution of peptide ratios for a protein to assess the precision of the expression level measured for each protein in that experiment, the isotope labeling approach was better able to identify small, but still biologically significant, changes in protein expression levels between the mutants and the wild-type strain.

In the case of spectral counting, when a large number of spectra were identified for each of the proteins, the results of the quantified relative protein expression differences correlated well with stable isotope labeling. The high spectral coverage required by spectral counting could be readily achieved when simple protein mixtures were analyzed or repetitive MudPIT analysis was performed on the same sample, and spectra from multiple analyses could be summed up. In those cases where redundant spectral coverage of proteins could be achieved, spectral counting becomes advantageous since it is easy to use and does not require prior isotopic labeling of the sample.

In conclusion, the study examined the effectiveness of two different quantitative proteomic strategies applied to the same proteomic data set. Significant protein expression changes were found and confirmed using both analysis methods. In those cases where a discrepancy was found between the methods, the discrepancy could typically be explained by too few identified spectra for that protein. Unusual distributions of peptide ratios between samples due to the inclusion of peptides that map to more than one protein (nonunique peptides) or modified peptides were another potential factor capable of skewing calculated protein abundance ratios when either of the quantitative approaches was used.

Acknowledgment. We thank Akira Motoyama for valuable discussions regarding MudPIT setup and suggestions on how to improve chromatography-based separation of peptides. This work was supported by the Danish Research Agency for Technology and Production and National Institutes of Health grants 5R01 MH067880 and P41 RR11823.

Supporting Information Available: Complete lists of proteins, for which abundances were found to be significantly ($P < 0.05$) changed (in the Δsnf1, Δsnf4, and Δsnf1 snf4 versus wild-type strain), based on the stable isotope labeling approach, the number of peptide pairs used for quantification, normalized relative protein abundance differences, and *t* test based significance of calculated abundance differences are included as supplementary data. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Humphery-Smith, I. A human proteome project with a beginning and an end. *Proteomics* **2004**, *4*, 2519–2521.
(2) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422* (13), 198–207.
(3) Washburn, M. P.; Wolters, D.; Yates, J. R., III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–7.
(4) Delahunty, C.; Yates, J. R., III. Protein identifcation using 2D-LC-MS/MS. *Methods* **2005**, *35*, 248–255.
(5) Washburn, M. P.; Ulaszek, R.; Deciu, C.; Schieltz, D. M.; Yates, J. R., III. Analysis of Quantitative Proteomic Data Generated via Multidimensional Protein Identification Technology. *Anal. Chem.* **2002**, *74*, 1650–1657.
(6) Gygi, S. P.; Rist, B.; Gerber, S. A.; Turecek, F.; Gelb, M. H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17* (10), 994–999.
(7) Patterson, S. D.; Aebersold, R. H. Proteomics: the first decade and beyond. *Nat. Genet.* **2003**, *33* (Suppl), 311–323.
(8) Pang, J. X.; Ginanni, N.; Dongre, A. R.; Hefta, S. A.; Opitek, G. J. Biomarker discovery in urine by proteomics. *J. Proteome Res.* **2002**, *1* (2), 161–169.
(9) Venable, J. D.; Wohlschlegel, J.; McClatchy, D. B.; Park, S. K.; Yates, J. R., III. Relative quantification of stable isotope labeled peptides using a linear ion trap-orbitrap hybrid mass spectrometer. *Anal. Chem.* **2007**, *79* (8), 3056–3064.
(10) Saito, A.; Nagasaki, M.; Oyama, M.; Kozuka-Hata, H.; Semba, K.; Sugano, S.; Yamamoto, T.; Miyano, S. AYUMS: an algorithm for completely automatic quantitation based on LC-MS/MS proteome data and its application to the analysis of signal transduction. *BMC Bioinformatics* **2007**, *8*, 15.
(11) Old, W. M.; Meyer-Arendt, K.; Aveline-Wolf, L.; Pierce, K. G.; Mendoza, A.; Sevinsky, J. R.; Resing, K. A.; Ahn, N. G. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* **2005**, *4* (10), 1487–1502.
(12) Kolkman, A.; Dirksen, E. H.; Slijper, M.; Heck, A. J. Double standards in quantitative proteomics: direct comparative assessment of difference in gel electrophoresis and metabolic stable isotope labeling. *Mol. Cell. Proteomics* **2005**, *4* (3), 255–266.
(13) Zybailov, B.; Coleman, M. K.; Florens, L.; Washburn, M. P. Correlation of relative abundance ratios derived from peptide ion chromatograms and spectrum counting for quantitative proteomic analysis using stable isotope labeling. *Anal. Chem.* **2005**, *77* (19), 6218–6224.
(14) Carlson, M. Glucose repression in yeast. *Curr. Opin. Microbiol.* **1999**, *2* (2), 202–207.
(15) Shirra, M. K.; Patton-Vogt, J.; Ulrich, A.; Liuta-Tehlivets, O.; Kohlwein, S. D.; Henry, S. A.; Arndt, K. M. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the *INO1*gene in a *snf1*mutant strain of *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **2001**, *21* (17), 5710–5722.
(16) Lo, W. S.; Duggan, L.; Emre, N. C.; Belotserkovskya, R.; Lane, W. S.; Shiekhattar, R.; Berger, S. L. Snf1--a histone kinase that works in concert with the histone acetyltransferase Gcn5 to regulate transcription. *Science* **2001**, *293* (553), 1142–1146.
(17) Carling, D. AMP-activated protein kinase: balancing the scales. *Biochimie* **2005**, *87* (1), 87–91.
(18) Usaite, R.; Patil, K. R.; Grotkjær, T.; Nielsen, J.; Regenberg, B. Global Transcriptional and Physiological Responses of *Saccharomyces cerevisiae* to Ammonium, L-Alanine, or L-Glutamine Limitation. *Appl. Microbiol. Biotechnol.* **2006**, *72* (9), 6194–6203.
(19) Usaite, R.; Nielsen, J.; Olsson, L. Physiological characterization of glucose repression in the strains with *SNF1*and *SNF4*genes deleted. *J. Biotechnol.* **2008**, *133*, 73–81.
(20) Yates, J. R., III; Cociorva, D.; Liao, L.; Zabrouskov, V. Performance of a linear ion trap-Orbitrap hybrid for peptide analysis. *Anal. Chem.* **2006**, *78* (2), 493–500.
(21) Ferea, T. L.; Botstein, D.; Brown, P. O.; Rosenzweig, R. F. Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. U.S.A* **1999**, *96* (17), 9721–9726.
(22) Verduyn, C.; Postma, E.; Scheffers, W. A.; Van Dijken, J. P. Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* **1992**, *8* (7), 501–517.
(23) Eng, J.; McCormack, A.; Yates, J. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989.
(24) Peng, J.; Elias, J. E.; Thoreen, C. C.; Licklider, L. J.; Gygi, S. P. Evaluation of multidimensional chromatography coupled with

tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. *J. Proteome Res.* **2003**, *2* (1), 43–50.
(25) Tabb, D. L.; McDonald, W. H.; Yates, J. R., 3rd. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* **2002**, *1* (1), 21–26.
(26) Liu, H.; Sadygov, R. G.; Yates, J. R., III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **2004**, *76* (14), 4193–4201.
(27) de Godoy, L. M.; Olsen, J. V.; de Souza, G. A.; Li, G.; Mortensen, P.; Mann, M. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biol.* **2006**, *7* (6), R50.
(28) Wei, J.; Sun, J.; Yu, W.; Jones, A.; Oeller, P.; Keller, M.; Woodnutt, G.; Short, J. M. Global proteome discovery using an online three-dimensional LC-MS/MS. *J. Proteome Res.* **2005**, *4* (3), 801–808.
(29) Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S. Global analysis of protein expression in yeast. *Nature* **2003**, *425* (6959), 737–741.
(30) McDonald, W. H.; Yates, J. R., III. Shotgun proteomics and biomarker discovery. *Dis. Markers* **2002**, *18* (2), 99–105.
(31) Schuller, H. J. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. *Curr.Genet.* **2003**, *43*, 139–160.
(32) Verdone, L.; Wu, J.; van Riper, K.; Kacherovsky, N.; Vogelauer, M.; Young, E. T.; Grunstein, M.; Di Mauro, E.; Caserta, M. Hyperacetylation of chromatin at the *ADH2*promoter allows Adr1 to bind in repressed conditions. *EMBO J.* **2002**, *21* (5), 1101–1111.
(33) Polge, C.; Thomas, M. SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control. *Trends Plant Sci.* **2007**, *12* (1), 20–28.
(34) Van Dijken, J. P.; Bauer, J.; Brambilla, L.; Duboc, P.; Francois, J. M.; Gancedo, C.; Giuseppin, M. L.; Heijnen, J. J.; Hoare, M.; Lange, H. C.; Madden, E. A.; Niederberger, P.; Nielsen, J.; Parrou, J. L.; Petit, T.; Porro, D.; Reuss, M.; van, R. N.; Rizzi, M.; Steensma, H. Y.; Verrips, C. T.; Vindelov, J.; Pronk, J. T. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb.Technol.* **2000**, *26* (9–10), 706–714.

PR700580M

Chapter 4:

Reconstruction of the yeast Snf1 kinase regulatory network reveals its role as a global energy regulator

Renata Usaite, Michael C. Jewett, Ana P. Oliveira, John R. Yates III, Lisbeth Olsson and Jens Nielsen

Submitted

Reconstruction of the yeast Snf1 kinase regulatory network reveals its role as a global energy regulator

Renata Usaite[1,2], Michael C. Jewett[1,3], Ana P. Oliveira[1], John R. Yates III[2], Lisbeth Olsson[1,4], Jens Nielsen[1,4*]

[1] Center for Microbial Biotechnology, BioCentrum-DTU, Technical University of Denmark, Soltofts Plads DTU - Building 223, DK 2800, Kgs. Lyngby, Denmark

[2] Proteomics Mass Spectrometry Lab, Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., CA 92037, La Jolla, USA

[3] Current address: Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, MA 02115, Boston, USA

[4] Current address: Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemigarden 4, SE 412 96, Gothenburg, Sweden

* Corresponding author: Lisbeth Olsson, Department of Chemical and Biological Engineering, Chalmers University of Technology, Kemigarden 4, Gothenburg, SE 412 96, Sweden. Tel. (+31) 772 38 05; Fax. (+31) 772 38 01, e-mail: lisbeth.olsson@chalmers.se

Highly conserved among eukaryotic cells, the yeast AMP-activated kinase (AMPK) Snf1 is a central regulator of carbon metabolism. To map the complete network of interactions around the protein kinase Snf1 and its regulatory subunit Snf4, we measured global gene expression, protein levels, and metabolite levels in wild-type, *Δsnf1*, *Δsnf4,* and *Δsnf1Δsnf4* knock-out strains. Integrating these measurements with global protein-protein-interactions, protein-DNA-interactions and the yeast genome-scale metabolic model reveals that Snf1 plays a far more extensive role in controlling both carbon and energy metabolism than previously understood. Similar to the function of AMPK in humans, our findings show that Snf1 is a low energy checkpoint. Our results indicate that it is possible to use yeast more extensively as a model system for studying the molecular mechanisms underlying the global regulation of AMPK in mammals, failure of which leads to aging and metabolic diseases, such as diabetes and obesity.

