This is to certify that the
dissertation entitled

RELATING STRESS TO STRAIN AT THE LEVEL OF GENE EXPRESSION

presented by

Stephen Jan Callister

has been accepted towards fulfillment
of the requirements for the

Doctoral degree in Environmental Engineering

Major Professor's Signature

1/28/2004

Date

RELATING STRESS TO STRAIN AT THE LEVEL OF GENE EXPRESSION

By

Stephen Jan Callister

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Civil and Environmental Engineering

2004

ABSTRACT

RELATING STRESS TO STRAIN AT THE LEVEL OF GENE EXPRESSION

By

Stephen Callister

The stress strain paradigm was investigated at the level of gene expression for *Saccharomyces cerevisiae* using tools developed and adapted from various fields including Ecology, Mathematics, and Engineering. Tools presented in this dissertation were developed to calculate strain, i.e., an aggregate measure of gene expression response, and to establish its relationship to stress, i.e. the perturbation causing the response. Strain for the Environmental Stress Response in *S. cerevisiae* was calculated using a Moment of Area and compared across eight environmental perturbations. Additionally, relative expression for a set of 169 genes making up the glycolysis, glycerolipid metabolism, cell cycle, and MAPK signaling biochemical pathways was observed for a series of applied osmotic stresses representing mild, hyper, and sever osmotic shock. To add biological significance to the stress strain paradigm, stability and its associated aspects such as *resistance*, which is the ability of a system to withstand the perturbation, *resilience*, which is rate at which a system returns to steady state, and *reactivity*, which is the rate at which a system departs from steady state were applied at the level of gene expression adding biological significance to the observed strain. Results showed that with increased perturbation magnitude a decrease in resilience, reactivity, and resistance

occurred. An overall decrease in stability resulted with increased strain as measured by the Moment of Area. An exponential relationship between perturbation magnitude and strain was observed on an individual gene basis, biochemical pathway basis, and aggregate gene basis. This dissertation presents the first application of relating stress to strain on a gene level.

DEDICATION

To my Father and Mother for instilling in me the desire to comprehend and achieve beyond my perceived expectations. Also, to Charlene for her patience and encouragement.

ACKNOWLEDGMENTS

The body of research contained in this dissertation was not possible without the help and guidance of my advisor Dr. Syed A. Hashsham. It has been an honor to work with him and I have thoroughly enjoyed observing his scientific and engineering creativity. His commitment and expressed excitement to this research has more than once carried me through periods of disappointment, and his willingness to take the time and give valuable criticism helped keep this research focused on the objectives.

I also need to express gratitude to my research committee, Dr. Bruce E. Dale for his flexibility and willingness to not "stand in the way" during the last few months of this research, Dr. Susan J. Masten for her perspective and realistic advise, and Dr. James M. Tiedje for his patience in helping an engineer learn and develop the molecular biology tools necessary for this research.

Mr. Sean M. Spellman gave hours of help with experimental set up and sampling. RNA extractions were performed with the help of Mr. John J. Parnell. Instrumentation set up and optimization was not possible without Dr. Annette P. Thelen of the Genomics Teaching and Support Facility. To these individuals I express my sincere gratitude for rolling up your sleeves and jumping in to keep things moving efficiently.

I need to express my appreciation to my parents, Jan and Jane Callister. My father is credited for introducing me to and instilling in me the desire to try and comprehend, at least to a small degree, my surroundings. Whether it was

observing the stratigraphy of the Colorado Plateau, scanning the horizon on top of Notch Peak, or fossil hunting in the Great Basin Desert he taught me as a boy to never be afraid to ponder the answers associated with the difficult interrogatives "How?" and "Why?". My mother is credited for teaching me that patience in oneself, hard work, trial and error, and faith are the required ingredients for accomplishing a difficult task. During the pursuit of this Ph.D. she continually reminded me of the long term benefits associated with what I was undertaking, which can often be difficult to discern after the n^{th} experimental failure.

Finally, I owe the deepest gratitude to my wife, Charlene. She is the only one, among all that have been mentioned, who has directly taken on the burden of supporting me in this pursuit while delaying her own educational goals, maintaining our home, and raising our two daughters. Thank you Charlene!

TABLE OF CONTENTS

LIST OF FIGURES

Some of the figures in this dissertation are in color.

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background and significance

High throughput genomic technologies have dramatically increased our ability to develop models to quantitatively predict the behavior of microorganisms in natural and engineered systems (Schena et al.,1995; Lockhart, et al., 1996; DeRisi et al., 1997). Many variations of the technology exist today, including the glass slide based cDNA and oligonucleotides microarrays (DeRisi et al., 1997; Agilent, Palo Alto, CA), in situ synthesized oligonucleotide arrays (Affymetrix, Santa Clara, CA; Xeotron, Houston, TX; NimbleGen, Madison, WI), and bead-based systems (Spiro et al., 2000). Typically, whole genome expression studies using the above technologies revolve around measuring relative changes in messenger RNA of thousands of genes together (referred to as genome expression pattern, response, or strain) to physical, chemical, or biological perturbations (referred to as perturbation or stress). Knowledge of this response is believed to be important in making new discoveries in many fields including oncology- to develop drugs for the control of cancerous cells (Perou et al., 2000; Swami et al., 2003), biotechnology- to produce proteins of commercial value, and environmental engineering- to study the degradation of xenobiotic compounds during bioremediation (Beliaev et al., 2002). In parallel to these, other low-throughput but more quantitative technologies such as quantitative real time polymerase chain reaction (RT-PCR) have also progressed towards higher capacities (Hernandez et al., 2000; DeFrancesco, 2003). Quantifying the relative

abundance of mRNA from thousands of genes is more economically feasible using the microarray platform. Quantifying the absolute abundance of a few hundred mRNAs is better accomplished by RT-PCR.

Since the inception of microarray technology, approximately a decade ago, more than one thousand studies have been reported covering dozens of whole genomes of various microorganisms (Wodicka et al., 1997; Cho et al., 1998; Hecker and Engelmann, 2000; Hinchliffe et al., 2003). Many of the data analysis tools applied to these studies were initially developed using experiments performed on the *Sacchaormyces cerevisiea* (yeast) genome because of experimental data publicly available from the laboratories of Pat Brown and Jospeh DeRisi (DeRisi et al., 1997; Gollub et al., 2003). The sharing of gene expression data and results, although initially lagging, has taken on new meaning in the fields of bioinformatics, genomics, and microbiology. A number of mathematical tools now exist to cluster and analyze the deluge of genome expression data under varying environmental conditions and extract qualitative (and often quantitative) information (Eisen et al., 1998; Xia and Xie, 2001; Autio et al., 2003; Laws et al., 2003).

A primary goal of data analysis for experiments involving environmental perturbation and genomic response is to establish (to the extent possible) predictive rules for the behavior of a microorganism for the system of interest (reactor, human body, soil, aquifer, surface water etc.). The predictive rules of behavior can be established on many organizational levels including community, population, proteome, and transcriptome. On a transcriptome level, none of the

analysis tools at present have the capability to find *a quantitative aggregate measure of relative gene expression for a group of genes; whether it be a cluster of genes, a pre-defined gene set, or whole transcriptome*. The aggregate measure of gene response is referred too as *strain* throughout this dissertation. A measure of strain is needed to compare the effect of different types and magnitudes of perturbations on the transcriptome. This information in turn may be useful in the direct control and management of gene expression and indirectly in the control and management of the microorganism. Past studies related to the existence of quantitative relationships, if any, between strain and applied perturbation (referred to as *stress* subsequently) have not been undertaken, because tools to calculate an aggregate measure of expression did not exist. This posed some problems for the research presented in this dissertation, because no published information on the aspect of relating stress to strain in a quantitative manner on a transcriptome level was found. Hence, many comparisons were made to studies that closely resembled the objectives and hypotheses of this study. It is of course possible to draw parallel from dose-response studies and toxicity evaluations at the organism level, but none of the studies related to analysis of gene expression patterns have adopted the approach presented here.

It is well known that many physical, chemical, social, and economic systems behave according to a stress-strain paradigm. Young's modulus of elasticity (extension in the length of a wire per unit area per unit of applied load) used to measure the strength of a material is an example of this type of

relationship important to comparing materials used in civil engineering. Ecologists routinely study ecosystems as a perturbation-response problem and characterize it as less or more stable. Economists forecast (to the extent possible) the change in economic activity due to an event such as a decrease in the interest rate. The hole in ozone layer is an example of the strain caused by intentional and unintentional release of chlorofluoro-carbons (CFCs). The foundation for this research is that "within limits", such a stress-strain relationship also exists for genomic systems and can be quantitatively studied by employing aggregate measures of "stress and strain". Such quantitative relationships will have potential utility in control and management of microorganisms in all areas of science.

1.2 Stress, Strain, and Stability

An aggregate measure of strain can be obtained from application of stability principles from ecology. Ecologists analyze and measure the response of a system by its response envelope (Nuebert and Caswell, 1997). A response envelope is simply the curve describing the behavior of a state variable in response to an applied stress. It is generally parabolic in shape but other responses are possible. Gene expression patterns can also be considered as response envelops. In ecology, the state variable is often the biomass or number of species, or the abundance of a nutrient. In gene expression studies, the response envelope will be attributable to the relative abundance of mRNAs associated with individual genes. Indeed, in genomics, not all envelopes are of the traditional parabolic shape. Five other common shapes of response are:

inverse of a parabola, a step-up response, a step-down response, and no response at all to the stress, implying no detectable deviation from pre-perturbed expression following the onset of a stress. It is obvious that genomic responses are more complex and may require additional tools for analysis compared to the response envelopes that are generally considered in ecology. However, as presented later in this dissertation, the mechanistic mathematical equations developed to analyze the response envelope can be applied to genome expression patterns of many types.

"Stability" is a system-associated property extracted from the response envelope. Its origins are rooted in the physics of perturbed motion (Merkin, 1997). In ecology, this property has been used in different manners (Grimm et al., 1992), requiring a specific definition for this research. *Stability is the ability of a system to remain at equilibrium following a perturbation* (Holling, 1973; Harrison, 1979; Sennhauser, 1991). Although equilibrium is commonly equated to stability in an ecological sense the use of steady state is more appropriate in the case of relative gene expression. Steady state is defined in this research as:

$$\text{Steady State} \equiv \frac{d\Delta x(t)}{dt} = 0 . \qquad (1.1)$$

Where, $\Delta x(t)$ represents the difference between the perturbed, $x_p(t)$, and pre-perturbed expression, $x_e(t)$, of an individual gene, or system of genes. In this case, a system is defined as a set of related genes found within a gene cluster or predefined group, such as those making up the glycolyis pathway for

Saccharomyces cerevisiae. Certain aspects of stability, or stability parameters, are commonly used to quantify this property. The two most common parameters are resistance and resilience. *Resistance* is a measure of the magnitude of response, i.e., the maximum change in the parameter of interest in response to perturbation. *Resilience* is often described either as the time required for the system to return to its pre-perturbation state or rate at which the system returns to its pre-perturbed state. The exact definition used in past studies has often depended on the author's prerogative. Resilience applied to gene response envelopes is described in more detail in Chapter 2. Since resistance and resilience are both quantitative terms (differing in units, however), they can be combined appropriately to yield a single measure for a single response envelope. And, for many response envelopes, strain calculated from the aspects of stability can be added together to describe the system as a whole (Grimm et al., 1992). Therefore, a value of strain made up of thousands of gene response envelopes is theoretically feasible. Such a summation of stability parameters calculated as Moments of Area for six response envelopes was used to describe the functional stability of an anaerobic microbial community (Hashsham et al., 2000). The research presented, proposes to use the same concept of calculating the Moment of Area of response envelopes extended to genome expression patterns.

It is well known that only a certain percentage of genes respond to an applied stress. These are generally those genes that are related to a group of proteins useful in averting the effect of the applied stress. It may belong to a

single pathway or spread over a few pathways. It is apparent that an aggregate measure of expression response (in terms of strain) is possible, but comes at the cost of losing resolution, i.e., information about the type of genes or pathways responding to that stress is lost. Such loss of resolution is common in all calculations that result in an index. The benefit of calculating an index becomes obvious when we consider that two very different types of perturbation (e.g., temperature and hydrogen peroxide) can be compared to each other via this index. This would not have been possible by cluster analysis or any other mathematical analysis that is currently in use in data analysis. It is possible, however, to calculate the strain for one gene, one pathway, a group of genes responding to the stress, or the whole genome. By keeping the gene identities included in an analysis constant, the comparison of strains may be made more biologically meaningful.

1.3 Choice of Monitoring Technology for Strain Quantification

After some initial analysis, it became clear that the cost of testing the applied concepts was prohibitive on microarrays. Experiments on genome expression patterns yield relative expression ratios. These ratios must be obtained preferably in triplicate for good data analysis. Adequate description of time series data often requires a minimum of five to seven temporal points, judiciously selected at times that represent important changes in the expression ratio. In addition, the experiments needed to be conducted at several magnitudes of stress to be linked to the magnitude of strain. In terms of microarrays,

considering six different magnitudes of stress including the control, and applying a factor of safety of 40% extra because some microarrays may not yield good results, at least one hundred microarrays will be needed. Combining the cost of labeling and hybridization buffer, the total cost was estimated at more than $50,000 in supplies. Therefore, an alternative approach was adopted. This approach used quantitative real-time PCR and reverse transcription of the mRNA. Quantitative real-time PCR is often used to validate expression ratios obtained from microarray experiments because it provides better quantitative information and can be conducted to yield absolute amounts of the mRNA. However, RT-PCR becomes uneconomical and more time consuming for more than a few hundred genes. The strategy of making quantitative RT-PCR economical stems from limiting genes to a smaller set that has been shown to be relevant to the type of stress. Yet, the cost saving strategy for a smaller number of genes can also be achieved for the microarray platform, especially if the microarray is printed in house. Therefore, the determining factor between a microarray of a smaller set of genes and RT-PCR was the reliability in determining the magnitude of strain. RT-PCR quantification of mRNA is more reliable with the possibility of absolute quantification. Hence, RT-PCT of a set of genes responding to the model stress was chosen as the best option for this research.

1.4 Choice of the Model Microorganism and Stress

The type of microorganisms or stress that can be used for validation of the stress-strain relationship is abundant. Therefore, organism selection for the purpose of this study was carried out based on the following two characteristics:

i) The microorganism's whole genome expression pattern should be well studied with respect to the type of stress and response so that an informed selection can be made about the set of genes that are relevant to the stress. This was essential because of the use of RT-PCR based quantification instead of the microarray-based quantification.

ii) Initial data on whole genome expression patterns must be available in the public domain so that mathematical tool development for quantifying the strain can be accomplished before conducting a more extensive stress-strain experiment.

S. cerevisiae satisfies the above two requirements. It is one of a few organisms in which the effects of and response to osmotic shock have been extensively studied (Hohmann and Mager, 1997). Briefly, upon encountering osmotic shock, growth ceases and the cell's volume decreases until the internal production of glycerol reduces the water potential sufficiently for growth to resume. Glycerol is the only compatible osmolyte produced by *S. cerevisiae* and its production in response osmotic shock results from transcription of GPD1 and a host of supporting genes (Hohmann and Mager, 1997; Martijn et al., 1999, 2000 & 2001; Mager and Siderius, 2002).

Increased levels of mRNA for GPD1 have been observed with greater osmotic shock, suggesting a relationship between the magnitude of the perturbation and response. The potential dependence between perturbation magnitude and response of this gene suggests that other genes related to glycerol production, or to maintaining cellular viability, should also exhibit greater response with greater osmotic shock. Currently, no quantitative description of this relationship exists for relative gene expression of *S. cerevisiae* in response to osmotic shock.

S. cerevisiae has been extensively used as a model organism to study the whole genome response to a number of stresses. Gasch and co-workers collected data on the relative gene expression of *S. cerevisiae* to a variety of environmental perturbations (Gasch et al., 2000). Eight of twelve perturbations, which are summarized in Table 1.1, were selected for addressing hypothesis 1 described below. Relative mRNA abundance was monitored by Gasch et al., (2000), over time using a full-scale genome microarray developed by Pat Brown and colleagues at Stanford University (Shalon et al., 1996; DeRisi et al., 1997). This microarray contained the PCR amplification products of approximately 6400 distinct genes along with positive controls for normalization and negative controls to monitor hybridization quality. The data from these experiments is publicly available from the Stanford Microarray Database (Gollub et al., 2003). Hierarchical clustering of differentially expressed genes (relative expression greater than 2-fold change) revealed two clusters of a combined total of approximately 900 genes. This set of genes was labeled by Gasch et al.,(2000),

Table 1.1. Description of the eight perturbations administered by Gasch et al.,(2000) in order to study relative gene expression in *Saccharomyces. cerevisiae*.

Environmental Perturbation (Cellular Stress)	*Description*
1. Temperature shock from 25° to 37° C	Cells were grown at 25°C collected by centrifugation, and suspended in media at 37°C. This temperature was maintained throughout the experiment.
2. Temperature shock from 37° to 25°	Cells were grown at 37°C collected by centrifugation, and suspended in media at 25°C. This temperature was maintained throughout the experiment.
3. Hyper-osmotic shock	Cells were grown to an OD_{600} of 0.6 and supplemented with media at containing 2 M sorbitol. Final concentration of sorbitol was 1.0 M.
4. Hypo-osmotic shock	Cells were grown in the presence of 1.0 M sorbitol, collected by centrifugation and suspended in media without sorbitol.
5. Hydrogen peroxide shock	Cells were grown to early log phase and H_2O_2 added for a final concentration of $3.0x10^{-4}$ M. Culture volume and concentration were maintained throughout the experiment.
6. Menadione shock	Menadione bisulfate previously suspended in water at a concentration of 1.0 M was added to the cell culture for a final concentration of 1 mM.
7. Diamide shock	Diamide was added to the cell culture to a final concentration of $1.5x10^{-3}$ M.
8. DTT shock	Cells were grown at 25°C and dithiothreitol was added for a final concentration of $2.5x10^{-3}$ M.

as the Environmental Stress Response (ESR). One cluster of approximately 600 genes showed repression in transcription followed by a return to relative expression prior to the perturbation. The other cluster of approximately 300 genes exhibited an induction in transcription also followed by a return to relative expression prior to onset of the perturbation. The ESR was induced as a transitional response to maintain cellular conditions optimal for growth and survival. Many of the genes responding to these perturbations were also confirmed by Causton et al., (2001), and labeled as the Common Environmental Response.

The above study by Gasch et al., (2000), and Causton et al., (2001), would serve as an excellent starting point to explore if genome expression patterns can be summarized in terms of strain. However, it would be important that the numerical values of strain retain meaningful differences for comparison of stress and developing the stress-strain relationships. It became clear that data from various types of stresses obtained by Gasch et al., (2000), would be useful for this purpose. However, it was also clear that the previously performed experiments were not specifically designed for the purpose of testing the existence of a stress-strain relationship. Therefore, a series of controlled experiments would need to be performed to fully explore the validity of the hypothesized relationships. The following sections describe the objectives and hypotheses for this study.

1.5 Objectives

Two objectives of this study are put forth. *Objective 1)*: To develop a mathematical tool to calculate strain, i.e., an aggregate measure of gene expression response, from the response envelopes of a set of genes responding to the applied stress, and *Objective 2)*: To study the relationship between stress and strain at the level of gene expression for a set of selected genes in *S. cerevisiae* under a series of applied osmotic stresses.

Completion of the first objective is necessary before the second objective can be met. The first objective uses existing data from Joe DeRisi's laboratory (Gash, et al., 2000) for initial tool development. This tool is then applied to the

experimental data obtained as part of this research to accomplish the second objective.

1.6 Hypotheses

The following three hypotheses will be tested as part of this study. They are essentially sub-hypotheses emanating from the more fundamental question of whether stress is related to strain in biological systems.

Hypothesis 1 *– The magnitude of strain for the set of genes labeled as ESR in S. cerevisiae, can be calculated using the Moment of Area method and will be a function of the type of stress. It will also be relatively insensitive to various methods employed to calculate it.*

Since all response envelopes have areas, calculation of strain is trivial. Whether this calculated area retains the resolution with respect to various types and magnitudes of stresses is key to determining the usefulness of this approach. Proof of this first hypothesis determines the fate of the second hypothesis that relates to the magnitude of stress. Also, the mathematical tools developed to test the first hypothesis are essential for the other two.

Hypothesis 2 *–The magnitude of strain is related to the magnitude of stress for the set of genes responding to that stress. This relationship is most likely non-linear and exists within a range given by minimum and maximum values of the stress.*

This range exists because strain may not be detectable below a certain lower value of stress due to limitations posed by the detection limit of the monitoring tools. Similarly, above a certain value of the stress an organism may not be able to respond in any manner except inactivity or death.

Hypothesis 3 *– Moment of Area is a better measure of the strain than resistance or resilience alone because it combines and emphasizes delayed response of the microorganisms under higher magnitudes of stress and it gives equal weight to induction and repression.*

This last hypothesis emanates from the need to reconcile the numerous parameters that have been used to study stability of a system in ecology. For example, resistance, resilience, reactivity, declivity, time of return, time of maximum response, and many other parameters represent various aspects of the response envelope (Grimm et al., 1992). The details of these parameters, the subtle differences, and their utility in obtaining strain are presented in Chapter 2. Testing this last hypothesis is also essential because of the need to analyze response envelopes that are generated by induction as well as repression. For example, Gasch et al., (2000), observed two clusters of genes within the ESR, although in opposite directions, one representing induction and the other repression. Both returned to the expression levels prior to the perturbation. The third hypothesis tests if both clusters can be combined to make a single cluster using the moment of the area approach. Assuming that the stability of the ESR is

additive, then the Moment of Area calculated from the set of both clusters should quantitatively agree with the Moment of Area from the single cluster.

Five tasks were completed in order to test the above hypotheses. They are:

1. Review the current technologies used to group related genes according to patterns of relative gene expression in response to environmental perturbation. This was essential to device methods to calculate Moment of Area and stability parameters.
2. Evaluate various parameters of stability theory in ecology, and adapt the most suitable option to obtain strain from response envelopes of relative gene expression.
3. Develop mathematical tools to automatically calculate the stability parameters from the response envelopes of relative gene expression.
4. Conduct stress-strain laboratory experiments using *S. cerevisiae* and at least five different magnitudes of osmotic stress (selected because maximum information is available for this type of stress on yeast). The experiment involved growing *S. cerevisiae*, applying the stress, and collecting mRNA for quantification by reverse transcription and real-time PCR.
5. Use the mathematical tool developed in Task 4 to calculate strain and relate it to the corresponding stress.

Completion of the above tasks was carried out in three phases. In the first phase, a review of the current qualitative analysis techniques used to analyze relative genome expression was conducted. This phase was important to develop an understanding of how expressed genes are grouped in a systematic manner. The second phase evaluated various methods to calculate stability parameters and applied the best option to develop mathematical tools for calculating strain from response envelopes of whole genome expression patterns. Existing data from Gasch et al., (2000), on yeast was used for this second phase. The third phase involved experimentation and application of the mathematical tools developed in phase 2. *S. cerevisiae* was subjected to varying magnitudes of stress (osmotic shock), messenger RNAs were harvested and quantified using quantitative PCR. The response envelopes obtained by this experiment were analyzed to calculate strain and study its relationship with the applied stress.

The remaining chapters of this dissertation are arranged in the following manner. Chapter 2 describes the pertinent literature available on stress, strain, and stability. Chapter 3 presents the materials and methods employed in this study including the mathematical tools. Chapters 4 and 5 present the results obtained and the associated discussion. Finally, conclusions are presented in Chapter 6. Appendices are provided to document the MathCad program used for conducting calculations, list of genes used in the analysis with corresponding pathways.

CHAPTER 2

LITERATURE REVIEW

2.1 Pulse and Press Perturbations

A perturbation is responsible for causing the transcriptional response necessary to counteract the resulting cellular stress. Two distinct types of experimental perturbations described by Bender et al., (1984), are: "pulse" and "press". Figure 2.1 graphically depicts both pulse and press perturbations. Pulse perturbations are instantaneous changes in an environmental condition followed by a return to steady state. Press perturbations are changes in an environmental condition that are continually maintained over the period of the study or for a specified time. Responses to pulse perturbations are commonly of short duration compared to the duration of press perturbations, which invoke long-term responses to the permanent change (Inchausti, 1995). Currently, perturbation experiments focusing on gene expression have focused on the press type (Siderius et al., 1997; Gasch et al., 2000; Causton et al., 2001; Alexandre et al., 2001; Zhang et al., 2002).

Quantification of the magnitude and duration of a perturbation is difficult for those occurring naturally, and often only the effects (direct and indirect) are quantified rather than the perturbation (Bender et al., 1984; Yodzis, 1988). Experimental perturbations resulting from artificial manipulation of the environment or community are quantifiable. Quantification of the pulse perturbation is a function of both the magnitude and duration. Because the

duration of press perturbation may continue to infinity (or the period of study), quantification is a function of magnitude only.