AMP-activated kinases (AMPKs) are highly conserved among yeast, plants, and mammals[1] and are central regulators involved in cellular development and survival. Mammalian AMPK, for example, is a master regulator of energy control.[2] Its function is linked to metabolic and aging diseases and it is a potential drug target against obesity and diabetes.[3] Through homology studies, yeast AMPK (Snf1) has been used as a model to study human AMPK. For example, the upstream kinases of Snf1, Elm1, Pak1, and Tos3, helped identify their mammalian counterparts, Lkb1 and CaMKK-β, that activate human AMPK.[4]

Snf1 regulates carbon metabolism during growth on various carbon sources.[5, 6] In a complex with its regulator Snf4 and scaffolding protein Gal83, Snf1 regulates the utilization of alternative carbon sources via the transcription factors (TFs) Mig1 and Cat8[7]. Moreover, two other Snf1 scaffolding proteins Sip1 and Sip2 determine distinct Snf1-substrate specificity and sub-cellular localization.[8]

There is, however, growing evidence that suggests a much broader role of Snf1 as a master regulator of both carbon and energy metabolism, probably in concert with the protein kinase Tor1. Genome-wide transcriptional profiling in yeast batch cultures has identified that active Snf1 is required for more than 400 of 1500 gene expression changes under glucose exhaustion.[9, 10] At the level of protein interactions (BioGRID database),[11] Snf1 associates with 209 proteins, only 10% of which are enriched (hypergeometric test: P = 1.5E-5) within *GO* carbohydrate metabolic process group (e.g. Adr1, Cat8, Sip4, Pho85, Gsy2, Reg1, Glc7). Moreover, Snf1 as well as mammalian AMPK has been found to respond to various nutrient and environmental stresses including oxidative stress,[12] implicating Snf1 as a global regulator in addition to controlling the utilization of various carbon sources.[13] Furthermore, the remarkable structural conservation of AMPKs' heterotrimeric complexes, specific upstream activators and downstream targets (at transcriptional, protein synthesis and degradation, and post-translational levels) in different kingdoms suggests a common AMPK ancestral function as a key regulator of energy homeostasis.[1]

Clarifying the organization and interactions of the Snf1 regulatory network is important for uncovering the complexity of global AMPK function and, ultimately, for using yeast as a model to study the role of AMPK in humans. However, neither transcriptional profiling, protein-protein interactions, nor ancestry

alone can adequately describe the global regulatory role of Snf1. For this, a systems approach combining global measurements across different levels of the cellular hierarchy (mRNAs, proteins, and metabolites) is required. Recently, Ishii *et al.* and Castrillo *et al*. demonstrated the utility of such an approach for mapping the cellular response of *Escherichia coli* and *Saccharomyces cerevisiae*, respectively, to genetic and environmental perturbations.[14, 15] Beyond mapping, we integrated data from genome-wide expression profiling and protein measurements with different networks comprising protein-protein interactions, protein-DNA interactions, and metabolic reaction stoichiometry. This systems approach enabled reconstruction of the Snf1 complex regulatory network.

Results

To construct a map of network interactions controlled by the Snf1 kinase, we first collected a global dataset for wild-type *S. cerevisiae* CEN.PK113-7D and three Snf1 complex knock-out mutants *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* (supporting information (SI) Table 1) grown in triplicate in carbon-limited chemostat cultivations at a fixed dilution rate D = 0.100 h^{-1}.[16] Abundances of gene, protein, and intracellular metabolites were quantified using Affymetrix GeneChip® Yeast Genome 2.0 Arrays,[17] multidimensional protein identification technology followed by quantitation using stable isotope labeling approach,[16, 18] and gas-chromatography coupled to mass spectrometry,[19] respectively. We quantified a total of 5667 transcripts, 2388 proteins, and 44 intracellular metabolites. At a threshold of $P < 0.05$, a total of 1651, 1810 and 2395 mRNAs, 381, 396 and 352 proteins and 20, 14 and 34 metabolites had significantly changed abundance levels in the knock-out *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* mutants compared to the wild-type, respectively (SI Table 2).

A systems approach to mapping Snf1 response pathways. To reveal how the biological system was reprogrammed as a result of deleting *SNF1*, *SNF4*, or both *SNF1* and *SNF4*, we applied several systems-wide methods that integrated our experimental measurements with data from protein-DNA binding,[20, 21] protein-protein interaction databases,[11] and the yeast genome scale metabolic model.[22] First, we used high-scoring subnetwork analysis[23] to identify co-regulatory circuits of directly connected

proteins and regulated genes that are significantly changing as a group in response to the loss of Snf1 kinase activity. *P*-values from a two-tail Student's t-test reflecting the significance of change in transcript levels between each mutant and the wild-type strain were converted into Z-scores using the inverse cumulative normal distribution function. Z-scores were subsequently mapped on to protein and DNA nodes of an interaction network defined by 57680 protein-protein interactions (BIOGRID-Saccharomyces_cerevisiae v.2.0.25)[11] and 10884 protein-DNA interactions,[20] and high-scoring subnetworks of coordinated biomolecular pathways were identified through a simulated annealing algorithm.[23] Second, to integrate our transcriptomics and proteomics measurements in the same analysis, we extended the high-scoring subnetwork analysis by mapping protein abundance data for protein nodes and included interaction edges between mRNA species and their corresponding proteins (see supporting materials and methods). We call this novel approach, which amplifies the significance of coordinated mRNA and protein expression, 'DOGMA subnetwork analysis'. Third, we developed and applied a 'Reporter Effector' algorithm[24] to identify TFs and regulatory proteins whose connected genes were most significantly affected and responded as a group to genetic disruptions of the Snf1 complex (see supporting materials and methods). Here, Z-scores for each effector were calculated based on the average of Z-scores of its adjacent genes (based on gene expression data) in a network of 3246 protein-DNA interactions and 484 effectors collected from ChIP-chip experiments and the YPD database.[20, 21] The cumulative Z-score was corrected for the size of the group. Finally, we applied the Reporter Metabolite algorithm[25] to our gene and protein expression data for discovering metabolic hot-spots that significantly responded to the loss of Snf1 kinase activity (see supporting materials and methods). Here, we queried a network comprising interactions derived from a genome scale metabolic model consisting of 708 enzymes, 584 metabolites and 1175 reactions.[22]

In total, our four different analyses identified 54 significant network interactions ($P < 0.05$) where Snf1 kinase plays a critical role (Fig. 1) and revealed the global regulatory network of the Snf1 kinase. High-scoring subnetwork analysis revealed three subnetworks comprising 301, 363, and 334 nodes and 651, 987, and 834 edges for the *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* mutants, respectively. DOGMA subnetwork analysis identified three networks comprising of 444, 450, and 376 nodes and 766, 740

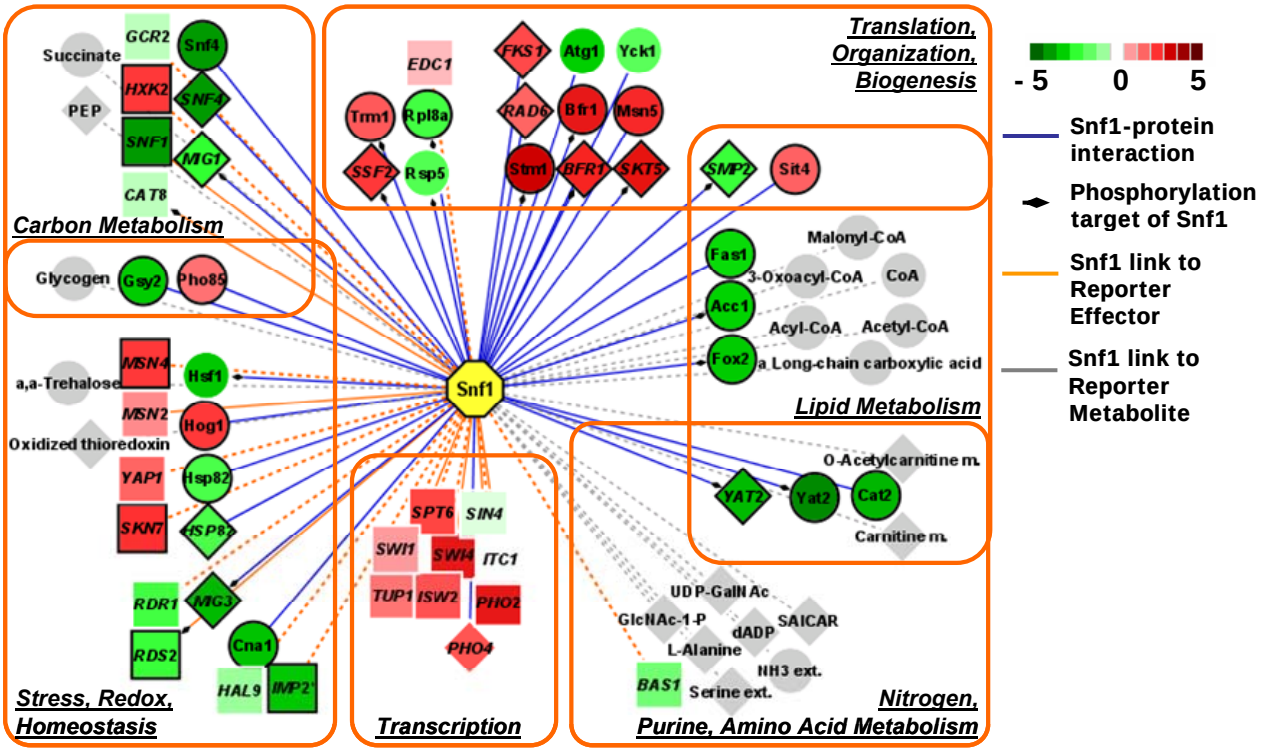


Figure 1. The reconstructed regulatory network of Snf1 kinase. The network was reconstructed by integrating mRNA and protein expression data for the *Δsnf1* mutant versus the wild-type strain with previously reported protein-DNA[20, 21] and protein-protein (BIOGRID-Saccharomyces_cerevisiae v.2.0.25)[11] interactions, and with protein-metabolite interactions provided by the yeast genome-scale metabolic model.[22] The network includes Snf1-interacting proteins that were identified by using high-scoring subnetwork and DOGMA analyses (blue connections to diamonds and circles, respectively). Diamonds show gene expression data and circles show protein expression data, which is colored according to $\log_2$-ratio color scale. The network also includes Reporter Metabolites, around which mRNA or protein abundance changes were significantly concentrated in response to the loss of *SNF1* (grey connections to diamonds and circles, respectively). Reporter Effectors of Snf1 (orange connections to squares) show gene expression data. The Reporter Effectors that are reported to associate to Snf1 kinase[11] are shown using solid orange connections. Nodes with black borders have significantly different ($P < 0.05$) mRNA or protein expression data for the *Δsnf1* mutant versus the wild-type strain. Genes and proteins are named according to the SGDatabase nomenclature. PEP = phosphoenolpyruvate, SAICAR = 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole, UDP-GalNAc = UDP-N-acetyl-D-galactosamine, GlcNAc-1-P = N-Acetyl-D-glucosamine 1-phosphate; m = mitochondrial, ext = extracellular. The reconstructed Snf1 kinase regulatory network based on data from the *Δsnf4* and *Δsnf1Δsnf4* strains versus wild-type strain are presented in SI Figs. 3 and 4. More detailed information describing the subnetwork, Reporter Effector and Reporter Metabolite analyses outputs can be found in SI Tables 3-7.

and 609 edges for the *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* mutants, respectively. Within these networks, focus was given towards first neighbors of Snf1 kinase. From these co-regulated circuits, 12, 19 and 13 first neighbor nodes of the Snf1 kinase were identified using high-scoring subnetwork analysis and 21, 14, and 16 first neighbor nodes of the Snf1 kinase were identified using DOGMA subnetwork analysis for

the *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* mutants, respectively (Fig. 1, SI Figs. 3 and 4). Our analysis identified a few proteins (e.g. Mig1, Snf4, Acc1, Gsy2) that were expected on the basis of previous studies, but we also identified many other proteins interacting with Snf1, including proteins involved in carnitine metabolism (Yat2, Cat2), lipid metabolism (Smp2, Fox2), and the stress response (Hog1, Cna1). Strikingly, ~ 85 % of the first neighbors of Snf1 play a primary role outside the carbon metabolism. Reporter Effector analysis (which identifies regulators whose targets are most significantly impacted by the loss of Snf1) identified 22, 16 and 22 effectors for the *Δsnf1*, *Δsnf4*, *Δsnf1Δsnf4* mutants respectively (Fig. 1, SI Figs. 3 and 4). Of importance, this method identified known TFs phosphorylated by Snf1 (e.g. Cat8), as well as TFs involved in redox, energy (Yap1, Skn7), nitrogen and amino acid (Bas1, Gcn4) metabolism that have not previously been implicated as being regulated by Snf1. Finally, corroborating our analyses, the Reporter Metabolite approach identified key changes within carbon, energy (acetyl-CoA, succinate, glycogen, malonyl-CoA, long-chain carboxylic fatty acids) and redox metabolism (oxidized thioredoxin, NAD+/NADH) (Fig. 1, SI Figs. 3 and 4), highlighting the significant Snf1 involvement in controlling energy homeostasis.