The administration and quantification of press perturbations is simpler, but inherent problems can result in selecting the time scale for observing the response (Yodzis, 1988). Conclusions based on press perturbations must be based within the boundaries of the observed response. On the other hand, pulse perturbations can be more difficult to administer and quantify, but the time scale for observations is easier to determine. Observations in the case of pulse perturbations are made until the environmental condition or community has returned to its pre-perturbed state.

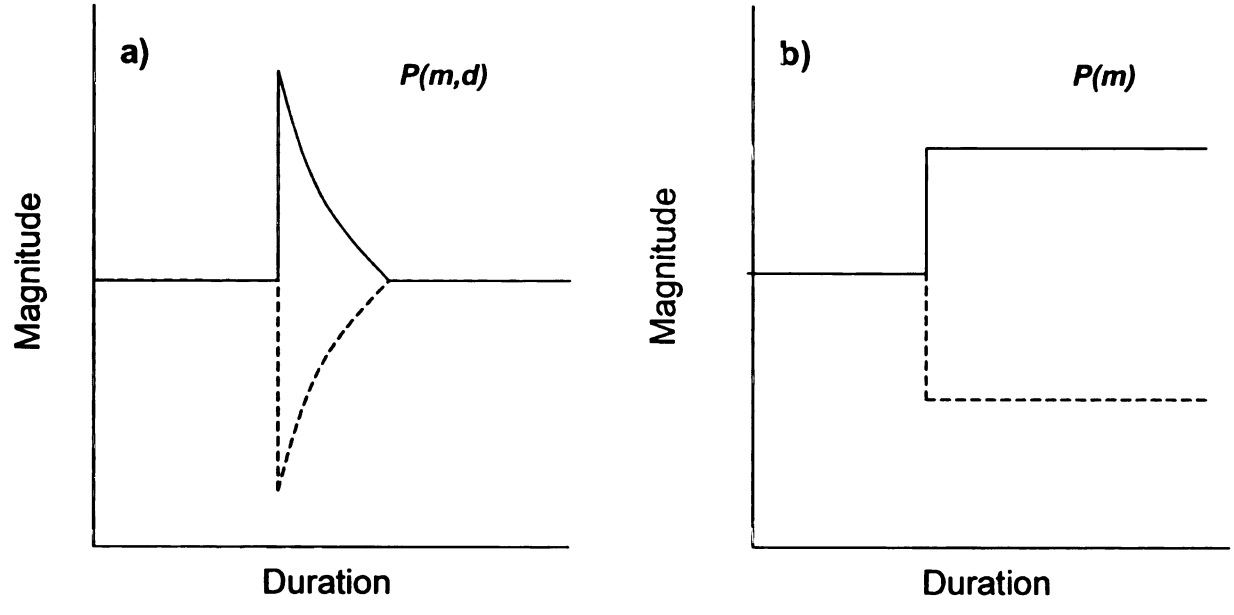

Figure 2.1. (**a**) Pulse perturbations in which an instantaneous change either by addition or removal of an environmental condition occurs followed by a return to the pre-perturbed condition. Pulse perturbations are a function of magnitude and duration. (**b**) Press perturbations in which change in an environmental condition is maintained. Press perturbations are a function of magnitude only.

2.2 Stability and Generalized Gene Expression Response Envelopes

Stability has generally been defined as the ability of a system to remain at equilibrium following a perturbation (Holling, 1973; Harrison, 1979; Sennhauser, 1991, Grimm et al., 1992). It is worth noting that in ecological literature, the concept of stability is often elusive, confusing, and subjective. Hence many other definitions can also be found. For the purpose of this research, the above definition will be used with the exception that "equilibrium" will be substituted with "steady state". A system that exhibits no displacement from steady state in terms of an observed parameter following a perturbation is deemed perfectly stable. However, once a system is displaced from steady state, it can either return to its pre-perturbed condition, or establish a new steady state. The former is known as asymptotic stability, and was first described by the Russian mathematician, Liapunov in the early 20^{th} century (referenced in Harrison, 1979; Merkin, 1997). The establishment of a new steady state is known as neighborhood stability and results from the perturbation permanently altering the environment (Lewontin, 1969; Sutherland, 1981; Murray, 1993; Hernandez, 2003). These two types of stability have been extensively discussed and debated by theoretical and empirical ecologists over several decades with demonstration to a few well-characterized ecosystems (Pim, 1982).

The above definitions associated with stability are applicable in general to any response envelope. Hence, they may be applied to the response envelope of gene expression pattern(s) for a single gene or a group of related genes that

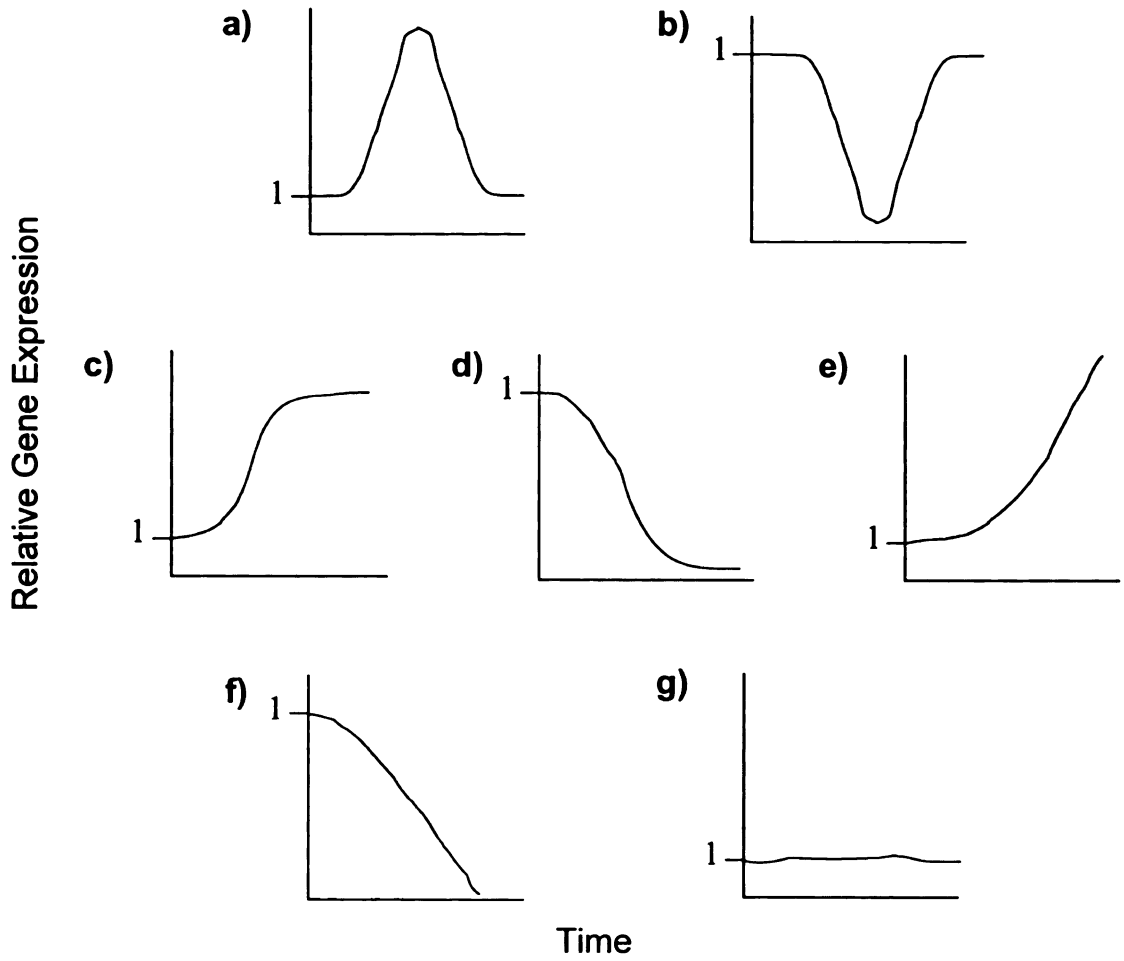

Figure 2.2. Generalized relative expression patterns in response to environmental perturbation. An induction **(a)** or repression **(b)** in transcription followed by a return to pre-perturbed expression levels. An induction **(c)** or repression **(d)** in transcription followed by an establishment of new relative expression levels. Patterns **(e)** and **(f)** could represent one of the other expression patterns given a longer time-scale. Pattern **(g)** represents no change in relative expression in response to perturbation.

result from an environmental perturbation. Figure 2.2 summarizes the generalized patterns of relative gene expression. Response **a** represents a transcriptional activation until suitable physiological conditions of the cell are restored followed by transcriptional repression at which point remaining mRNAs are translated or degraded to pre-perturbed expression levels. Contrasting response **a**, is response **b** in which transcriptional repression is followed by activation to achieve pre-perturbed expression levels. Responses **a** and **b** might occur when the perturbation momentarily changes environmental conditions after which the environmental conditions returns to its pre-perturbation state. Responses **c** and **d** denote induction and repression, respectively, until a new equilibrium expression level is achieved to maintain suitable physiological conditions. These responses might occur when the perturbation results in the establishment of new environmental conditions. It is unclear whether responses **e** and **f** will ultimately look like responses **a** and **c** or **b** and **d**, respectively, due to a shortened time scale of observation. Because all systems tend to a recognizable equilibrium following small perturbations (Pimm, 1982), the continuing trend of responses **e** and **f** is unlikely. Finally, response **g** represents no change in expression from the resulting perturbation.

Evidence of these generalized expression patterns was demonstrated by Wen et al. (1998) while studying the temporal gene expression of the central nervous system. Euclidean clustering identified five basic "waves" of gene expression, comparable to responses **a**, **c**, **e**, **f** and **g** that characterized central nervous system development. Evidence for responses **a** and **b** were presented

by Gasch et al. (2000) following the study of gene expression in *S. cerevisiae* to a number of perturbations including temperature and osmotic shock.

2.3 Measures of Stability used in Ecology

Figure 2.3a shows asymptotic stability for a cluster of genes responding to osmotic shock and neighborhood stability for a cluster of genes responding to a shift in temperature from 37° to 25° C (Gasch et al., 2000). These two sets of response envelopes are similar to the generalized response envelopes presented in Figure 2.1 **a** and **c**. The stability parameters resilience, and resistance (Figure 2.3b) can be used to describe specific aspects of asymptotic and neighborhood stability. An additional stability parameter, reactivity, is also shown in Figure 2.3b. Until now, these aspects of stability have predominately been applied to analyze response envelopes of parameters at the organism level (Viragh, 1989; Mittelbach et al., 1995; Grandpre and Bergeron, 1997). Application at the genome expression level will be attempted in this research for the first time.

Resilience and reactivity have units of inverse time and describe the relative change in abundance of mRNAs governed by the rates of transcription, translation, and translocation in Eukaryotes. The more resilient a cluster of genes, the more rapidly it returns to pre-perturbation state following the perturbation. Further, those genes that did not respond at all, as represented by the generalized response envelope **i** in Figure 2.2, are considered perfectly resilient. Mathematically, resilience is defined as (Neubert and Caswell, 1997):

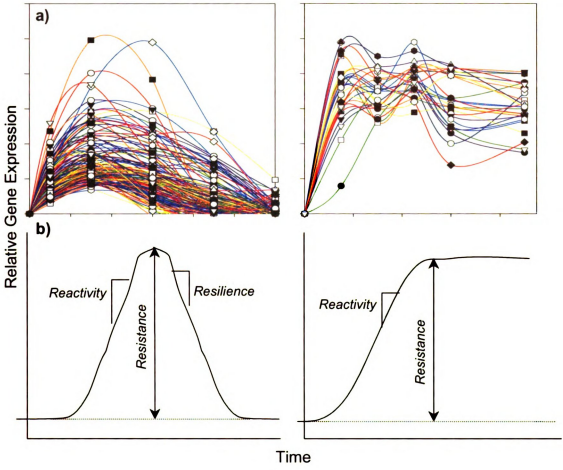

Figure 2.3. (a) Relative expression of *Saccharomyces cerevisiae* exhibiting asymptotic and neighborhood stability to hyper-osmotic and temperature perturbations **(b)** Application of the stability parameters to asymptotic and neighborhood stability.

$$\text{Resilience (1/time)} \equiv \lim_{t \to \infty} \left(\frac{1}{x(t)} \frac{d\Delta x(t)}{dt} \right). \qquad (2.1)$$

Where, $\Delta x(t)$ represents the difference between the perturbed, $x_p(t)$, and pre-perturbed expression, $x_e(t)$, state of the system. According to Equation 2.1, as t approaches ∞, the greater the negative value of $d\Delta x(t)/dt$, the more rapidly $x_p(t)$ approaches $x_e(t)$ and the greater the stability of the system in terms of resilience (Harrison, 1979, DeAngelis, 1980; Nuebert and Caswell, 1997).

Reactivity is a less commonly described stability parameter, but is complementary to resilience. Mathematically, reactivity is defined as (Nuebert and Caswell, 1997):

$$\text{Reactivity(1/time)} \equiv \lim_{t_0 \to t\max} \left(\frac{1}{x(t)} \frac{d\Delta x(t)}{x(t)} \right). \qquad (2.2)$$

Where, $t_0 \rightarrow tmax$ is the limit taken from the time at which the perturbation begins, t_0, to the maximum displacement from steady state, $tmax$. The larger the positive value of $d\Delta x(t)/dt$ the more rapidly $x_e(t) \rightarrow x_p(t)$ as $t_0 \rightarrow tmax$. As with resilience, a larger reactivity value indicates greater stability in that the system reacts more quickly to remove the affects of the perturbation. The more reactive a cluster of genes the more rapidly the cluster reacts to the perturbation to restore steady state expression. A comparably larger resilience and reactivity denote greater stability (Harrison, 1979; Sutherland, 1981; Nakajima, 1992; Ives, 1995).

Resistance is defined as the maximum displacement from the pre-perturbed state (Harrison, 1979; Pimm, 1984) and is defined mathematically as (Nuebert and Caswell, 1997):

$$\text{Resistance (unitless)} \equiv \max\left(\frac{\Delta x(t)}{x(t_0)}\right). \tag{2.3}$$

Resistance is the ability of the gene cluster to withstand the perturbation. It has been compared to the concept of buffering capacity in ecology (Grimm et al., 1992). On a molecular scale, the less resistant a gene cluster the larger the magnitude the displacement.

2.4 The Moment of Area as a Tool to Calculate Strain for Generalized Response Envelope

A dilemma originates when dealing with multiple response envelopes and multiple stability measures (i.e., number of parameters). Consider a case where resistance is high but reactivity and resilience are small because the slopes of each limb of the response envelope are small. A system could be more stable with respect to one parameter and less stable with respect to another making an overall determination of stability difficult. The problem is compounded when trying to compare multiple response envelopes Therefore, a single measure of stability that integrates multiple envelopes and multiple stability parameters is required. Hashsham et al., (2000), presented the Moment of Area as a measure

of overall stability incorporating resilience, resistance, and reactivity. They proved the utility of this measure in comparing the functional stability of anaerobic reactor communities perturbed by glucose. Mathematically, the Moment of Area is defined as:

$$\text{Moment of Area (time}^2) \equiv \sum_{0}^{i} A_i t_i \,. \tag{2.4}$$

Where, A_i is the area under the response envelope for the ith compound and t_i is the moment arm of the envelope (Figure 2.4). Modifying the original definition, A_i can represent the area of the response envelope and t_i its corresponding moment arm, for a single gene, an aggregate of clustered genes, or set genes involved in a biochemical pathway.

The moment arm is calculated as the distance from the center of mass of the response envelope to the perturbation axis, a vertical line extending from the starting point of the perturbation (Hashsham et al., 2000). The center of mass represents a point of symmetry on which the stability parameters act, similar to describing the center of mass of an object on which distributed forces act. As a component of the Moment of Area, the moment arm accounts for possible lag time from the onset of the perturbation to the gene cluster response, effect of the shape of different amplification envelopes, and emphasizes the response that occurs over a longer duration. Units associated with different types of perturbations, such as temperature or concentration does not appear in the Moment of Area. Therefore, differences between Moments of Areas are a

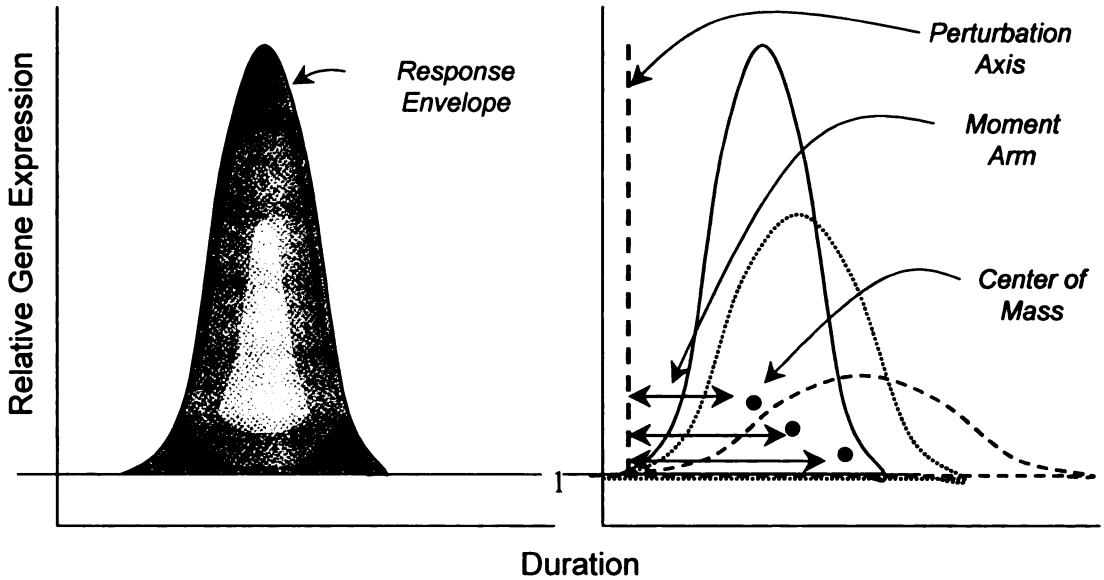

Figure 2.4. The Moment of Area encompasses the the stability parameters and is measure of overall stability. Components of the Moment of Area applied to asymptotic stability are shown. Components include response envelope, perturbation axis, moment arm, and center of mass.

reflection of differences in stability, which allow for comparisons between i) commonly expressed genes across different environmental perturbations, ii) the response of a set of genes to different magnitudes of the same perturbation, or iii) the variations just described but with different genomes. As with the other stability parameters a larger Moment of Area indicates a lower degree of stability.

The response envelope encompasses all the stability parameters described previously. To demonstrate this, consider the asymptotic stable response of a single gene, *i*. Dividing the response of *i* at *tmax,* the area of the response envelope for each half can be found by integrating (Figure 2.5):

$$A_i = \int_{t_0}^{t\max} \frac{x_{pi}(t) - x_{ei}(t)}{x_i^*(t_0)} dt + \int_{t\max}^{\infty} \frac{x_{pi}(t) - x_{ei}(t)}{x_i^*(t_0)} dt . \qquad (2.5)$$

Where, $x_{pi}(t)$ is the perturbed expression, $x_{ei}(t)$ is the expression prior to the perturbation and $x_i^*(t)$ describes the initial expression prior to the perturbation of gene *i*. If expression is at steady state prior to the perturbation, then $x_i^*(t) = x_{ei}(t)$. Making the substitution, $\Delta x_i(t,) = x_{pi}(t) - x_{ei}(t)$, taking the anti-derivative of Equation 2.5, and assigning the appropriate limits results in Equation 2.6.

$$\left(\frac{1}{x_i^*(t)}\right)\frac{d\Delta x_i(t)}{dt} = \lim t_0 \to t\max \left(\frac{1}{x_i^*(t)}\frac{d\Delta x_i(t)}{dt}\right) + \lim t\max \to t_\infty \left(\frac{1}{x_i^*(t)}\frac{d\Delta x_i(t)}{dt}\right).$$

Equation 2.6 describes the amplification envelope of gene *i* as the overall rate of change of relative expression governed by reactivity for $t_0 < t \leq tmax$ and resilience for $tmax \leq t < t_\infty$. If the relative response of gene *i* exhibited neighborhood stability, then there is no return to pre-perturbed steady state. Hence the resilience portion of Equation 2.6 is non-existent and the overall rate of change of relative expression is governed by reactivity only.

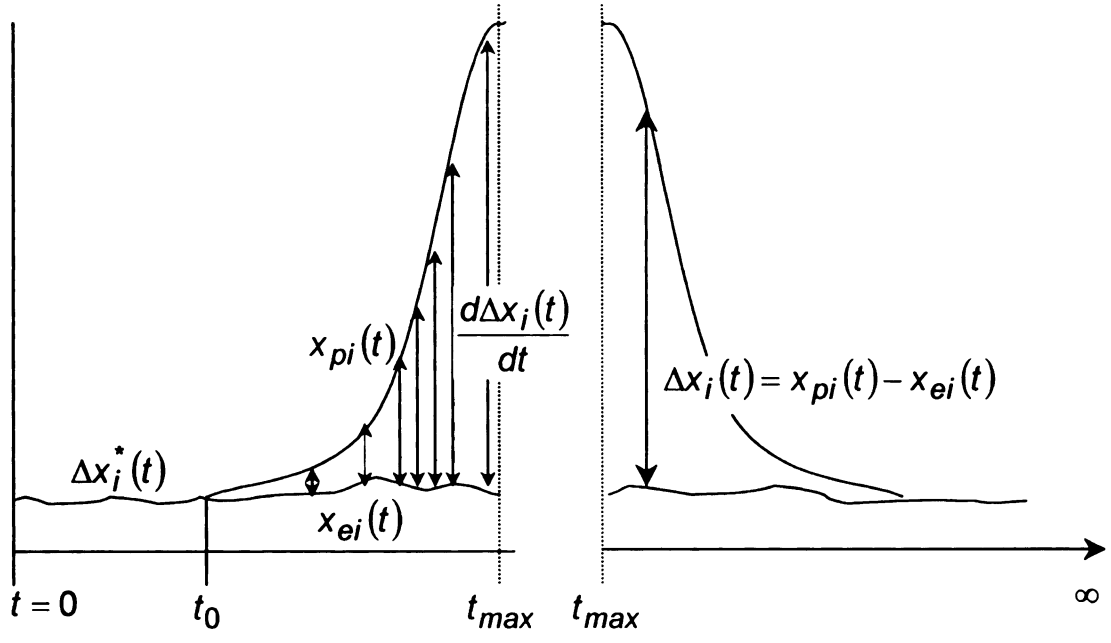

Figure 2.5. Symbols applied to the response envelope showing how it encompasses the stability parameters reactivity, resilience, and resistance.

2.5 Equations used to Calculate the Aggregate Response Envelopes

The above exercise for a single gene can be expanded to consider an aggregate response of a group of genes, with the Moment of Area being the sum of aggregate response envelopes for multiple gene sets or clusters multiplied by their respective moment arms. Two approaches for calculating the aggregate response are shown mathematically by Equations 2.7 and 2.8 The cumulative response equation (Equation 2.7) sums the variance of each gene over all genes in the set or cluster at each measured time point according to:

$$C(t) = \sum_{1}^{i} \sqrt{\frac{\left[x_{pi}(t) - x_i^*(t)\right]^2}{\left[x_i^*(t)\right]^2}} \,. \qquad (2.7)$$

Where, $X_{pi}(t)$ is the perturbed gene expression and $X^*_i(t)$ is the expression prior to the perturbation. The relative displacement equation is similar to equation 2.7 with the exception that the square root is taken after summing the variances according to:

$$D(t) = \sqrt{\sum_{1}^{i} \frac{\left[x_{pi}(t) - x_i^*(t)\right]^2}{\left[x_i^*(t)\right]^2}}. \qquad (2.8)$$

Equation 2.8, designated in this dissertation as the relative displacement equation, was initially proposed to compare the resilience of energy flow among different ecosystems (O'Neill, 1976), and has found applications in determining the stability of nutrient cycling (DeAngelis, 1980; DeAngelis et al., 1989; DeAngelis 1992; Cottingham and Carpenter, 1994). Its application to the stability of gene expression as an aggregate measure is novel and represents a new quantitative tool for analyzing microarray data.

2.6 Physiological Response of *S. cerevisiae* to Osmotic Shock

S. cerevisiae, like all other unicellular organisms, has developed mechanisms to counteract changes in transmembrane water potential due to natural fluctuations in external osmolytes. Because of its commercial importance, the response of *S. cerevisiae* to osmotic shock has consistently been a focus of attention. Exposure to salt solutions, used in de-watering processes during mass production, subject *S. cerevisiae* to high osmotic pressures (Attfield, 1997).