Validation of previously described Snf1 regulation. At the core of the Snf1 regulatory network (Fig. 1), there are players known to be impacted by Snf1 (e.g. glucose repression regulatory cascade nodes): Mig1, Hxk2, Snf4, Cat8. For example, Cat8 regulates the glyoxylate cycle by controlling expression of the *ICL1*, *MLS1*, and *MDH2* genes.[7] Consistent with the importance of the Snf1-Cat8 interaction, the expression of these genes, their coding proteins and the measured level of intracellular glyoxylate were significantly lower in the *Δsnf1* mutant relatively to the wild-type strain ($P < 0.05$) (SI Fig. 5, SI Tables 8 and 9). Overall, the results demonstrated a high degree of correlation for glucose repression related genes, their coding proteins and pathway metabolites. This confirms prior knowledge of glucose repression regulatory cascade studies[7, 13] indicating that transcriptional regulation is the primary glucose repression control mechanism and this validates our approach.

Snf1 is revealed as a low energy check-point. Mammalian AMPK is described as a low energy checkpoint that mediates the energy state of the cell by regulating catabolic and anabolic reactions.[4] Recent reviews have also implied that the yeast Snf1 protein kinase is a global energy regulator.[1, 26] However, global evidence across multiple levels of the cellular hierarchy is still lacking. If this ancestral function is conserved, yeast Snf1 kinase would be expected to induce energy generating and repress energy consuming reactions under carbon limited growth conditions as used in this study. Indeed, our data support this hypothesis. High-scoring subnetwork analysis identified the most significant factors associated with Snf1 to be Fox2, Acc1 and Fas1 (Fig. 1, SI Figs. 3 and 4). To explore how these

pathways were impacted, we built a pathway model linking all measurement types and known protein-protein interactions in Cytoscape[27] (Fig. 2). Genes and proteins (Cta1, Pox1, Fox2 and Pot1) involved

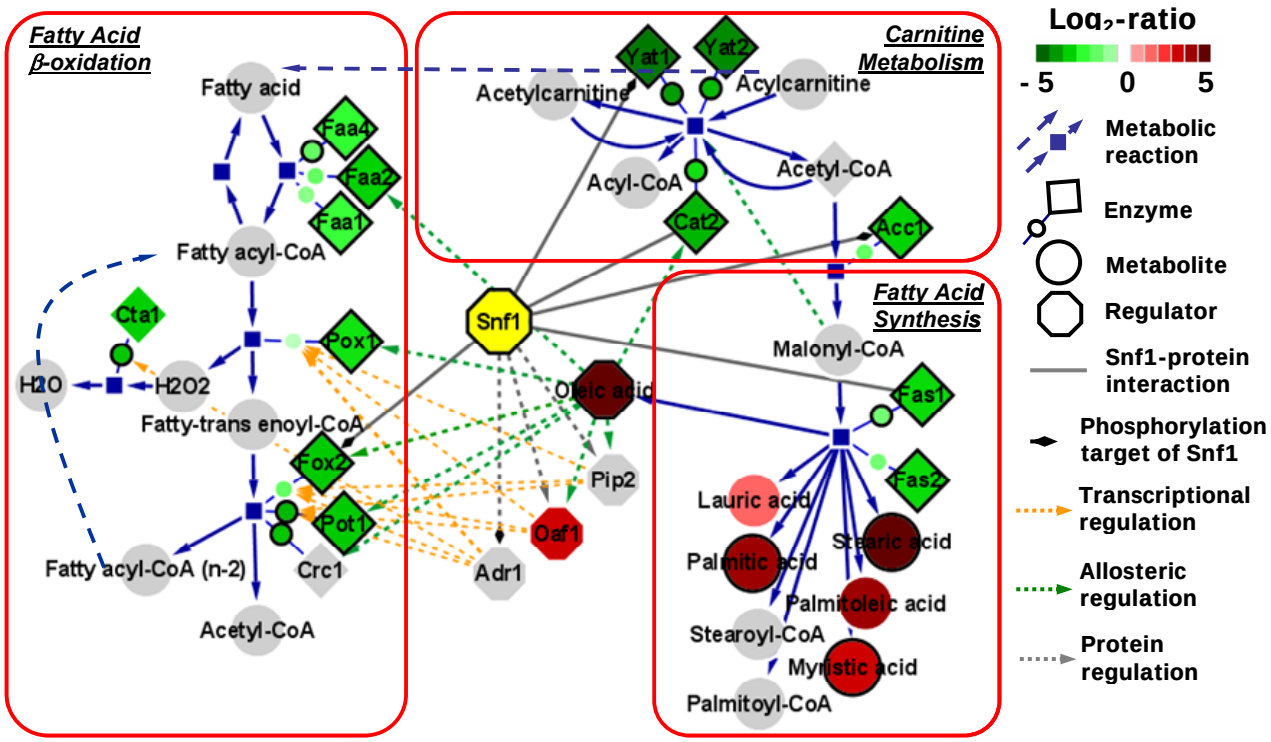


Figure 2. The impact of Snf1 kinase loss on fatty acid metabolism demonstrates its role as a global energy regulator. This figure comprises information of the yeast metabolic network,[22] the reconstructed Snf1 regulatory network, and raw mRNA, protein and metabolite abundance data for the snf1 mutant compared to the wild-type strain (Fig. 1, SI Tables 8 and 9). This figure demonstrates that the loss of Snf1 represses energy producing reactions (e.g. β-oxidation). In other words, the presence of Snf1 induces energy generating reactions, pointing to its role as a low energy checkpoint. Enzymes are mapped using protein (in diamonds) and mRNA (in small circles) expression data, which is colored according to log2-ratio color scale. Available protein and metabolite relative abundance data is mapped on the regulators and metabolites, accordingly. Nodes with black borders have significantly different ($P < 0.05$) expression data for the *Δsnf1* mutant versus the wild-type strain. Grey nodes represent components that were not measured. Five Snf1-protein interactions (solid grey lines) were identified using subnetwork analyses. Colored dotted lines indicate previously reported protein, transcriptional and allosteric regulations.[7, 28]

in *β*-oxidation had lower expression in the Snf1 mutants relative to the wild-type (Fig. 2). In agreement, quantitative metabolome analysis revealed that free fatty acids (oleic, palmitoleic, myristic, palmitic, and stearic acid) accumulated in the Snf1 kinase knock-out mutants relatively to the wild-type strain, rather than being catabolised by *β*-oxidation to generate energy (SI Tables 8 and 9). Loss of Snf1 therefore leads to a decrease in energy producing pathways. This global regulation is likely through Snf1 induction of *β*-oxidation gene expression via the TFs Adr1, Pip2 and Oaf1.[7, 28]

Further evidence for the role of Snf1 in controlling energy metabolism through β-oxidation activation comes from exploring the fatty acid transfer system. In humans, the carnitine metabolic and transfer system plays a rate limiting role for fatty acid β-oxidation. Notably, it is found to be regulated by AMPK through controlling the level of malonyl-CoA.[29] Analogous to humans, and previously not observed in yeast, our data indicate that the carnitine metabolism and transfer system are clearly under the control of the Snf1 kinase. First, carnitine transfer associated proteins Cat2 and Yat2 were identified as key targets of Snf1 by subnetwork analyses and shown to have lower protein and gene expression levels when the Snf1 kinase complex was disrupted. Second, malonyl-CoA and carnitine derivatives were identified as reporter metabolites (indicating that significant changes in gene and protein expression are centred around these metabolites). Third, the malonyl-CoA generating enzyme Acc1 had lower expression in the mutants versus the wild-type strain. Therefore, either through Snf1 association to Cat2 and Yat2, or through regulating production of malonyl-CoA, yeast Snf1 kinase controls the carnitine metabolic and transfer system and consequently fatty acid β-oxidation.

Substantiating the hypothesis that Snf1 is a low energy checkpoint, repression of energy consuming (fatty acid and sterol biosynthesis) pathways was observed under carbon limitation in the mutants relative to the wild-type strain (Fig. 2). Lipid biosynthesis may also be decreased, as the expression of *GUT1* and its coding protein catalyzing the formation of glycerol-3-phosphate (the structural backbone of many lipids) synthesis was decreased. This result is consistent with previous experiments showing *GUT1* expression is activated by Snf1.[28, 30] Given the observed increase in the levels of free fatty acids, serine and glycine, Snf1-mediated control of *GUT1* expression might, in part, be regulating the biosynthesis of phospholipids (SI Fig. 5, SI Table 8). Corroborating this hypothesis, our DOGMA subnetwork analysis identified several components involved in *de novo* lipid biosynthesis such as Acc1, Fas1 and Smp2, which all have AMPK-regulated human analogs.

Energy storing pathways are also downregulated through a Snf1 dependent mechanism. Glycogen is one of the major storage depots of the cell, and both the mRNA expression and protein levels of Gsy2, glycogen synthase, are decreased in the Snf1 mutants. Importantly, Gsy2 and Pho85 (which regulates the activity of Gsy2)[31] are identified in DOGMA subnetwork analysis as being some of the most

important network players impacted by the loss of Snf1 kinase activity. Snf1 control of Msn2,4, which regulates the expression of *GSY1* encoding glycogen synthase,[31, 32] was also implicated (SI Fig. 4, SI Table 9). These results are consistent with the identification of glycogen as a Reporter Metabolite.

Collectively, our systems approach identified that energy generating β-oxidation pathways, energy consuming fatty acid synthesis and energy storing glycogen synthesis pathways were significantly affected by the loss of Snf1 kinase activity, indicating that Snf1 is mimicking the role of its homolog AMPK in mammalian cells as a low energy checkpoint.

Snf1 as a regulator of redox metabolism. Energy balancing not only depends on energy consumption and generation reactions, but also on the redox state of the cell. In this study, the oxidoreductase activity *GO* molecular function category was enriched (hypergeometric test: *P* = 5E-07, *P* = 6E-03) for both mRNA and proteins whose abundance were found to be significantly changed. The systems analysis also implicated multiple genes, proteins, and metabolites that respond to redox change or are involved in redox maintenance (Yap1, Skn7, Msn2,4, Bas1, Pho2, Ssa1, Hsf1, Gts1, Fas1, Fox2, oxidized thioredoxin, NAD^+/NADH, glutathione) (Fig. 1, SI Figs. 3 and 4). Based on a previously observed Yap1-Sip2 interaction,[33] we suggest that Snf1 kinase may contribute to redox homeostasis through a non-glucose repression signalling mechanism. Specifically, the lower expression was found among genes (e.g. *CTT1*, *SOD1*, *SOD2*, *GPX2*) that are involved in maintaining the redox balance and are regulated by oxidative stress through, Yap1 (SI Table 9). The data indicate that a lower oxidative stress, thus, a less induced oxidative stress defense system, is present in the mutants versus the wild-type in this study.

The role of Snf1 in controlling longevity. Noting major changes in carbon, energy, and redox metabolism (previously implicated in aging in yeast, *C. elegans*, and humans), we also considered the impact of Snf1 on cell longevity. Earlier studies have shown that overproduction of Snf1 kinase causes accelerated aging[34] and that a *Δsnf4* strain age slower.[35] Our study results demonstrated that deletion of the Snf1 kinase resulted in lower induction of gluconeogenesis and glycogen biosynthesis (SI Fig. 5), mimicking the biochemical and gene expression profile of a slower aging *Δsnf4* strain described by Gordon *et al.*[35, 36] Furthermore, we find a significant effect on redox metabolism and oxidative stress,

which has also been shown to be linked to aging.[37] Recently the protein kinase Tor1 was linked to aging.[38] Interestingly our analysis shows that Snf1 interacts with many pathways that may also be linked to the protein kinase Tor1 and thus our results suggest that Snf1 and Tor1 both play a role in integrating information on the nutritional state and in concert control energy and redox metabolism, and thereby also aging.