During dough fermentation the osmotic pressure experienced by *S. cerevisiae* has been shown to nearly double (Benítez et al., 1996). Thus, development and identification of new yeast strains with increased osmotolerance is an ongoing research interest.

Three physiological events occur in *S. cerevisiae* upon encountering sudden increases in osmotic conditions. In the first event, a decrease in turgor from loss of intercellular water results in the reduction of the cell volume (Blomberg, 2000). Depending on the severity of the osmotic shock the cell volume has been shown to decrease by as much as 30% to 60% its original volume (Benítez et al., 1996; Hohmann, 1997; Klis et al., 2002). The ability of the cell wall to withstand such a dramatic change is mainly attributed to the presence of the elastic properties associated with the membrane protein β 1,3-Glucan, which has a shape comparable to a wire spring (Klis et al., 2002).

In the next event, growth ceases until adaptation to the new environment. For *S. cerevisiae* adaptation primarily occurs by accumulation of glycerol in the cytosol through internal production and by restricting transport to the outside through the glycerol-specific channel protein, fps1p (Andreishcheva and Zvyagilskaya, 1999; Blomberg, 2000). Trehalose, another osmolyte, has also been observed to accumulate in the cytosol. However, the level of accumulated trehalose compared to glycerol indicates that it does not have a substantial impact on osmotic resistance (Blomberg, 2000). The accumulation of inorganic osmolytes such as KCl and NaCl to counteract osmotic stress has not been observed in *S. cerevisiae*, and is rarely observed in any other unicellular

organism, because of severe impacts on metabolic function due to ionic properties associated with these inorganic compounds (Yancey et al., 1982).

In the final physiological event, the cell volume expands and growth resumes. Yet, the imposed stress does have its costs. The cell volume does not return to its pre-stressed state, growth rate is retarded, and population viability is impacted. The extent of these costs varies with *S. cerevisiae* strains. Hohmann and Meger, (1997), indicate that for the commonly studied W303-1A and YPH strains growth has been observed at 1.7 M NaCl; whereas, some strains show no growth in as little as 0.35 M NaCl. In terms of viability, most strains exhibit a 10% survival rate after addition of NaCl to a final concentration ranging from 1.0 M to 1.38 M (Blomberg, 1997). However, viability has also been shown to depend on the stage of growth. Yeast populations are less viable after an osmotic shock during early log phase growth compared to an osmotic shock during late log phase growth. The reason for increased sensitivity during early log phase has not been clearly explained, but could be a result of thermodynamic or genetic considerations.

2.7 Transcriptional Response of *S. cerevisiae* to Osmotic Shock

The transcriptional response of *S. cerevisiae* to osmotic shock can be grouped into two categories: the general stress response, which is associated with the expression of a number of stress genes across many different perturbations, and the specific stress response associated with expression of genes to a specific perturbation. In the case of osmotic shock, if NaCl is used as the osmolyte generating the perturbation, a slightly expanded specific response

is observed compared to the specific response generated following the addition of sorbitol. This occurs because metal toxicity associated with Na^+, generates an additional response compared to an equivalent stress caused by addition of sorbitol. In spite of this, NaCl is generally preferred over sorbitol because of lesser amounts required to generate the genomic response. A 1.2 M NaCl is equivalent to 1.8 M sorbital for generating a water activity of 0.960 (Hohmann and Mager, 1997).

The general stress response in *S. cerevisiae* is connected to a cis-regulatory element common to genes showing expression to heat shock, oxidative stress, nutrient starvation, and other encountered stresses. This cis-regulatory element was first identified by comparison of promoter regions for CTT1, (a catalase), DDR2 (involved in DNA damage repair), and HSP12 (heat shock chaperon) as a five base pair consensus sequence, CCCCT, and was subsequently named the stress response element (STRE) (Ruis and Schüller, 1995). Targeting this consensus sequence are the transcriptional factors Msn2/Msn4p that are negatively regulated by the cAMP-PKA signaling pathway. A search of the yeast genome for this element resulted in 186 potentially regulated STRE genes with roughly 50 of these known to be functionally related to stress response (Kobayashi et al., 1993; Marchler et al., 1993; Martinez-Pastor et al., 1996; Eustruch, 2000.).

Despite the presence of the STRE, genes containing this element often show different patterns of expression (Estruch, 2000). This has been explained by the presence of additional stress induced regulatory sequences in the

promoter regions. For example, Gasch et al., (2000), observed that expression of the TRX2 cluster was induced for a variety of stresses, but super-induced upon change of the cellular redox potential. Genes in this cluster contained an additional promoter sequence targeting the transcription factor Yap1p, responsible for inducing transcriptional response under oxidative stress (Fernandes et al., 1997). Therefore, the STRE in combination with additional condition specific promoter sequences result in enhanced expression of genes related to the specific stress. This redundancy adds to the assurance that the appropriate genes necessary to mediate the effects of the stress are transcribed.

The use of high throughput technologies for studying gene expression has of lately expanded the notion of the general stress response to beyond those genes solely containing the STRE. Gasch et al., (2000), observed differential expression in roughly 1200 genes (constituting ~19% of the genome) to different stresses. Out of this, the expression of 180 genes was significantly impacted by either Msn2/Msn4p double mutants or over-expression of Msn2/Msn4p. Similarly, Causton et al., (2001), observed differential expression in roughly 10% of the genome, with 47 of 216 induced genes having the STRE site. Additionally, of the 283 genes exhibiting repression, 133 contained a previously unknown consensus sequence, GATGAG, which occurred within 600 bp of the transcription start site. All of this points to a much more complicated general stress response composed of multiple and redundant transcriptional regulatory sites, the products of which may be directly and indirectly involved in mediating the effects imposed by the stress.

The main mechanism for the accumulation of glycerol results from induction of GPD1, GPP2, and a host of other supporting genes with numbers estimated from 200 to 1000 (Mager et al., 2002). Translation of GPD1 and GPP2 results in a NAD^+ dependent glycerol-3-phosphate dehydrogenase and a glycerol-3-phosphatase, respectively. As shown in Figure 2.6, dihydroxyacetone phosphate (DHAP) produced from the isomerization of glyceraldehyde-3-phosphate during glycolysis is catalytically reduced by Gpd1p to glycerol-3-phosphate (Gly-3-P). Gly-3-P is then dephosphorylated by Gpp2p producing glycerol.

Two homologous genes for GPD1 and GPP2 have also been identified and named GPD2 and GPP1, respectively. Although GPD1 and GPD2 produce glycerol-3-phosphate dehydrogenases their physiological assignments have been shown to differ. Evidence for this was demonstrated by protein deletions in which Gpd1pΔ mutants exhibited normal growth under anaerobic conditions, but poor glycerol production at high osmolarity (Larsson et al., 1993). Whereas, Gpd2pΔ mutants exhibited poor growth under anaerobic conditions when glycerol was needed for redox regulation, but normal growth under osmotic shock (Hohmann and Mager, 1997). Differential expression of GPP1 and GPP2 have been demonstrated in the presence of increased external osmolytes indicating functional similarity (Norbeck et al., 1996; Martijn et al., 2000).

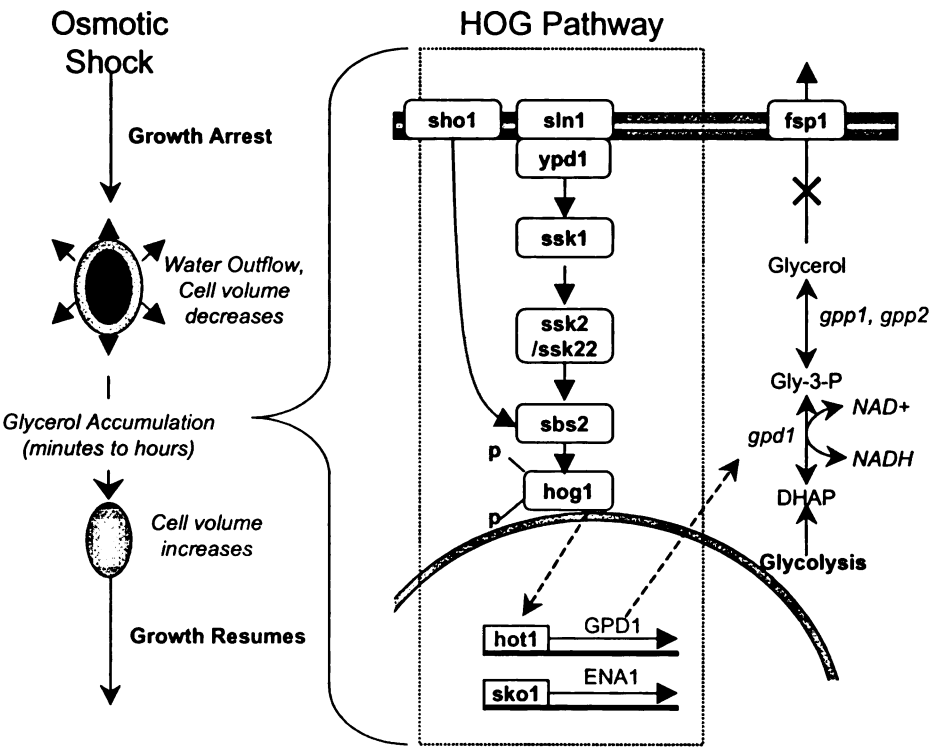

Figure 2.6. Physiological and specific transcriptional responses to osmotic shock. Upon encountering an osmotic shock cell growth ceases until suitable glycerol accumulation, triggered by the high osmotic glycerol (HOG) pathway, allows for growth to resume.

2.8 Regulation of Specific Transcriptional Response to Osmotic Shock

GPD1 contains three transcriptional activation sites demonstrating its redundancy for osmotic shock. One site contains the stress response element connected to the general stress response, another site corresponds to the Msnp1 transcriptional factor of which little is known as to its regulation. The final transcriptional activation site corresponds to the Hot1p activator. This activator is regulated by the high osmotic glycerol pathway (HOG), which is one of four known pathways associated with the mitogen activated phosphorylation kinase (MAPK) signaling pathway. Another important pathway that is related to osmotic

shock, and also a part of MAPK, is the PKC pathway, which is responsible for cell wall integrity (Hohmann and Mager, 1997). All pathways associated with MAPK have a characteristic cascade of kinase activity.

As demonstrated in Figure 2.6, following an osmotic perturbation a two component osmo-sensor (Sln1p & Ypd1p) in the cell wall activates the control kinase (Ssk1p) by dephosphorylation, which in turn directly interacts with the MAPKKK (Ssk2p/Ssk22p). A cascade of kinase activity proceeds with activation of MAPKK (Pbs2p) and activation by dual phosphorylation of the MAPK (Hog1p). Active Hog1p is translocated across the nuclear membrane where it directly interacts with Hot1p. Although Hot1p has not been directly observed to bind with the ORF encoding GPD1, glycerol production has been shown to dramatically decrease with a HOT1 mutant (Rep et al., 1999). Hog1p has also been shown to regulate other transcriptional factors including Sko1p, which activates transcription of ENA1 encoding an ATPase involved in Na^+ extrusion (Andreishcheva and Zvyagilskaya, 1999). Alternative activation of Hog1p was demonstrated bypassing the control kinase and MAPKKK by way of the synthetic high-osmolarity sensitive sensor, Sho1p (Maeda et al., 1995).

2.9 Monitoring of mRNA using Reverse Transcriptase Real-Time PCR

Real-time PCR allows for either the absolute or relative quantification of mRNA by monitoring the PCR amplification process in real-time using a fluorescent reporter. The mRNA is first converted to cDNA by reverse transcription. This cDNA is used as template in the amplification process. As the

PCR reaction progresses amplified DNA increases exponentially over the number of selected reaction cycles until one or more components of the reaction mixture become limiting. At this point the rate of amplification decreases and the amplification curve levels off. Quantification occurs by extending an arbitrary horizontal line through the linear portion of the amplification curve, which is an S-shaped curve depicting amplification of the target with increasing cycle number; and finding the corresponding reaction cycle value (x-value) defined as the cycle threshold, Ct. The cycle threshold is compared to a standard curve equating the initial amount of target to Ct. Samples containing a larger initial amount of target, or larger copy number, have smaller Ct values. Whereas, samples with a relatively smaller amount of initial target, or copy number have larger Ct values.