Discussion

Using our systems approach, novel Snf1 targets and their regulation on gene or protein level (in response to the loss of active Snf1) were identified. Highlighting the importance of measuring both mRNAs and proteins, Snf1 neighbours identified only in DOGMA subnetwork analysis implied important post-transcriptional regulation effects. For example, by only identifying Acc1 in DOGMA analysis, our results indicate that Acc1, which is phosphorylated and inactivated by Snf1,[39] is regulated on the protein expression level when the Snf1 kinase complex is inactive. As Pho85 and Gsy2 were also only identified as Snf1 first neighbours in DOGMA subnetwork analysis, combined gene expression and protein level data indicate that post-transcriptional control through Pho85 and Gsy2 regulate glycogen metabolism when Snf1 kinase is inactive. Because of the mode of action of Snf1 kinase activity, measuring both gene expression and protein levels is an appropriate strategy for identifying regulatory structure. Intracellular metabolome data was further used to validate changes in metabolic pathways which, through our network analysis, were identified to be Snf1-controlled (e.g. glyoxylate as described above) (Fig. 2, SI Fig. 5). Measured free fatty acids (SI Table 8) highlighted the importance of available metabolome data and contributed to our understanding of role of Snf1 in controlling lipid metabolism and energy homeostasis. Overall, our results indicate the beneficial contribution of utilizing measurements from multiple cellular levels to reconstruct regulatory networks.

We have combined global data measurements from three levels of the cell (mRNAs, proteins, and metabolites) to construct a regulatory map of Snf1 kinase. By integrating our measurements with different network based interactions and metabolic structures, our systems approach substantially increased information content and minimized the appearance of false positives from global-scale

datasets. The regulatory map reconstructed here identifies new Snf1 targets and confirms previously described connections, validating the power of our systems approach. Overall, our analysis reveals that Snf1 kinase is involved in multiple cellular pathways and, in particular, acts as a low energy checkpoint. Corroborating the data presented in the main text, data from the *Δsnf4* and *Δsnf1Δsnf4* strains are consistent and strengthen the conclusion that Snf1 is a global energy regulator. We suggest that the Snf1 kinase mediates energy and redox balancing for optimal yeast growth: (i) the Snf1 kinase in complex with Snf4 and Gal83 regulates carbon and energy metabolism,[7, 13] (ii) the Snf1 kinase in complex with Snf4 and Sip2 is involved in the regulation of redox balancing and longevity.[33, 35] Thus, our work strengthens the homology in function between yeast Snf1 and mammalian AMPK and opens the door for further using yeast as a model organism to study AMPK as a potential route to better understand and ultimately address metabolic disorders.

Materials and Methods

Strains and cultivation. The *S. cerevisiae* strains used in this study were a prototrophic strain CEN.PK 113-7D (*MAT***a** *MAL2-8*c *SUC2),*[40] its derivatives *Δsnf1* and *Δsnf4* supplied by Koetter (Frankfurt, Germany) and *Δsnf1Δsnf4* generated by Usaite et al.[41] The only genotypic difference among strains used is summarized in SI Table 1. Steady state aerobic chemostat cultures were grown at 30 ºC in 2-liter bioreactors (Braun B) using a dilution rate of D = 0.100 (± 0.005) h^{-1}. Chemostat cultivation ensured that metabolic and regulatory changes observed were specific to disruptions of the Snf1 complex, and not complicated by external effects resulting from the specific mutant physiology (e.g. different growth rates). Detailed description of the cultivations performed and the composition of the carbon limited minimal medium used was summarized previously.[16] After steady state was reached, the cell samples for metabolome, transcriptome and proteome analyses were collected.

Transcriptome analysis and data acquisition. Samples for RNA isolation were taken from chemostat cultivations as previously described.[42] Total RNA was extracted by using a FastRNA Pro Red Kit (BIO 101© Systems, Inc., Vista, CA). The cDNA synthesis, cRNA synthesis, labeling and

cRNA hybridization on the oligonucleotide array Yeast_2.0 (Affymetrix, CA) were performed as described in the Affymetrix GeneChip expression analysis manual that was downloaded from the Affymetrix website in October 2004. The Yeast_2.0 arrays were scanned with the GeneChip® 3000 7G Scanner. The Affymetrix microarray suite v5.0 was used to generate CEL image files of the arrays. The array images were then normalized and the transcript levels of all *S. cerevisiae* probe sets were calculated with the perfect-match model in dChip v1.2.[43] The gene expression data is available on the ArrayExpress. The accession number is E-MEXP-1407.

Proteome analysis and data acquisition. Quantitative proteome data was generated and described by us previously.[16] Briefly, samples for total protein were collected from chemostat cultivations, cells were lysed and total protein was extracted. Protein concentration per sample was determined, and ^{14}N-labeled and ^{15}N-labeled samples were mixed 1:1 by protein weight. A total of 200 μg of total protein was chemically modified and digested by tripsin and endoproteinase LysC. The protein pool digest was analyzed using Multidimensional Protein Identification Technology (MudPIT).[18] A tandem mass spectrum was analyzed and relative protein abundance was quantified using stable isotope labeling as previously described.[16, 44] Complete list of proteins, for which abundances were found to be significantly ($P < 0.05$) changed (in the *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* versus wild-type strain), based on the stable isotope labeling approach is available at http://pubs.acs.org as supplementary data to Usaite et al.[16]

Metabolome analysis and data acquisition. Cells from chemostat cultivations were rapidly quenched according to de Koning and van Dam.[45] Cells were centrifuged at 10,000 x g for 3 minutes in -20°C to separate the cells from the quenching solution. Chloroform: methanol: buffer (CMB) extraction and pure methanol extraction (MEOH) were carried out.[46] Samples were freeze-dried at -56°C using a Christ-Alpha 1-4 freeze dryer.[19]. Amino and non-amino organic acid levels were determined by GC-MS analysis according to Villas-Boas et al[19] except that a Finnegan FOCUS gas chromatograph coupled to single quadrupole mass selective detector (EI) (Thermo Electron Corporation, Waltham, MA, USA) was used. Peak enumeration was conducted with AMDIS (NIST, Gaithersburg, MD) with default parameters, and identification of conserved metabolites was conducted

with SpectConnect,[47] using default parameters and a support threshold of 3. Because SpectConnect is unable to resolve metabolite peaks that have similar MS spectra and are close in time, hand curation of the AMDIS output files was also performed. Samples were normalized by an internal standard chlorophenylalanine (30 mL of a 4 mM solution was added prior to extraction) and by the biomass weight per sample. The identified and quantified metabolites are listed in SI Table 8.

Data analysis. Reporter Metabolite analysis was used to identify metabolic hot spots that significantly responded to the Snf1 kinase complex disruption at the gene or protein expression level.[25] Reporter Effector analysis was used to identify TFs and regulatory proteins whose connected genes were most significantly affected and responded as a group to genetic disruptions of the Snf1 complex.[24] High-scoring subnetwork analysis [23] and in this study developed DOGMA subnetwork analysis were used to identify co-regulatory circuits of directly connected proteins and regulated genes that are significantly changing as a group in response to the loss of Snf1 kinase activity. Reporter Metabolite, Reporter Effector, high-scoring and DOGMA subnetwork analyses are described in detail in supporting materials and methods.

Acknowledgements

We thank James Wohlschlegel, John D. Venable and Sung K. Park from Yates's laboratory for help in generating the proteome dataset. We thank Kiran R. Patil and Intawat Nookaew for valuable discussions on data analysis, and Jerome Maury for discussion on sterol metabolism. This work was supported by the Danish Research Agency for Technology and Production and the National Institutes of Health grants 5R01 MH067880 and P41 RR11823. M.C.J. is grateful to the NSF International Research Fellowship Program for supporting his work.

References

1. Polge, C. & Thomas, M. SNF1/AMPK/SnRK1 kinases, global regulators at the heart of energy control? *Trends in Plant Science* **12**, 20-28 (2007).

2. Kahn, B.B., Alquier, T., Carling, D. & Hardie, D.G. AMP-activated protein kinase: Ancient energy gauge provides clues to modern understanding of metabolism *Cell Metabolism* **1**, 15-25 (2005).
3. Hardie, D.G. AMP-activated protein kinase as a drug target *Annual Review of Pharmacology and Toxicology* **47**, 185-210 (2007).
4. Hardie, D.G. & Sakamoto, K. AMPK: a key sensor of fuel and energy status in skeletal muscle. *Physiology* **21**, 48-60 (2006).
5. Celenza, J.L. & Carlson, M. A yeast gene that is essential for release from glucose repression encodes a protein kinase. *Science* **233**, 1175-1180 (1986).
6. Carlson, M. Glucose repression in yeast. *Curr.Opin.Microbiol.* **2**, 202-207 (1999).
7. Schuller, H.J. Transcriptional control of nonfermentative metabolism in the yeast *Saccharomyces cerevisiae*. *Curr.Genet.* **43**, 139-160 (2003).
8. Vincent, O., Townley, R., Kuchin, S. & Carlson, M. Subcellular localization of the Snf1 kinase is regulated by specific beta subunits and a novel glucose signaling mechanism. *Genes Dev.* **15**, 1104-1114 (2001).
9. Young, E.T., Dombek, K.M., Tachibana, C. & Ideker, T. Multiple pathways are co-regulated by the protein kinase Snf1 and the transcription factors Adr1 and Cat8. *J.Biol.Chem.* **278**, 26146-26158 (2003).
10. DeRisi, J.L., Iyer, V.R. & Brown, P.O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686 (1997).
11. Stark, C. et al. BioGRID: a general repository for interaction datasets. *Nucleic Acids Res.* **1**, D535-D539 (2006).
12. Hong, S.P. & Carlson, M. Regulation of snf1 protein kinase in response to environmental stress. *J.Biol.Chem.* **282**, 16838-16845 (2007).
13. Gancedo, J.M. Yeast carbon catabolite repression. *Microbiol.Mol.Biol.Rev.* **62**, 334-361 (1998).
14. Ishii, N. et al. Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations. *Science* **316**, 593-597 (2007).
15. Castrillo, J.I. et al. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol.* **6**, 4 (2007).
16. Usaite, R. et al. Characterization of global yeast quantitative proteome data generated from the wild type and glucose repression *Saccharomyces cerevisiae* strains: the comparison of two quantitative methods. *J Proteome Res* **7**, 266-275. (2008).
17. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. & Lockhart, D.J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat.Biotechnol.* **15**, 1359-1367 (1997).
18. Washburn, M.P., Wolters, D. & Yates, J.R., 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat.Biotechnol.* **19**, 242-247 (2001).
19. Villas-Boas, S.G., Moxley, J.F., Akesson, M., Stephanopoulos, G. & Nielsen, J. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem.J.* **388**, 669-677 (2005).
20. Harbison, C.T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99-104 (2004).
21. Hodges, P.E., McKee, A.H., Davis, B.P., Payne, W.E. & Garrels, J.I. The Yeast Proteome Database (YPD): a model for the organization and presentation of genome-wide functional data. *Nucleic Acids Res.* **27**, 69-73 (1999).
22. Forster, J., Famili, I., Fu, P., Palsson, B.O. & Nielsen, J. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* **13**, 244-253 (2003).
23. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**, S233-S240 (2002).
24. Oliveira, A.P., Patil, K.R. & Nielsen, J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Systems Biology* **2**, 17 (2008).
25. Patil, K.R. & Nielsen, J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc.Natl.Acad.Sci.U.S.A* **102**, 2685-2689 (2005).
26. Hardie, D.G. AMP-activated/SNF1 protein kinases: conserved guardians of cellular energy. *Nat Rev Mol Cell Biol.* **8**, 774-785 (2007).
27. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504 (2003).
28. Young, E.T., Kacherovsky, N. & Van, R.K. Snf1 protein kinase regulates Adr1 binding to chromatin but not transcription activation. *J.Biol.Chem.* **277**, 38095-38103 (2002).
29. Folmes, C.D.L. & Lopaschuk, G.D. Role of malonyl-CoA in heart disease and the hypothalamic control of obesity. *Cardiovascular Research* **73**, 278-287 (2007).
30. Grauslund, M., Lopes, J.M. & Ronnow, B. Expression of *GUT1*, which encodes glycerol kinase in *Saccharomyces cerevisiae*, is controlled by the positive regulators Adr1p, Ino2p and Ino4p and the negative regulator Opi1p in a carbon source-dependent fashion. *Nucleic Acids Res.* **27**, 4391-4398 (1999).

31. Hardy, T.A., Huang, D. & Roach, P.J. Interactions between cAMP-dependent and SNF1 protein kinases in the control of glycogen accumulation in *Saccharomyces cerevisiae*. *J.Biol.Chem.* **269**, 27907-27913 (1994).
32. Unnikrishnan, I., Miller, S., Meinke, M. & LaPorte.D.C. Multiple Positive and Negative Elements Involved in the Regulation of Expression of *GSY1* in *Saccharomyces cerevisiae*. *J.Biol.Chem.* **278**, 26450–26457 (2003).
33. Wiatrowski, H.A. & Carlson, M. Yap1 Accumulates in the Nucleus in Response to Carbon Stress in *Saccharomyces cerevisiae*. *Eukaryot.Cell* **2**, 19-26 (2003).
34. Lin, S.S., Manchester, J.K. & Gordon, J.I. Sip2, an N-myristoylated beta subunit of Snf1 kinase, regulates aging in *Saccharomyces cerevisiae* by affecting cellular histone kinase activity, recombination at rDNA loci, and silencing. *J.Biol.Chem.* **278**, 13390-13397 (2003).
35. Ashrafi, K., Lin, S.S., Manchester, J.K. & Gordon, J.I. Sip2p and its partner snf1p kinase affect aging in *S. cerevisiae*. *Genes Dev.* **14**, 1872-1885 (2000).
36. Lin, S.S., Manchester, J.K. & Gordon, J.I. Enhanced gluconeogenesis and increased energy storage as hallmarks of aging in *Saccharomyces cerevisiae*. *J.Biol.Chem.* **276**, 36000-36007 (2001).
37. Kaeberlein, M., Burtner, C.R. & Kennedy, B.K. Recent Developments in Yeast Aging. *PLoS.Genet.* **3**, e84 (2007).
38. Kaeberlein, M. et al. Regulation of Yeast Replicative Life Span by TOR and Sch9 in Response to Nutrients. *Science* **310**, 1193-1196 (2005).
39. Shirra, M.K. et al. Inhibition of acetyl coenzyme A carboxylase activity restores expression of the *INO1* gene in a *snf1* mutant strain of *Saccharomyces cerevisiae*. *Mol.Cell Biol.* **21**, 5710-5722 (2001).
40. Van Dijken, J.P. et al. An interlaboratory comparison of physiological and genetic properties of four *Saccharomyces cerevisiae* strains. *Enzyme Microb.Technol.* **26**, 706-714 (2000).
41. Usaite, R., Nielsen, J. & Olsson, L. Physiological characterization of glucose repression in the strains with *SNF1* and *SNF4* genes deleted. *J.Biotechnol* **133**, 73-81 (2008).
42. Usaite, R., Patil, K.R., Grotkjær, T., Nielsen, J. & Regenberg, B. Global Transcriptional and Physiological Responses of *Saccharomyces cerevisiae* to Ammonium, L-Alanine, or L-Glutamine Limitation. *Appl.Microbiol.Biotechnol.* **72**, 6194–6203 (2006).
43. Li, C. & Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc.Natl.Acad.Sci.U.S.A* **98**, 31-36 (2001).
44. Venable, J.D., Wohlschlegel, J., McClatchy, D.B., Park, S.K. & Yates, J.R., 3rd. Relative quantification of stable isotope labeled peptides using a linear ion trap-orbitrap hybrid mass spectrometer. *Anal.Chem.* **79**, 3056-3064 (2007).
45. de Koning, W. & van Dam, K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal.Biochem.* **204**, 118-123 (1992).
46. Villas-Boas, S.G., Hojer-Pedersen, J., Akesson, M., Smedsgaard, J. & Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**, 1155-1169 (2005).
47. Styczynski, M.P. et al. Systematic identification of conserved metabolites in GC/MS data for metabolomics and biomarker discovery. *Anal.Chem.* **79**, 966-973 (2007).

Chapter 5:

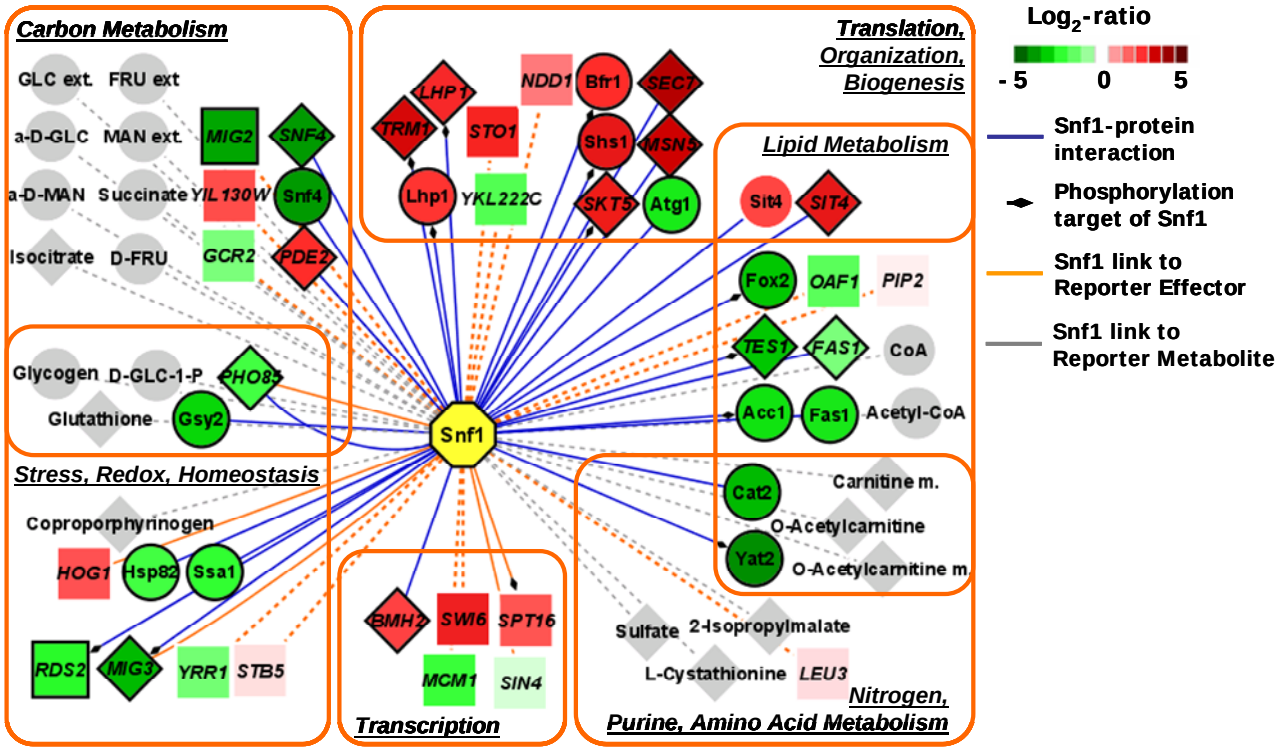
Supplementary Information for the Chapter 4:

Supplementary Figures and Legends

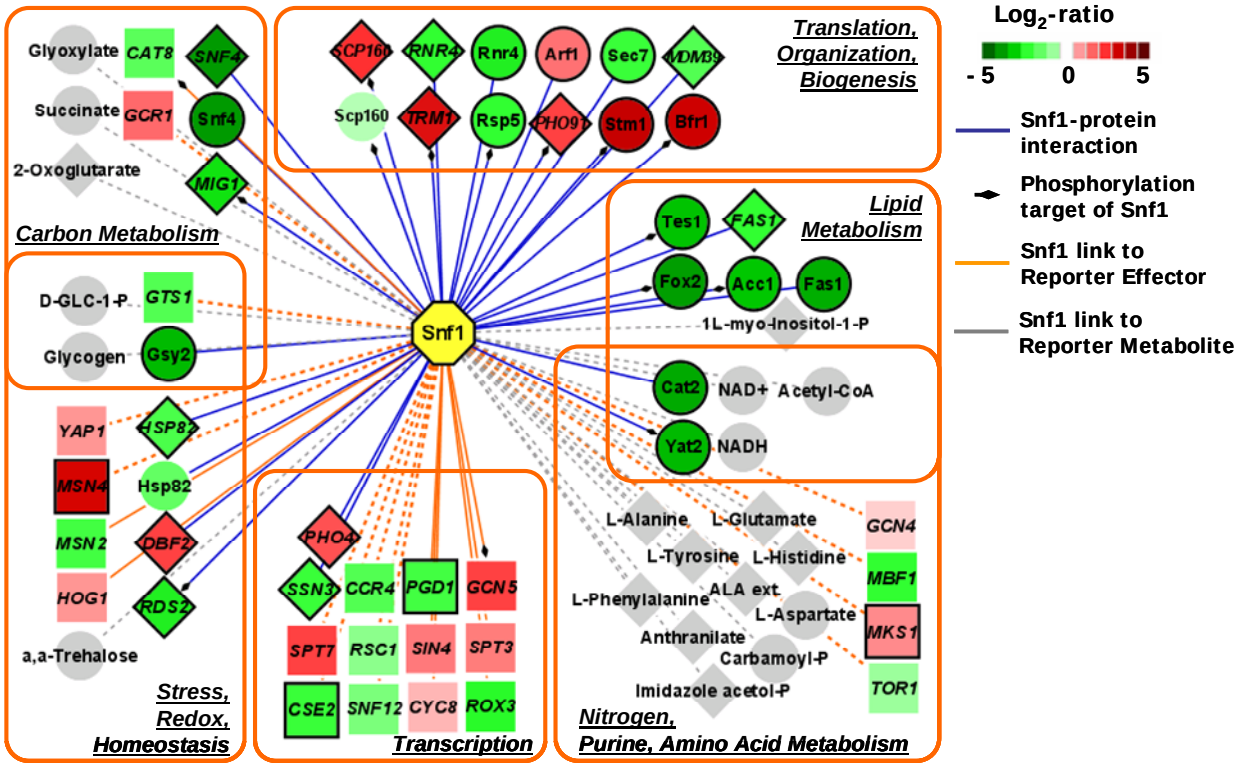
Supplementary Methods

Supplementary Tables

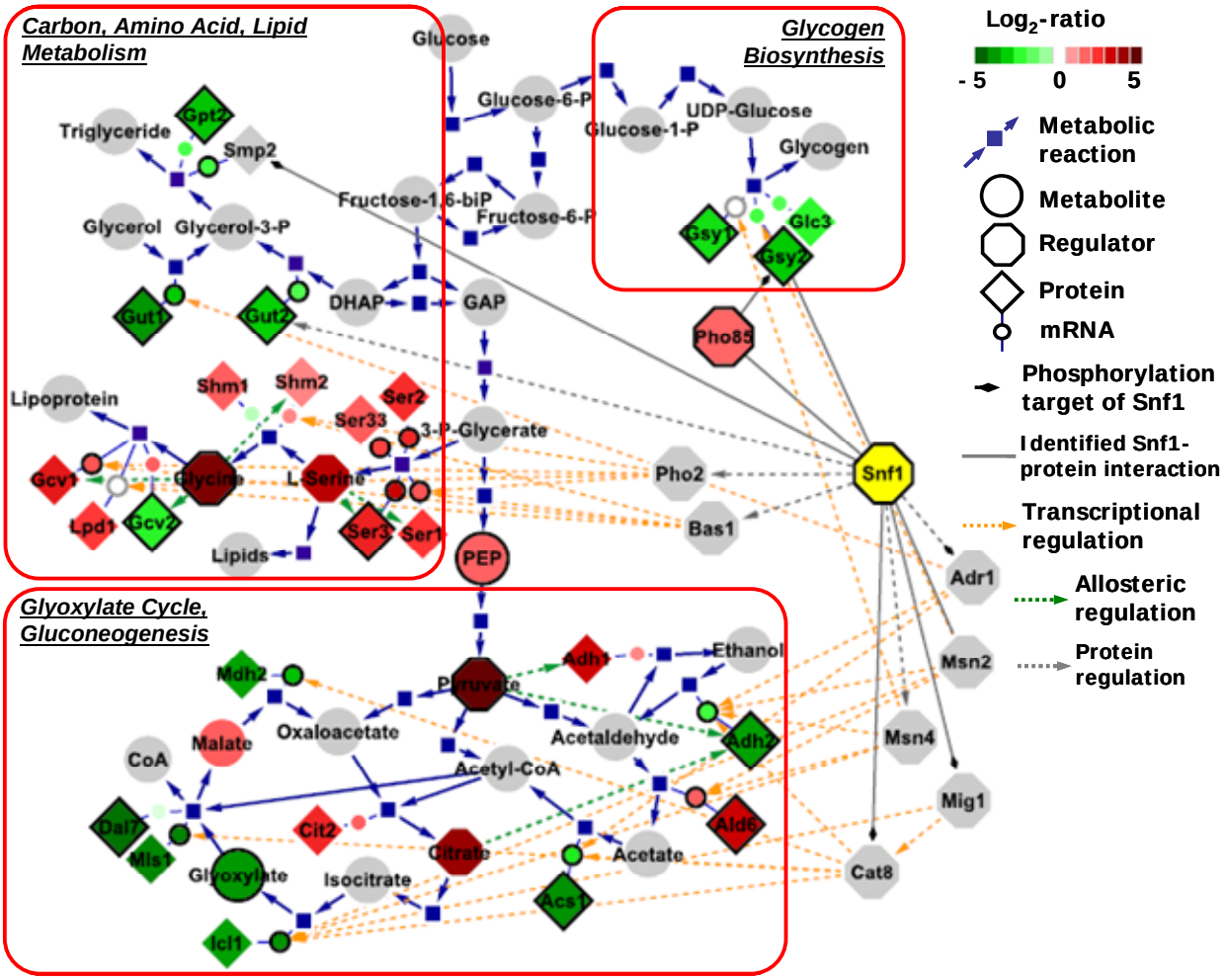
5.1 Supplementary Figures and Legends



SI Figure 3. The reconstructed regulatory network of Snf1 kinase. The network was reconstructed by integrating mRNA and protein expression data for the *Δsnf4* mutant versus the wild-type strain with previously reported protein-DNA (Harbison, *et al* (2004), *Nature*; Hodges, *et al* (1999), *Nucleic Acids Res.*) and protein-protein (BIOGRID-Saccharomyces_cerevisiae v.2.0.25) (Stark, *et al* (2006), *Nucleic Acids Res.*) interactions, and with protein-metabolite interactions provided by the yeast genome-scale metabolic model (Forster, *et al* (2003), *Genome Res.*). The network includes Snf1-interacting proteins that were identified by using high-scoring subnetwork and DOGMA analyses (blue connections to diamonds and circles, respectively). Diamonds show gene expression data and circles show protein expression data, which is coloured according to $\log_2$-ratio colour scale. The network also includes Reporter Metabolites, around which mRNA or protein abundance changes were significantly concentrated in response to the loss of *SNF4* (grey connections to diamonds and circles, respectively). Reporter Effectors (orange connections to squares) show gene expression data. The Reporter Effectors that are reported to associate to Snf1 kinase (BIOGRID-Saccharomyces_cerevisiae v.2.0.25) are shown using solid orange connections. Nodes with black borders have significantly different ($P < 0.05$) mRNA or protein expression data for the *Δsnf4* mutant versus the wild-type strain. Genes and proteins are named according to the SGDatabase nomenclature. GLC, FRU and MAN stand for glucose, fructose and mannose, respectively; ext = extracellular, m. =mitochondrial. More detailed information describing the subnetwork, Reporter Effector and Reporter Metabolite analyses outputs can be found in SI Tables 3-7.



SI Figure 4. The reconstructed regulatory network of Snf1 kinase. The network was reconstructed by integrating mRNA and protein expression data for the *Δsnf1Δsnf4* mutant versus the wild-type strain with previously reported protein-DNA (Harbison, *et al* (2004), *Nature*; Hodges, *et al* (1999), *Nucleic Acids Res.*), and protein-protein (BIOGRID-Saccharomyces_cerevisiae v.2.0.25) (Stark, *et al* (2006), *Nucleic Acids Res.*) interactions, and with protein-metabolite interactions provided by the yeast genome-scale metabolic model (Forster, *et al* (2003), *Genome Res.*). The network includes Snf1-interacting proteins that were identified by using high-scoring subnetwork and DOGMA analyses (blue connections to diamonds and circles, respectively). Diamonds show gene expression data and circles show protein expression data, which is coloured according to $\log_2$-ratio colour scale. The network also includes Reporter Metabolites, around which mRNA or protein abundance changes were significantly concentrated in response to the loss of both *SNF1* and *SNF4* (grey connections to diamonds and circles, respectively). Reporter Effectors (orange connections to squares) show gene expression data. The Reporter Effectors that are reported to associate to Snf1 kinase (BIOGRID-Saccharomyces_cerevisiae v.2.0.25) are shown using solid orange connections. Nodes with black borders have significantly different ($P < 0.05$) mRNA or protein expression data for the *Δsnf1Δsnf4* mutant versus the wild-type strain. Genes and proteins are named according to the SGDatabase nomenclature. GLC = glucose, ALA = alanine, ext = extracellular. More detailed information describing the subnetwork, Reporter Effector and Reporter Metabolite analyses outputs can be found in SI Tables 3-7.



SI Figure 5. The impact of Snf1 kinase loss on central carbon metabolism. This figure comprises information of the yeast metabolic network (Forster, *et al* (2003), *Genome Res.*), the reconstructed Snf1 regulatory network, and raw mRNA, protein and metabolite abundance data for the *Δsnf1* mutant compared to the wild-type strain (Fig. 1, SI Tables 8, 9). This figure demonstrates that the loss of Snf1 represses glycogen biosynthesis, gluconeogenesis and biosynthesis of lipid precursors. In other words, the Snf1 kinase regulates carbon and energy metabolism. Enzymes are mapped using protein (in diamonds) and mRNA (in small circles) expression data, which is coloured according to $\log_2$-ratio colour scale. Available protein and metabolite relative abundance data is mapped on the regulators and metabolites, accordingly. Nodes with black borders have significantly different ($P < 0.05$) expression data for the *Δsnf1* mutant versus the wild-type strain. Grey nodes represent components that were not measured. Seven Snf1-protein interactions (solid grey lines) were identified using subnetwork and Reporter Effector analyses. Coloured dotted lines indicate previously reported and by Reporter Effector analysis identified protein, transcriptional and allosteric regulations (Grauslund, *et al* (1999) *Nucleic Acids Res.*; Young, *et al* (2002) *J Biol Chem*).

5.2 Supplementary Methods

5.2.1 Reporter Metabolite analysis

Gene or protein expression changes in response to *SNF1* or *SNF4* gene deletion were mapped on the genome-scale metabolic model of *S. cerevisiae* (Forster, *et al* (2003), *Genome Res.*) in order to identify metabolic hot spots that significantly responded to the Snf1 kinase complex disruption at the gene or protein expression level (Patil & Nielsen (2005), *Proc. Natl. Acad. Sci. USA*). The genome-scale model of yeast was first represented as a graph in which each metabolite is connected to all enzymes that catalyze a reaction involving that particular metabolite. Each enzyme involved in this graph was then scored based on the significance of the change in the expression level of the corresponding gene or protein. This significance score was calculated by using a t-test and transforming the resulting *P* value to a Z-score using the inverse normal cumulative distribution function. Each metabolite was assigned the average score of its *k* neighbouring enzymes, and this score was then corrected for the background by subtracting the mean and dividing by the standard deviation of average scores of 10,000 enzyme groups of size *k* selected from the same data set. These corrected scores were then converted back to *P* values by using the normal cumulative distribution function. The 10 top-scoring metabolites were identified as reporter metabolites and selected for further analysis. Thus, the reporter metabolites are those around which transcriptional (SI Table 6) or protein expression (SI Table 7) changes are significantly concentrated.

5.2.2 Reporter Effector analysis

The Reporter Effector algorithm is an integrative method that combines the topology of the regulatory network (effector – gene) with gene expression levels, in order to identify those effectors (transcription factors, other regulatory proteins) whose connected genes are most significantly responsive as a group to a perturbation (Oliveira, *et al*, accepted). Gene expression for a total of 484 effectors and 968 their targets, and 3246 protein-DNA interactions collected from ChIP-chip experiments (Harbison, *et al* (2004), *Nature*) and YPD database (Hodges, *et al* (1999), *Nucleic Acids Res.*) were used to perform the Reporter Effector analysis. Each effector was scored based on the average of scores of its adjacent genes (corrected for the size of the group of connected genes), and the high-scoring effectors are termed Reporter Effectors. Reporter Effectors highlight the regulatory pathways affected following a perturbation, and thus uncover the functional links between the perturbation and the consequent regulatory mechanisms invoked in the cell. Many transcription factors and regulators do not respond at transcriptional level *per se* (but through post-translational regulation instead): Reporter

Effector analysis provides a powerful tool for reconstruction of regulatory circuits without *a priori* requirement of change in the transcription level of the regulators. In this study, the Reporter Effector analysis was performed three times using the complete gene expression dataset, using only lower expression, or using only higher expression having gene expression datasets for any two-strain comparison to account for repressor, activator or the dual role having transcription factors. The top-scoring reporter effectors ($P < 0.05$) were selected for further analysis and are presented in the SI Table 5.

5.2.3 Subnetwork analysis

To search for the high-scoring subnetworks that describes highly active regulatory modules of connected proteins and regulated genes that are significantly changing in response to a perturbation, we used the previously proposed simulated annealing algorithm (Ideker, *et al* (2002), *Bioinformatics*) implemented with an additional heuristics: the probability of a certain node being marked visible in the initialization was proportional to (1 – *p*-value). Briefly, the algorithm takes as inputs a graph *G* (i.e., one of the interaction networks) and a list of *p*-values (in this case, from a two-tail Student's t-test, reflecting the changes in transcript/protein levels between each mutant and the reference strain). *P*-values are converted into *z*-scores using the inverse cumulative distribution function. *Z*-scores are then mapped into the graph, and the score of a given subnetwork *SG* is calculated as the average sum of all node-elements of *SG*, corrected for background and for the size of *SG*. In order to find the highest-score subnetwork, a simulated annealing algorithm is used. As referred by Ideker *et al*, and because the problem of finding the highest-score connected subnetwork is NP-hard, it is not guaranteed to find the overall maximum using this algorithm (Ideker, *et al* (2002), *Bioinformatics*). Therefore, each pair of network/data was analyzed 10 times, and the further analysis was based on the results from the merged 10 high-scoring subnetworks.

Here, a large network comprising of 57680 protein-protein interactions and 10884 protein-DNA interactions was used for this analysis. From the BioGRID database (BIOGRID-Saccharomyces_cerevisiae v.2.0.25), a downloaded list of protein-protein interactions was curated by removing duplicated information and by selecting protein physical interactions generated using Affinity Capture-MS, Affinity Capture-RNA, Affinity Capture-Western, Biochemical Activity, Co-crystal Structure, Co-fractionation, Co-localization, Co-purification, FRET, Far Western, Protein-RNA, Protein-peptide, Reconstituted Complex and Two-hybrid methods. This resulted in obtaining 57680 protein-protein interactions, covering 3868 unique proteins. A total of 10884 high-confidence (p-value < 0.001) protein-DNA interactions derived from ChIP-chip data (Harbison, *et al* (2004), *Nature*) were included in high-scoring subnetwork analysis.

High-scoring subnetwork analysis contained two different types of interactions (protein-protein and protein-DNA) and only changes in transcript levels were used to score the nodes. First Snf1 kinase neighbours (Snf1-interacting proteins), that were identified using high-scoring subnetwork analysis, are listed in SI Table 3.