Absolute quantification has been demonstrated in studies measuring the abundance of isolates from mixed communities and mRNAs in gene expression studies (Rantakokko-Jalava and Javala, 2001; Ayala-del-Río, 2002; Wickert et al., 2002). For abundance of isolate(s), standards composed of quantified target sequences are used to prepare the standard curve. For mRNA quantification, however, the standard curve must account for the efficiency of reverse transcription of mRNA to cDNA (Bustin, 2000). This is accomplished by generating target DNA sequences then in-vitro transcribing them back to RNA. The RNA is then reverse transcribed to cDNA along with the RNA of the unknowns.

Measuring the relative abundance of gene expression is less complicated and two approaches are commonly used. In the first approach, a standard curve

is created, but the absolute abundance of target sequence does not need to be determined. Only the dilution number associated with each standard in the curve is required. The dilution number of the unknown is determined from the standard curve and divided by the calibrator i.e., the dilution number of the gene in which the unknown will be relative too. The other approach of measuring relative gene expression from real-time PCR data uses a mathematical computation known as the delta delta Ct ($2^{-\Delta\Delta Ct}$) computation (Livak and Schmittgen, 2001). The $\Delta\Delta Ct$ refers to the difference between $\Delta C_{T,q}$ - $\Delta C_{T,cb.}$. Where, $\Delta C_{T,q}$ is the difference between the cycle threshold of a reference, known as an endogenous control, and unknown sample, q, and $\Delta C_{T,cb}$ is the difference between the cycle threshold of the endogenous control and cycle threshold of the gene selected as the calibrator. This computational method relies on the assumption that the amplification efficiencies of the reference, calibrator, and unknown are approximately equal. An endogenous control must be included for both absolute and relative quantification to account for minor differences in the amount of starting template of unknown samples to be compared. A gene that does not exhibit a change in expression for the condition studied may serve as an endogenous control. Common endogenous controls include ribosomal DNA, GAPDH, and β-actin (Toshihide et al., 2000).

Monitoring real time amplification commonly occurs by way of one of two generalized approaches. The first approach uses DNA binding dyes that directly react with double stranded DNA. SYBR green I is the most commonly used DNA binding dye for monitoring PCR amplification (Karsai et al., 2001). This dye emits

a fluorescent signal upon intercalating with double stranded DNA. As the amount of amplified DNA increases the fluorescence emitted by SYBR green I is proportional. However SYBR green I is unable to discriminate between target and non-target amplified DNA resulting from non-specific binding of primers. Therefore, results are sensitive to false positives. Similarly, the presence genomic DNA contamination in reverse transcribed samples can result in high background fluorescence. Identification of false positives can be accomplished by generating a melting temperature (T_m) curve (Ririe et al., 1997). The T_m is the temperature at which the double stranded DNA of the amplicon separates and is dependent of the nucleotide composition. A peak is generated that should correspond to the specific T_m of the amplicon as the temperature of the reaction mixture is raised. Additional peaks, or broad peaks indicate the presence and significance of non-target amplified DNA.

The second approach depends on hybridization of specific probes containing fluorescent reporter systems to the amplicon of interest. Three techniques make up the second approach for monitoring amplification. The Taqman assay (Perkin Elmer and Applied Biosystems) utilizes a third probe specific to a region on the amplicon bound by the specific template primers. This probe contains a fluorescent reporter on the 5' end and a fluorescent quencher on the 3' end. During the annealing and extension step the probe hybridizes to the amplicon and 5'-exonuclease activity of the DNA polymerase cleaves the probe separating the reporter from the quencher allowing for the fluorescent emission of the reporter to be monitored. The Molecular beacons (Stratagene)

assay also relies on a reporter and quencher system except the probe containing the two dyes forms a hairpin loop due to self-complimentarity on both ends. In the hairpin loop, the dyes are close enough so that the emission from the reporter is quenched. At the correct annealing temperature, conformational transition separates the ends of the probe, which then binds to its complimentary sequence. The separation of the reporter and quencher is large enough so that emission is no longer quenched. The final technique relies on two separate hybridization probes one of which contains a fluorescein dye on the 3' end and the other contains an acceptor fluorophore on the 5' end. The excitation spectrum of the acceptor fluorophore overlaps the emission spectrum of the fluorescein donor resulting in fluorescence resonance energy transfer when the 3' end is brought into close contact of the 5' end. The resulting emission is a red fluorescence. When the two probes are separated, only the background fluorescein donor emission is present. This last technique results in the greatest target specificity because of the combined specificities of the template primers and hybridization probes (Bustin, 2000).

2.10 Current Analysis Tools for Gene Expression Data

Current tools for analysis of relative gene expression data rely heavily on systematic methods originally designed for inferring evolutionary history from hierarchy relationships (Planet et al., 2001). The use of systematic methods, applied to expressed genes, assumes that meaningful insights into molecular function can be inferred from similar patterns of expression (Planet et al., 2001). This assumption is also valid when describing the stability of clusters of genes to

environmental perturbation. Because all systematic methods contain biases, it is important to have a basic understanding of the systematic method selected when grouping genes for stability analysis.

All data obtained from gene expression studies is grouped by placing each gene in a row and each condition studied, or each observed time point for a single condition in a column. This forms the gene expression matrix in which expression is studied by comparing profiles of genes by comparing rows, or profiles of samples by comparing columns (Brazma and Vilo, 2000; Altman and Raychaudhuri, 2001). Each entry in the expression matrix is a gene expression vector thought of as a point in *m*-dimensional space (Brown et al., 2000). Once the expression matrix has been established, analysis proceeds by calculating similarity, dissimilarity, or confidence between expression vectors using a statistical metric, then establishing relationships among the expression vectors using clustering.

Among the numerous metrics available for calculating similarity, dissimilarity, or confidence, Pearson's correlation, including its modifications, and Euclidean distance have found broad application to studying relative gene expression. Pearson's correlation is a measure of directional similarity between two gene expression vectors in which each vector is treated as unit length (Sherlock, 2000). The directional similarity can often be taken around the mean of the two vectors (Figure 2.7a), or around another arbitrarily defined reference

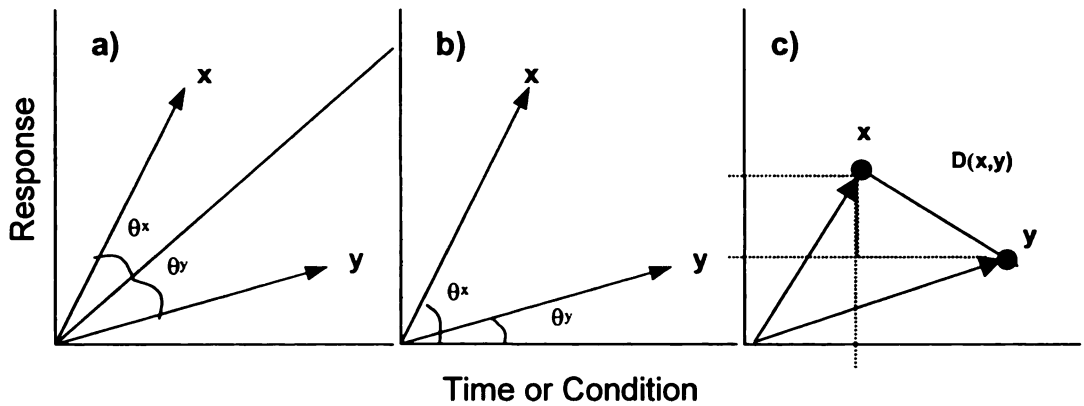

Figure 2.7. Graphical representation of different comparison metrics used for analysis of microarray data. **(a)** Pearson's correlation in which the angle between the mean of the gene expression vectors *x* and *y* is used to calculate similarity. **(b)** Pearson's correlation around zero or often known as the standard correlation, in which similarity for the gene expression vectors *x* and *y* is calculated as the angle from zero. **(c)** Euclidean distance is calculated as the distance between the expression vectors *x* and *y*.

line (Figure 2.7b). Although Pearson's correlation excels at capturing similarity in terms of shape without regard to magnitude it is not robust to outliers (Eisen et al., 1998; Sherlock, 2000). Heyer et al., (1999), found that this lack of robustness resulted in false positives; genes receiving a high score of similarity in terms of function or co-regulation, but in reality are dissimilar.

Euclidean distance is a measure of dissimilarity and takes into account both direction and magnitude (Legendre and Legendre, 1998). As shown in Figure 2.7c, it is calculated as the distance between two points making up the hypotenuse of an arbitrary right triangle using the Pythagorean formula. It is often overly robust by giving high dissimilarity scores to expressed genes with the same shape, but different magnitudes (Heyer et al., 1999). Alternative metrics to Pearson's correlation and Euclidean distance are described in Appendix A. There

is no "absolute" for choosing the best metric, rather consideration should be given to whether the shape of the expression pattern or the magnitude is of more importance (Brazma and Vilo, 2000).

The result of calculating a similarity, dissimilarity, or confidence value for each pair of genes for all genes in the gene expression matrix, is a n^2 matrix, known as a similarity matrix. Values in the similarity matrix are clustered so that important inferences about gene function, and regulation can be made. Most clustering analysis is categorized as supervised or unsupervised. Supervised clustering uses learning algorithms such as support vector machines (SVM) (Brown et. al., 2000), Bayesian networks (Long et al., 2001), and others to define classifiers that represent certain groups of functional genes. Functionality of unknown expressed genes is determined by comparing the classifiers to expression data.

Unsupervised clustering is performed without a priori knowledge to classification and consists of hierarchical and non-hierarchical methods. Hierarchical algorithms form a similarity tree by grouping related genes into clusters and connecting related clusters at nodes. Nodes of related clusters are connected using branches, with the length of a branch indicating the distance between adjoining clusters. Algorithms building a similarity trees in the fashion are described as agglomerative, meaning the tree is formed from top to bottom. Spellman et al., (1998), in a seminal example of applying agglomerative clustering to microarray analysis, associated clusters of genes with their corresponding transcription factors for *S. cerevisiae* during its cell cycle.

The relevancy of genes to their associated clusters when progressing higher up the similarity tree is one disadvantage of agglomerative clustering (Sherlock, 2000). When genes are clustered together, an average expression profile of the profiles making up the cluster is used to represent that cluster (Eisen et al., 1998). As a cluster grows from top to bottom, the average expression profile of the cluster may not accurately reflect the profiles contained in the smaller clusters (Sherlock, 2000). To compensate for this, Alon et al., (1999), demonstrated the use of a divisive clustering algorithm, or a "bottom to top" approach of building a similarity tree. In this approach, genes are randomly assigned to two clusters and then re-assigned to maximize the probability of each cluster. Each cluster is then divided and the processes of re-assignment continues until each cluster is made up of a single expression profile.

Non-hierarchical clustering is useful for grouping gene expression data, collected over time, into families of patterns associated with important events. Common algorithms include K-means (Hartigan, 1975; Heyer et al., 1999), Self-organizing maps (Tamayo et al., 1999; Sherlock, 2000), CAST clustering (Ben-dor et al., 1999), and QT clustering (Heyer et al., 1999). An important example of non-hierarchical clustering was demonstrated by Cho et al., (1998), in characterizing genes associated with the periodic events of DNA replication, chromosome segregation, and mitosis for *S. cerevisiae*. Unlike hierarchical clustering, the user must specify values influencing the number of resulting clusters in the output. This gives the user more control over the clustering process and alleviates the problem of relevancy associated with agglomerative

hierarchical clustering. However, the user should have some external criteria or notions as to the number of clusters desired (Planet et al., 2001). If the number of resulting clusters is too few, important groups of genes are not separated. If the resulting number of clusters is too many, genes that should be grouped together are separated. User control associated with non-hierarchical clustering makes this approach ideally suited for grouping similar responses according the characteristic patterns of gene expression in response to environmental perturbation.

CHAPTER 3

MATERIALS AND METHODS

3.1 Sorting of Gene Expression Data for ESR

Raw gene expression data for the eight environmental perturbations listed in Table1.1 was obtained publicly from the Stanford Microarray Database (www.dnachip.org). ESR genes and their corresponding expression values were located and removed from the raw data by comparison to a downloadable list of these genes, also publicly available on the Stanford Microarray Database. Relative expression was calculated and zero transformed according to (Eisen, 1998):

$$\text{Relative expression} = \frac{CH2DN - CH1D}{CH1D}. \qquad (3.1)$$

The symbols used in equation 3.1 correspond to the data format associated with the software, GeneScan™. Where, *CH2DN* is the background subtracted and normalized emission from fluorescently labeled nucleotides; labeled with the carbocyanine dye derivative, Cy-3, incorporated into cDNA during reverse transcription. Fluorescent emission from this fluorophore represented the transcriptional response of ESR to the perturbations. *CH1D* is the background subtracted emission from cDNA containing incorporated nucleotides labeled with another derivative of the fluorescent carbocycanine dyes, Cy-5, and corresponds to transcription of genes prior to the perturbation.