Dogma subnetwork analysis contained three types of interactions: protein-protein, protein-DNA and 'mRNA to protein' translation interactions. Since we quantified both mRNA and protein abundances, the network of bio-molecular interactions were extended to accommodate both changes in transcript and protein levels. Namely, we had a chance to expand network analysis including the translational relationship between each transcript *i* and the corresponding protein *i*. Changes in proteome levels were used to score protein-nodes, while transcriptome data was used to score gene-nodes. Therefore, the resulting high-scoring subnetworks revealed connected circuits being significantly regulated at gene, translation and protein levels. First Snf1 kinase neighbours (Snf1-interacting proteins), that were identified using DOGMA subnetwork analysis, are listed in SI Table 4.

5.3 Supplementary Tables

SI Table 1. Yeast *Saccharomyces cerevisiae* strains used in this study

Strain[a]	Name in text	Genotype	Source/Reference
CEN.PK 113-7D	wild-type	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8[C]	Provided by P. Kötter[b]
CEN.PK 506-1C	*Δsnf1*	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8[C] snf1 (4,1899)::loxP-Kan-loxP	Provided by P. Kötter
CEN.PK 507-5B	*Δsnf4*	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8[C] snf4 (4,966)::loxP-Kan-loxP	Provided by P. Kötter
IBT100072	*Δsnf1Δsnf4*	MATa URA3 HIS3 TRP1 LEU2 SUC2 MAL2-8[C] snf1 (1,1903) snf4 (4,966)::loxP-Kan-loxP	Usaite *et al*[c]

[a] The strains were derived from the parental laboratory CEN.PK background strain (Van Dijken, *et al* (2000), *Enzyme Microb Technol*)

[b] Institut für Mikrobiologie, Frankfurt, Germany

[c] Strain design and characteristics have been previously published (Usaite, *et al* (2008), *J Biotechnol*)

SI Table 2. Numbers of significantly changed proteins, transcripts and metabolites[a]

	Δsnf1	*Δsnf4*	*Δsnf1Δsnf4*	Total quantified
Transcripts[b]	1651	1810	2395	5544
Proteins[c]	381	396	352	2388
Metabolites[d]	20	14	34	44

[a] Significantly changed abundance was determined based on a t-test and a threshold of $P < 0.05$ for the data of the *Δsnf1, Δsnf4* and *Δsnf1Δsnf4* strains versus the data of the wild-type strain

[b] Gene expression data is available on The ArrayExpress (accession number is E-MEXP-1407)

[c] Proteome dataset is previously described (Usaite, *et al* (2008), *J Prot Res*) and available on http://pubs.acs.org

[d] Metabolites, detected in one out of two strains of comparison are also included among significantly changed (SI Tables 8)

SI Table 3. Snf1 kinase first neighbours identified using high-scoring subnetwork analysis

Protein	ORF	Gene expression log_2(fold)[a]			Interaction (BioGRID)[b]
		Δsnf1	*Δsnf4*	*Δsnf1Δsnf4*	
Bfr1	YOR198C	**0.63**			*BA*
Bmh2	YDR099W		**0.42**		AC-MS
Dbf2	YGR092W			**0.42**	*AC-MS*
Fas1	YKL182W		**-0.15**	**-0.50**	SR
Fks1	YLR342W	**0.37**			*SL*
Hsp82	YPL240C	**-0.37**	-0.24	**-0.35**	*SL*
Lhp1	YDL051W		**0.59**		*BA*
Mdm39	YGL020C			**-0.31**	*SGD, PhE*
Mig1	YGL035C	**-0.53**	**-0.72**	**-0.78**	AC-W, 2H, BA, PhE, SR
Mig3	YER028C	**-1.36**	**-1.33**		BA, DL, SR
Msn5	YDR335W		**1.01**		DR
Nrg1	YDR043C		-0.54		AC-W, SR
Pde2	YOR360C		**0.56**		DR, SGD
Pho4	YFR034C	0.30		**0.32**	*AC-MS*
Pho85	YPL031C		**-0.40**		SL, SR
Pho91	YNR013C			**0.41**	*BA*
Rad6	YGL058W	**0.20**			*SGD, PhE*
Rds2	YPL133C		**-0.55**	**-0.68**	*BA*
Rmd7	YER083C		**-0.23**		*SGD, PhE*
Rnr4	YGR180C			**-0.53**	PhE
Scp160	YJL080C			**0.52**	*BA*
Sec7	YDR170C		**1.44**		*AC-MS*
Sit4	YDL047W		**0.73**		SR
Skt5	YBL061C	**0.80**	**0.71**		*BA*
Smp2	YMR165C	**-0.43**			*BA*
Smt3	YDR510W		**0.46**		*AC-MS*
Snf4	YGL115W	**-**	**-**	**-**	AC-MS, RC, 2H, SGD, SR, DR, CoP
Ssf2	YDR312W	**0.47**			*BA*
Ssn3	YPL042C			**-0.45**	SR, AC-W, 2H
Tes1	YJR019C		**-1.16**		*BA*
Trm1	YDR120C		**0.89**	**0.83**	*BA*
Yat2	YER024W	**-1.33**			*BA*

[a] Gene expression change is expressed as a log_2 of the fold change, calculated for a mutant versus the wild-type strain. Fold change marked in bold represent by t-test identified significant ($P < 0.05$) expression change

[b] Listed Snf1 interactions are based on BioGRID database (BIOGRID-Saccharomyces_cerevisiae v.2.0.25). Types of interactions: 2H – two-hybrid, AC-MS – affinity capture mass spectrometry, AC-W – affinity capture western, BA – biochemical activity, CoP – co-purification, DL – dosage lethality, DR – dosage rescue, PhE – phenotypic enhancement, RC – reconstructed complex, SGD – synthetic growth defect, SL – synthetic lethality, SR – synthetic rescue. Interactions, which were identified using only global-scale studies, are marked in italic. References that support each interaction can be found at http://www.thebiogrid.org/SearchResults/summary/32529

SI Table 4. Snf1 kinase first neighbours identified using DOGMA subnetwork analysis

Protein[a]	ORF	Protein expression log_2(fold)[b]			Interaction (BioGRID)[c]
		Δsnf1	*Δsnf4*	*Δsnf1Δsnf4*	
Acc1	YNR016C	**-0.95**	**-0.79**	**-1.11**	SR
Arf1	YDL192W			**0.18**	*AC-MS*
Atg1	YGL180W	-1.00	**-0.72**		DR
Bfr1	YOR198C	**0.76**	**0.57**	**0.98**	*BA*
Cat2	YML042W	**-1.41**	**-1.36**	**-1.70**	SR
Cna1	YLR433C	**-1.24**			SGD
Fas1	YKL182W	**-0.89**	**-0.74**	**-1.63**	SR
Fox2	YKR009C	**-1.09**	**-1.13**	**-1.91**	*BA*
Gsy2	YLR258W	**-1.05**	**-0.91**	**-1.33**	PhS
Hog1	YLR113W	**0.46**			PhE
Hsf1	YGL073W	-1.11			2H, BA
Hsp82	YPL240C	**-0.36**	**-0.38**	-0.23	*SL*
Lhp1	YDL051W		**0.49**		*BA*
Msn5	YDR335W	**0.60**			DR
Pho85	YPL031C	**0.18**			SL SR
Rnr4	YGR180C			**-0.69**	PhE
Rpl8a	YHL033C	**-0.42**			*BA*
Rsp5	YER125W	-0.29		**-0.50**	*BA*
Scp160	YJL080C			-0.07	*BA*
Sec7	YDR170C			**-0.44**	*AC-MS*
Shs1	YDL225W		**0.76**		*BA*
Sit4	YDL047W	**0.22**	0.39		SR
Snf4	YGL115W	**-**	**-**	**-**	AC-MS, RC, 2H, SGD, SR, DR, CoP
Ssa1	YAL005C		**-0.51**		AC-MS
Stm1	YLR150W	**1.01**		**1.20**	*BA*
Tes1	YJR019C			**-1.18**	*BA*
Trm1	YDR120C	**0.26**			*BA*
Yat2	YER024W	**-2.86**	**-2.72**	**-1.36**	*BA*
Yck1	YHR135C	-0.26			DR

[a] Proteins marked in bold were also identified in high-scoring subnetwork analysis (SI Table 3)

[b] Protein expression change is expressed as a log_2 of the fold change, calculated for a mutant versus the wild-type strain. Fold change marked in bold represent by t-test identified significant ($P < 0.05$) expression change

[c] Listed Snf1 interactions are based on BioGRID database (BIOGRID-Saccharomyces_cerevisiae v.2.0.25). Types of interactions: 2H – two-hybrid, AC-MS – affinity capture mass spectrometry, BA – biochemical activity, CoP – co-purification, DR – dosage rescue, PhE – phenotypic enhancement, PhS – Phenotypic Suppression, RC – reconstructed complex, SGD – synthetic growth defect, SL – synthetic lethality, SR – synthetic rescue. Interactions, which were identified using only global-scale studies, are marked in italic. References that support each interaction can be found at http://www.thebiogrid.org/SearchResults/summary/32529

SI Table 5. Reporter Effectors identified for the *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains

Reporter Effector[a]	ORF	Gene expression log_2(fold)[b]		
		Δsnf1	*Δsnf4*	*Δsnf1Δsnf4*
Bas1[c]	YKR099W	-0.17		
Cat8	YMR280C	-0.07		-0.25
Ccr4	YAL021C			-0.30
Cse2	YNR010W			**-0.47**
Cyc8	YBR112C			0.07
Edc1	YGL222C	0.07		
Gcn4	YEL009C			0.05
Gcn5	YGR252W			0.40
Gcr1	YPL075W			0.20
Gcr2	YNL199C	-0.07	-0.16	
Gts1	YGL181W			-0.30
Hal9	YOL089C	-0.10		
Hog1	YLR113W		0.32	0.11
Hxk2	YGL253W	**0.47**		
Imp2'	YIL154C	**-1.10**		
Isw2	YOR304W	0.33		
Itc1	YGL133W	0.00		
Leu3	YLR451W		0.04	
Mbf1	YOR298C-A			-0.49
Mcm1	YMR043W		-0.44	
Mig2	YGL209W		**-1.78**	
Mig3	YER028C		**-1.33**	
Mks1	YNL076W			**0.13**
Msn2	YMR037C	0.13		-0.38
Msn4	YKL062W	**0.47**		**0.95**
Ndd1	YOR372C		0.16	
Oaf1	YAL051W		-0.27	
Pgd1	YGL025C			**-0.42**
Pho2	YDL106C	0.73		
Pho85	YPL031C		**-0.40**	
Pip2	YOR363C		0.02	
Rdr1	YOR380W	-0.40		
Rds2	YPL133C	**-0.47**		
Rox3	YBL093C			-0.55
Rsc1	YGR056W			-0.13
Sin4	YNL236W	-0.03	-0.04	0.16
Skn7	YHR206W	**0.53**		
Snf1	YDR477W	**-**		
Snf12	YNR023W			-0.15
Spt16	YGL207W		0.31	
Spt3	YDR392W			0.15
Spt6	YGR116W	0.40		
Spt7	YBR081C			0.42
Stb5	YHR178W		0.03	
Swi1	YPL016W	0.10		
Swi4	YER111C	0.73		
Swi6	YLR182W		0.66	
Tor1	YJR066W			-0.10
Tup1	YCR084C	0.17		
Yap1	YML007W	0.20		0.11
Yil130w	YIL130W		0.36	
Yrr1	YOR162C		-0.18	

[a] Reporter Effectors were identified using Reporter Effector analysis (see Supplementary Methods)

[b] Gene expression change is expressed as a log_2 of the fold change, calculated for a mutant versus the wild-type strain. Fold change marked in bold represent by t-test identified significant ($P < 0.05$) expression change

[c] Reporter Effectors marked in italic were identified in Harbison *et al* Chip-CHIP study (Harbison, *et al* (2004), *Nature*), the rest have confirmation in YPD database at www.proteome.com/databases/