3.2 Clustering of Relative Gene Expression Data

To separate genes according to the two clusters observed by Gasch et al., (2000), relative expression data was imported into GeneSpring™ software (Silicon Genetics, Redwood City, CA) and a similarity matrix created using a modified Pearson's correlation (Legendre and Legendre, 1998). The non-hierarchical K-means clustering algorithm was used to sort genes into patterns of induction and repression (Brazma and Vilo, 2000). Clustering results were exported to Excel™ and manually sorted to remove genes with missing data points. These results were verified against results obtained by both Gasch et al., (2000), and Causton et al., (2001).

3.3 Calculation of Strain from the Response Envelope

Least squares regression was used in conjunction with the general linear model, $y(t)=\beta_0 + \beta_1 t_1+\beta_2 t_2+\cdots\beta_k t_k+\varepsilon$, to estimate the partial slopes of the response envelopes. The area of each amplification envelope was determined by integrating the general linear model, from the start of the perturbation to the last observation. The center of mass for each amplification envelope was also determined from the general model according to (Fishbane et al., 1993):

$$x\ coordinate = \frac{\int_a^b t \cdot y(t)dt}{\int_a^b y(t)dt}, \quad y\ coordinate = 0.5\frac{\int_a^b (y(t))^2 dt}{\int_a^b y(t)dt}. \tag{3.2}$$

Where, *y(t)* is the specific model for each response envelope, *b* is the time of the last observation, and *a* is the starting time of the perturbation. The aggregate response equations (Equations 2.7 and 2.8) were used to calculate the relative displacement as described previously. Results of the two methods were compared for the ESR genes.

3.4 Calculation of the Strain for Large Data Sets

To expedite the process of calculating the Moment of Area for a large number of response envelopes a program was developed using MathCad™. A flow diagram for the calculation approach is displayed in Figure 3.1. The program was made up of four parts. Data representing amplification envelopes of cluster aggregates, or individual genes were imported as a text file to part one. Here, linear regression takes place using the least squares method to calculate the coefficients (partial slopes) of the regression model (Longnecker and Ott, 2001). These coefficients were then exported as a text file, and used in parts two and three.

In part two, a statistical analysis of the regression model was done by calculating the sum of squares of the residuals, sum squares of regression, and total sum of squares. The above sum of squares was used to calculate the residual standard deviation, and coefficient of determination. These statistical measures are used to evaluate the error associated with the regression model, goodness of fit, and predictive capability, respectively. Statistical equations to calculate these parameters can be found in Longnecker and Ott, (2001).

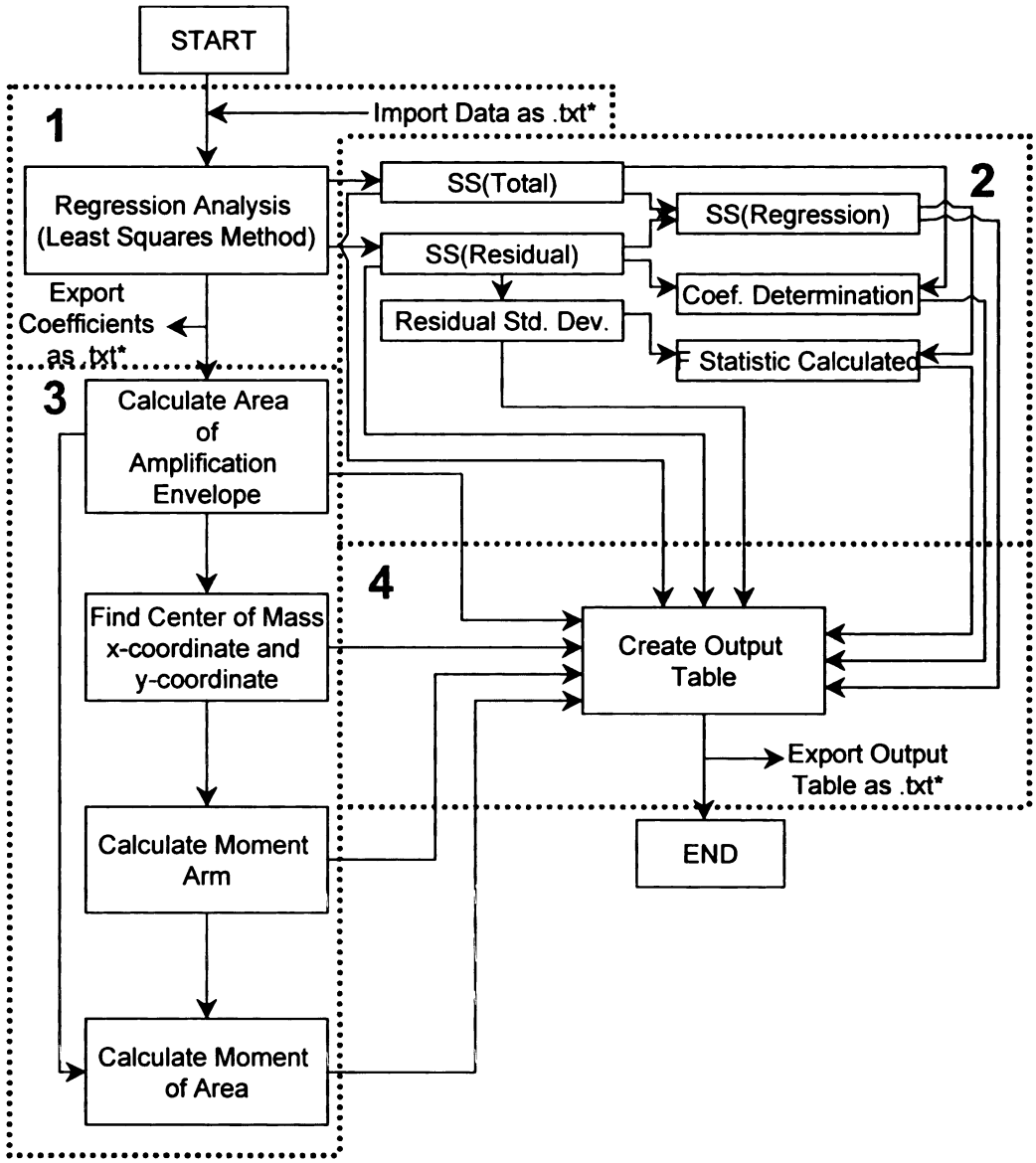

Figure 3.1. Schematic of MathCad™ program used to calculate the Moment of Area for large data sets. The program is divided into four sections. Section one uses linear regression to fit the data to a statistical mode. Section two calculates the statistics associated with the regression analysis. Section three finds the components leading to the Moment of Area calculation. Section four creates an output table containing all of the results.

In part three, the coefficients of regression were used to find the area, x-coordinate, and y-coordinate of the center of mass according to Equation 3.2, and moment arm of the response envelopes. Using the areas and moment arms, the Moments of Area were calculated using Equation 2.4. Results from parts one, two, and three were exported to part four of the program where an output table was generated and exported as a text file. A detailed description of each part of the MathCad™ program is found in Appendix B.

3.5 Selection of Genes and Primer Design

One hundred sixty nine genes were selected for observation to osmotic shock. The selection criteria were based on: 1) whether the genes had previously exhibited significant expression to osmotic shock from published microarray data (Gasch et al., 2000; Martijn et al., 2000; Causton et al., 2001), 2) whether the genes had been previously described in connection to osmotic shock, but did not exhibit significant expression indicated by microarray results (Hohmann and Mager, 1997), and 3) whether the genes were connected with one of the four biochemical pathways noted as important to osmotic shock. These pathways included one regulatory pathway-the MAPK signaling pathway, and three metabolic pathways- glycolysis/gluconeogenesis, glycerolipid metabolism, and cell cycle. Identification of genes associated with the four biochemical pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000).

In addition to the 169 genes, RDN18-1 that encodes 18S rRNA was included as an endogenous control for normalization. Components of the

ribosome in eukaryotes are often used for normalization (Toshihide et al., 2000). However, post-transcriptional modifications can make designed primers ineffective for amplification of cDNA (Venema and Tollervey, 1999). Therefore, YBR011C (IPP1) that encodes a pyrophosphatase was also included as an endogenous control (Kurilova et al., 1993). Changes in expression of YBR011C to shifts osmotic shock have not been observed, and its use for normalization has been demonstrated (Norbeck and Blomberg, 1997; Martijn et al., 1999).

Primers for each gene were designed using Primer Express™ 2.0 (Perkin Elmer Applied Biosystems, Foster City, CA). Design parameters included a melting temperature ranging from 58° C to 62° C, GC content ranging from 45% to 55%, length ranging from 20 bp to 22 bp, and amplicon length (size of the amplified PCR product) ranging from 100 bp to 110 bp. Short amplicons are ideally suited for quantitative PCR because of better amplification efficiency due to an increased likelihood for complete denaturing. This allows effective binding of primers to their target sequences. PCR efficiency also increases due to shorter extensions times required by polymerization (Bustin, 2000). If possible, at least one primer was designed to cross an intron; hence, increasing primer specificity for the target (Wickert et al., 2002). Additional design parameters not listed here, but included in the software were set according to the manufacturer's suggestions. Appendix C describes the 169 genes, including primers, according to the four biochemical pathways mentioned above.

Specificity of primers was evaluated by: 1) using stand alone BLAST and the yeast nucleotide database contained in GenBank (www.ncbi.nlm.nih.gov), 2)

ethidium bromide staining of the amplified products using gel electrophoresis, and 3) analyzing a dissociation curve for each primer pair following RT-PCR using an ABI Prism 7900HT (Perkin Elmer Applied Biosystems, Foster City, CA). Primers showing fewer than 3 mis-matches to non-target sequences as indicated by BLAST, were discarded. Primers resulting in the formation of multiple amplicons as indicated by double bands on gels following electrophoresis were also discarded. Finally, dissociation curves with multiple peaks were matched to their primer sets and disregarded in the final data analysis.

3.6 Experimental Approach for Quantitative Stress-Strain Response

An overview of the experimental approach used to study osmotic shock in *S. cerevisiae* is outlined in Figure 3.2. Osmotic shock was induced using a 5.0 M NaCl solution added to batch cultures resulting in press perturbations of increasing magnitude (0.5, 0.7, 1.0, 1.2, and 1.4 M). Previous experiments showed a final concentration of 0.4 M NaCl was required to induce osmotic shock; whereas a 1.4 M NaCl concentration resulted in no resumption of growth (Wuytswinkel et al., 2000). Biological replication of each perturbation was conducted in triplicate. In addition, negative controls were included for each perturbation to monitor gene expression under non-perturbed growth conditions.