SI Table 6. Reporter Metabolites identified based on gene expression data

***Δsnf1* versus wild type**			***Δsnf4* versus wild type**			***Δsnf1Δsnf4* versus wild type**		
Metabolite[a]	No. of genes[b]	*P* value[c]	Metabolite	No. of genes	*P* value	Metabolite	No. of genes	*P* value
UDP-GalNAc	4	1E-02	O-Acetylcarnitine m.	2	7E-03	L-Alanine	10	2E-03
O-Acetylcarnitine m.	2	2E-02	Carnitine m.	2	7E-03	L-Glutamate	41	2E-03
Carnitine m.	2	2E-02	Glutathione	10	9E-03	Anthranilate	4	5E-03
PEP	3	3E-02	Sulfate	5	1E-02	α-ketoglutarate	18	8E-03
L-Alanine	10	3E-02	Succinate	7	1E-02	L-Tyrosine	10	9E-03
SAICAR	2	3E-02	Isocitrate	5	1E-02	L-Histidine	7	9E-03
SER ext.	5	4E-02	2-Isopropylmalate	2	1E-02	L-Phenylalanine	6	1E-02
GlcNAc-1-P	2	4E-02	L-Cystathionine	3	2E-02	ALA ext.	5	1E-02
dADP	4	5E-02	Coproporphyrinogen	2	2E-02	Imidazole acetol-P	2	1E-02
Oxidized thioredoxin	3	5E-02	O-Acetylcarnitine	2	3E-02	1L-myo-Inositol 1-P	2	2E-02

[a] UDP-GalNAc stands for UDP-N-acetyl-D-galactosamine, PEP - 3-Phosphonooxypyruvate, SAICAR – 1-(5'-Phosphoribosyl)-5-amino-4-(N-succinocarboxamide)-imidazole, GlcNAc-1-P – N-Acetyl-D-glucosamine 1-phosphate; Imidazole acetol-P - 3-(Imidazol-4-yl)-2-oxopropyl phosphate; m = mitochondrial, ext = extracellular

[b] "No. of genes" indicates the number of genes with altered transcript levels which encode enzymes involved in the synthesis or degradation of the reporter metabolite

[c] Metabolites were ordered according to the *P* value scored in the genome-scale metabolic model

SI Table 7. Reporter Metabolites identified based on protein expression data

***Δsnf1* versus wild type**			***Δsnf4* versus wild type**			***Δsnf1Δsnf4* versus wild type**		
Metabolite[a]	No. of proteins[b]	*P* value[c]	Metabolite	No. of proteins	*P* value	Metabolite	No. of proteins	*P* value
Acetyl-CoA	17	2E-04	Acetyl-CoA	16	1E-03	Succinate	4	5E-04
CoA	24	2E-03	GLC ext.	10	2E-03	NADH	28	2E-03
NH_3 ext.	2	5E-03	FRU ext.	10	2E-03	NAD^+	29	2E-03
Acyl-CoA	4	9E-03	MAN ext.	10	2E-03	Carbamoyl P	3	5E-03
Glycogen	4	1E-02	D-Glucose 1-P	4	2E-03	Glycogen	3	6E-03
α,α-Trehalose	3	1E-02	D-Fructose	12	3E-03	Acetyl-CoA	14	1E-02
Succinate	5	2E-02	α-D-Mannose	13	3E-03	D-Glucose 1-P	4	1E-02
Malonyl-CoA	2	2E-02	Glycogen	4	3E-03	α,α-Trehalose	2	1E-02
3-Oxoacyl-CoA	2	2E-02	CoA	20	7E-03	Glyoxylate	5	1E-02
LCCA	3	2E-02	α-D-Glucose	18	9E-03	L-Aspartate	6	1E-02

[a] LCCA stands for a Long Chain Carboxylic Acid, ext. = extracellular

[b] "No. of proteins" indicates the number of proteins with altered abundance levels which are enzymes involved in the synthesis or degradation of the reporter metabolite

[c] Metabolites were ordered according to the *P* value scored in the genome-scale metabolic model

SI Table 8. Detected and relatively quantified intracellular metabolites

***Δsnf1* versus wild type**		***Δsnf4* versus wild type**		***Δsnf1Δsnf4* versus wild type**	
Metabolite	log_2(fold)[a]	Metabolite	log_2(fold)	Metabolite	log_2(fold)
citramalate	*-10*	glyoxylate	*-10*	2-isopropylmalate	*-10*
oleic acid	*10*	oleic acid	*10*	Citramalate	*-10*
Pyruvate	*10*	pyruvate	*10*	lauric acid	*-10*
stearic acid	*10*	stearic acid	*10*	oleic acid	*10*
asparagine	**0.88**	palmitic acid	**3.18**	Pyruvate	*10*
isoleucine	**1.01**	aspartate	**1.53**	stearic acid	*10*
Fumarate	**0.60**	asparagine	**0.88**	α-ketoglutarate	**2.51**
Glycine	**3.59**	P-enolpyruvate	**-0.85**	asparagine	**2.20**
glyoxylate	**-2.06**	itaconic acid	**1.09**	NADP/NADPH	**1.38**
myristic acid	**0.94**	glycine	3.55	succinate	**1.70**
phenylalanine	**0.33**	myristic acid	1.02	threonine	**3.30**
palmitic acid	**3.07**	palmitoleic acid	2.47	fumarate	**1.28**
P-enolpyruvate	**0.22**	citrate	1.81	isoleucine	**2.08**
Leucine	**0.99**	lysine	2.63	glycine	**5.25**
itaconic acid	**1.34**	malate	-0.46	malate	**1.09**
Aspartate	1.85	isoleucine	0.96	itaconic acid	**2.10**
Serine	1.33	serine	1.94	serine	**2.45**
Threonine	1.43	α-ketoglutarate	0.38	glutamine	**0.69**
palmitoleic acid	2.46	threonine	1.62	lysine	**3.90**
Lysine	2.38	leucine	0.80	glyoxylate	**-4.74**
Citrate	1.99	2-isopropylmalate	0.67	valine	**2.04**
Alanine	0.64	phenylalanine	1.29	glutamate	**0.99**
glutamine	-0.21	succinate	-0.26	tyrosine	**1.67**
2 aminobutyric a.	0.85	2 aminobutyric a.	0.86	citrate	**1.65**
succinate	0.31	alanine	0.79	alanine	**1.30**
a-ketoglutarate	0.93	lauric acid	0.35	P-enolpyruvate	**-0.38**
Valine	0.58	glutamate	0.53	proline	**1.15**
proline	0.47	tyrosine	0.54	aspartate	3.23
glutamate	0.53	valine	0.67	histidine	1.56
lauric acid	0.16	NADP/NADPH	0.42	palmitic acid	2.44
Malate	0.25	glutamine	0.52	2 aminobutyric a.	0.96
Tyrosine	0.23	proline	0.10	phenylalanine	1.39
Ornithine	0.33	ornithine	0.24	ornithine	0.97
2-isopropylmalate	0.12	histidine	0.12	palmitoleic acid	2.00
Histidine	-0.05	citramalate	-0.30	myristic acid	0.20
NADP/NADPH	-0.07	fumarate	0.07	leucine	-0.06

[a] Peak enumeration was conducted with AMDIS (NIST, Gaithersburg, MD) with default parameters, and identification of conserved metabolites was conducted with SpectConnect, spectconnect.mit.edu, (Styczynski, *et al* (2007), *Anal.Chem.*), using default parameters and a support threshold of 3. Metabolite abundance difference is expressed as a log_2 of the fold change, calculated for a mutant versus the wild-type strain. Fold change marked in bold represents by t-test identified significant ($P < 0.05$) metabolite abundance difference between two strains of comparison. Fold change marked in *italic* represents assigned metabolite abundance difference between two strains of comparison; in such case, the metabolites indicating mass spectrometry peaks were detected in one strain, but not the other strain of comparison.

SI Table 9. Expression changes of selected ORFs[a] for *Δsnf1*, *Δsnf4* and *Δsnf1Δsnf4* strains

GO category	Gene/Protein name		*Δsnf1*		*Δsnf4*		*Δsnf1Δsnf4*	
	ORF	Common	Protein expression	Gene log_2(fold)[b]	Protein expression	Gene log_2(fold)	Protein expression	Gene log_2(fold)
Biological process								
Glycogen biosynthesis	YEL011W	GLC3	-0.57	-0.30	**-0.67**	0.18	**-1.67**	**-0.98**
	YFR015C	GSY1	**-0.82**	0	**-0.62**	0.81	-	**-0.40**
	YLR258W	GSY2	**-1.05**	-0.40	**-0.91**	0.17	**-1.33**	**-0.86**
Glyoxylate cycle	YCR005C	CIT2	0.63	0.23	0.87	-0.17	**1.18**	**0.94**
	YER065C	ICL1	-1.85	**-2.33**	**-2.21**	**-2.86**	**-2.28**	**-1.10**
	YOL126C	MDH2	-1.82	**-1.26**	-1.93	**-1.26**	**-1.53**	**-1.77**
	YNL117W	MLS1	-3.06	**-3.12**	**-4.02**	**-3.17**	**-2.85**	**-1.52**
Gluconeogenesis	YHL032C	GUT1	**-2.17**	**-1.03**	**-1.58**	-0.34	-	**-1.24**
	YIL155C	GUT2	**-0.96**	**-0.33**	**-0.98**	-0.07	**-1.64**	**-0.50**
	YMR303C	ADH2	**-1.85**	**-0.43**	**-2.00**	-0.07	**-4.25**	**-1.42**
Amino acid metabolism	YOR184W	SER1	0.49	**0.30**	**0.26**	0.17	**1.10**	**0.63**
	YGR208W	SER2	0.57	**0.63**	0.63	0.48	1.44	**0.99**
	YER081W	SER3	**0.67**	**1.03**	**0.83**	**0.74**	**1.27**	**1.21**
	YIL074C	SER33	0.25	**0.43**	**0.31**	0.24	**0.57**	**0.50**
β-oxidation of fatty acids	YDR256C	CTA1	-0.98	**-1.13**	**-1.38**	**-0.99**	**-2.28**	**-1.88**
	YOR317W	FAA1	**-0.44**	-0.13	**-0.47**	0.03	**-0.92**	-0.13
	YER015W	FAA2	**-0.96**	-0.20	**-0.80**	-0.45	**-1.55**	**-0.48**
	YMR246W	FAA4	**-0.53**	**-0.20**	**-0.37**	-0.10	-0.40	**-0.34**
	YKR009C	FOX2	**-1.09**	-0.23	**-1.13**	**-0.47**	**-1.91**	-0.63
	YIL160C	POT1	**-1.05**	**-1.36**	**-1.24**	**-1.25**	**-2.05**	**-1.61**
	YGL205W	POX1	**-0.78**	-0.07	**-0.76**	-0.37	**-1.61**	-0.44
Fatty acid synthesis	YNR016C	ACC1	**-0.95**	-0.20	**-0.79**	-0.21	**-1.11**	-0.26
	YKL182W	FAS1	**-0.89**	**-0.17**	**-0.74**	**-0.15**	**-1.63**	**-0.50**
	YPL231W	FAS2	**-0.87**	-0.17	**-0.77**	-0.22	**-1.61**	**-0.43**
	YML075C	HMG1	-0.32	**-0.33**	-0.33	**-0.29**	**-0.64**	**-0.28**
Carnitine transfer and metabolic process	YML042W	CAT2	**-1.41**	**-0.83**	**-1.36**	**-0.89**	**-1.70**	**-0.58**
	YAR035W	YAT1	**-3.79**	**-1.89**	**-2.06**	**-2.47**	**-1.60**	-0.38
	YER024W	YAT2	**-2.86**	**-1.33**	**-2.72**	**-1.37**	**-1.36**	0.29
	YOR100C	CRC1	-	**-1.10**	-	**-0.64**	-	**1.34**
Molecular function								
Oxidoreduction activity	YGR088W	CTT1	**-**	**-1.69**	**-**	**-1.49**		**-3.16**
	YBR244W	GPX2	**-**	**-0.63**	**-**	**-0.54**		-0.23
	YJR104C	SOD1	-0.61	**-0.30**	-0.49	**-0.14**	0.06	**-0.17**
	YHR008C	SOD2	-0.49	**-0.60**	**-0.78**	-0.17	-0.40	**-0.76**

[a] 'selected ORFs' are those, which are specifically discussed in the study or used in Figure 2 and SI Figure 5.

[b] Protein and gene expression difference is expressed as a log_2 of the fold change, calculated for a mutant versus the wild-type strain. Fold change marked in bold represent by t-test identified significant ($P < 0.05$) metabolite abundance difference between two strains of comparison.