An average of eight time point samples, including time zero, were collected for each replicate and negative control for the six perturbations resulting in roughly 180 samples to be processed in preparation for relative gene expression quantification. Total RNA was extracted and quantified followed by detection of genomic DNA contamination using no RT control procedures. If the

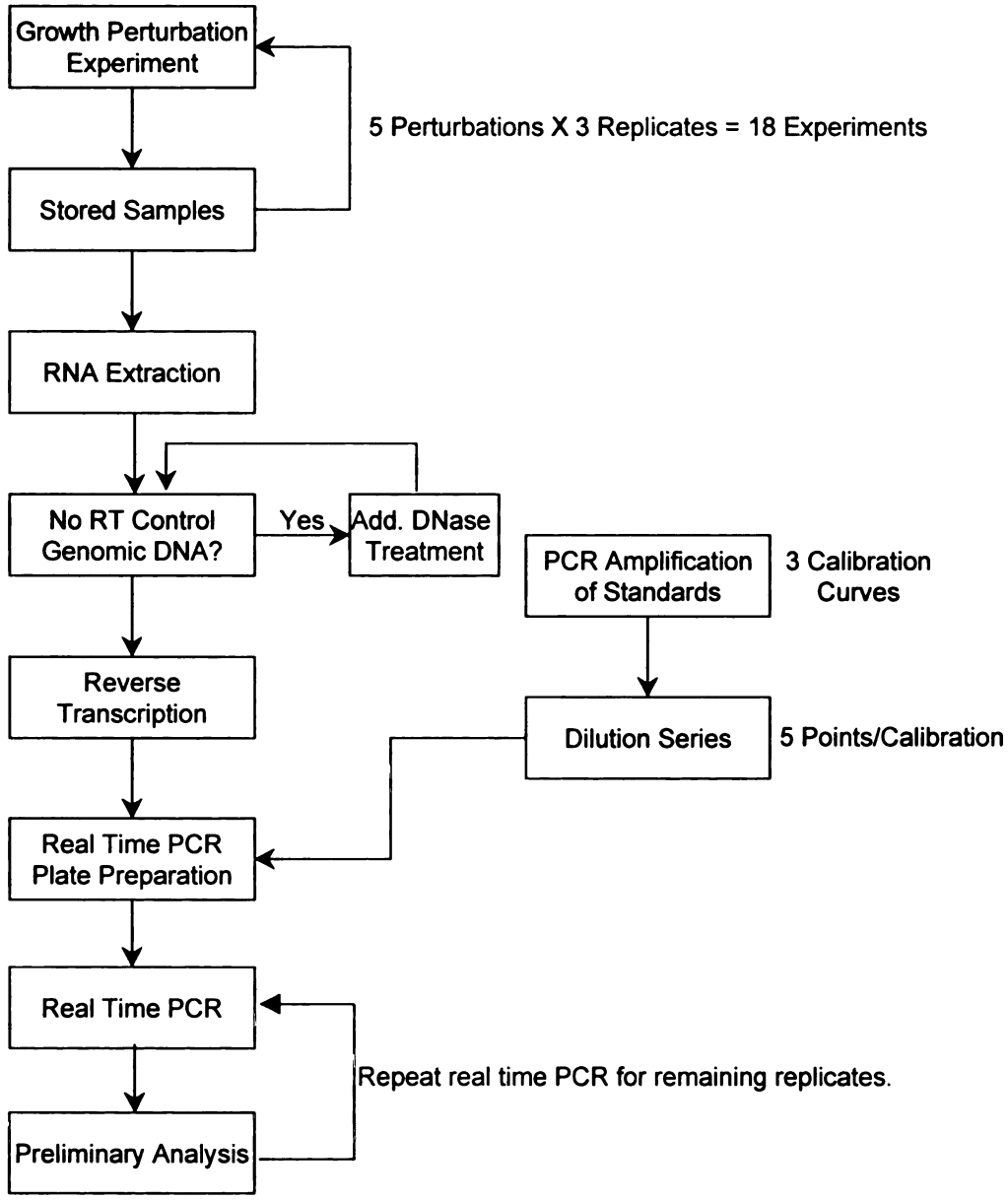

Figure 3.2. Design approach used for conducting osmotic shock experiments. RNA was extracted from roughly one hundred eighty samples checked for genomic DNA contamination and reverse transcribed into cDNA. Standard curves for real time PCR were developed in parallel to the RNA extractions. In all, roughly 18,000 PCR reactions were carried out in 384 well optical plates.

presence of genomic DNA was significant, additional enzymatic treatment occurred until the presence of genomic DNA could not be detected using no RT controls. Total RNA was then reverse transcribed to cDNA, purified, and quantified in preparation for real time PCR.

Real-time PCR was conducted on six of the eight time point samples using SYBR I green as the fluorescent reporter to quantify gene expression relative to accompanying calibration curves. Each time point sample for each perturbation acted as template for the 169 genes shown previously to be differentially expressed to osmotic shock. Preliminary analysis showed success in the experimental design for the five perturbations. In all, more than 18,000 PCR reactions were carried out in roughly 50 384 well plates, which included calibration standards, negative controls, and no template controls.

Growth of S. cerevisiae and application of osmotic stress. *S. cerevisiae* strain W303 was obtained from the American Tissue and Culture Collection (ATCC 200060). Cells were grown in batch culture at 30° C in YDP medium containing 2% bacto-peptone (DIFCO, Detroit, MI), 1% yeast extract (DIFCO, Detroit, MI), and 2% glucose. A 10x solution containing the peptone and yeast extract was prepared separate from a 10x solution of glucose and then combined following sterilization to prevent a darkening of the medium.

Application of osmotic shock to *S. cerevisiae* proceeded by harvesting an overnight culture by centrifugation at 3500 rpm for approximately five minutes. Cells were suspended in YDP without glucose and diluted to an OD_{600} of 0.2.

The cell suspension was divided into four equal parts of 360 ml and 40 ml of a 1.11 M glucose solution was added. Final OD_{600} of the cell suspension with glucose was approximately 0.16. The four cell suspensions were placed on an environmental orbital shaker set at 100 rpm and 30° C. Growth was monitored using a Shimazdu UV-160 spectrophotometer until early log phase growth (approximate OD_{600} 0.4). An appropriate volume of a 5 M NaCl solution was added to three of the cell suspensions to create a final NaCl concentration of either 0.5 M, 0.7 M, 1.0 M, 1.2 M, or 1.4 M. The fourth cell suspension not receiving any NaCl served as a negative control for gene expression not related to osmotic shock.

Harvesting of cells during osmotic shock. Just prior to the onset of osmotic shock a 15 ml sample was collected from each cell suspension and used to represent gene expression at time 0. During the perturbation cells were harvested a regular intervals from all cell suspensions until growth resumed, indicated by three consecutive increases in OD_{600}. To avoid gene expression not related to osmotic shock, harvesting proceeded according to established protocols (Causton et al., 2001.). Briefly, 15 ml volumes were withdrawn from each cell suspension and pelleted by centrifugation at 3500 rpm for 3 minutes at room temperature. The medium was decanted and pellets were immediately plunged into liquid nitrogen then stored at –80° C until RNA extraction, which occurred within one month of harvesting to avoid degradation.

RNA extraction. Total RNA extraction of the stored samples proceeded according to modifications of the hot acid phenol extraction method developed by Schmitt et al., (1990). While being kept on ice, cell pellets were suspended in 400 µl AE buffer (50 mM sodium acetate, 10 mM EDTA [pH 5.3]) and transferred to 1.5 ml centrifuge tubes. Immediately, 40 µl of a 10% SDS solution was added followed by addition of an equal volume of double distilled phenol at pH 5.3 (Ambion, Austin, TX) equilibrated in AE buffer and prepared within one week of use. Samples were mixed, incubated at 65° C for 6 minutes, then rapidly chilled in a dry ice ethanol bath until the appearance of phenol crystals. Following the chilling step, samples were allowed to thaw slightly then centrifuged at room temperature for 2 minutes at 14,000 rpm. The aqueous phase was transferred to a new tube and extracted again with acid phenol.

To the transferred aqueous phase, an equal volume of phenol/chloroform/Isoamyl alcohol (Sigma, St. Louis, MO) was added, mixed by vortexing and centrifuged at room temperature for 5 minutes at 14,000 rpm. The aqueous phase was transferred to a new tube. DNase I (Ambion, Austin, TX) was added to the above aqueous sample to achieve a 10 units per ml solution and the sample was incubated at 37° C for 30 minutes. To inactivate the enzyme, an additional acid phenol extraction was performed and the aqueous phase was transferred to a new tube.

Total RNA was precipitated by addition of 0.1 volume 3 M sodium acetate, [pH 5.3], and 2.5 volumes ice-cold 95% ethanol. Samples were placed at –20° C overnight. RNA was harvested by centrifuging at 4° C for 30 minutes at 14,000

rpm. The supernatant was decanted and RNA washed once in ice cold 80% ethanol, and centrifuged at 4° C for 20 minutes at 14,000 rpm. The supernatant was decanted and the RNA pellet was allowed to dry, then suspended in 50 µl Tris-EDTA [pH 8.0] (Ambion, Austin, TX). A 5 µl sub-sample was removed and reserved for RNA quantification, and to check for the presence of genomic DNA contamination.

All solutions and supplies used for RNA extractions were either purchased RNase free or treated with DEPC (diethyl pyrocarbonate) according to standard protocols (Fedorcsak and Ehrenberg, 1966). All glassware was Placed in a 180° C furnace for at least 8 hr to destroy residual RNase (Sambrook et al., 1989). Additionally, prior to each extraction, a hood dedicated to RNA extraction was cleaned to reduce or destroy residual RNase activity using either a 10% bleach solution or RNase wipes (Ambion, Austin TX).

Check for Genomic DNA Contamination. To check for genomic DNA contamination in the extracted RNA samples, a no RT control was performed. Briefly, 20 ng of RNA template obtained from each of the RNA sub-samples and added to a reaction cocktail (Karsai et al., 2002) containing 10 mM Tris-HCl (pH 8.5), 50 mM KCl, 2 mM $MgCl_2$, 0.15% Triton x-100 (Sigma, St. Louis, MO) and 40 µM dNTPs (Promega, Madison, WI). One unit of taq polymerase (Promega, Madison, WI) was added to this mixture along with 12 pmoles of both forward and reverse primers targeting a 104 bp region of the RDN18-1 (18s rDNA) open reading frame. Design of the forward primer (5'-

TGAACCCAATCATCCAAGACAC-3') and reverse primer (5'-TTGTGGGAAAGCACCATAGTTG-3') followed guidelines described in the gene selection and primer design section (Section 3.6).

Final reaction volume of the above components was 25 μl, which underwent PCR following a protocol described previously (Ayala-del-Río, 2002). Briefly, it consisted of a pre-dwell for 2 minutes at 94° C followed by a denaturing step for 30 seconds at 94° C, an extension step for 30 seconds at 55° C, and an elongation step for 1 minute at 72° C. The denaturing, extension, and elongation steps were repeated 40 times. A final post-dwell of 1 minute at 72° C occurred following completion of the repeated steps. A positive control containing genomic DNA template instead of RNA template, and negative control lacking any template were also included for quality control.

Amplified products from PCR were detected using electrophoresis on a 3% (wt/vol) agarose gel (Gibco BRL) in 1x TAE (Tris-acetate-EDTA). The presence of a band of the same size compared to the positive control indicated genomic DNA contamination. It was suspected that poor digestion of genomic DNA resulted from the presence of NaCl, which can significantly reduce activity of the DNase I enzyme. Total RNA in contaminated samples was precipitated as described above and suspended in 500 μl DEPC treated water to which 55 μl DNase buffer (Ambion, Austin, TX) and 6 units of Turbo DNase I (Ambion, Austin, TX) were added. This modified enzyme is more robust to NaCl than traditional DNase I. The mixture was incubated at 37° C for 30 minutes and enzyme inactivated with a phenol/chloroform/isoamyl alcohol extraction.

Precipitation, harvesting, and quantification of the treated samples proceeded as described previously. RNA samples free of genomic DNA were reverse transcribed using a High Capacity cDNA Conversion Kit (Applied Biosystems, Foster City, CA) and purified using Qiagen PCR purification columns (Qiagen, Valencia, CA). Purified cDNA was quantified by measuring absorbency at 260 nm and stored at –80° C.

Real-time PCR. The reaction cocktail used for SYBR™ Green I real time PCR was adapted from Karsai et al., (2002). It consisted of 1x SYBR™ Green I buffer described above. To the reaction cocktail was added 1.0 unit of Taq Polymerase, 0.375 μl of SYBR™ Green I (Molecular Probes, Eugene, OR) diluted from 10,000x to 1000x prior to addition, 2.3 μM 6-ROX (6-carboxy-x-rhodamine, Molecular Probes, Eugene, OR), 10 ng cDNA, and 12.5 pmoles of forward and reverse primers.

Thirteen μl of the reaction cocktail minus primers was distributed to 384 well optical plates (Applied Biosystems, Foster City, CA) using a Biomek 2000 workstation (Beckman Coulter, Fullerton, CA). Two μl of each primer were also added to each well using the workstation. PCR was performed on an ABI 7900HT (Applied Biosystems, Foster City, CA) using the reaction conditions described in the previous section.

Production of Taq. Because the consumption of Taq was expected to be cost-prohibitive, it was produced in the laboratory using *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli* INValphaF' (Invitrogen, San Diego, CA) following protocols established by Engelke et al., (1990), and Pluthero, (1993). *E. coli* containing the taq polymerase gene on pTaq plasmid under the control of the lac promoter was grown in LB broth containing 80 mg l^{-1} ampicillin. Induction of Taq polymerase occurred by addition of IPTG (iso-propyl-1-thio-β-D-galactopyranoside, Sigma, St. Louis, MO) to a concentration of 125 mg l^{-1}. Cells were harvested, lysed, and soluble proteins separated by centrifuging. Soluble proteins containing the polymerase were salted out using ammonium sulfate and dissolved in 50 mM Tris-HCl [pH 7.9], 50 mM glucose, and 1 mM EDTA. Crude purification of the protein extract occurred by dialysis using a Spectra/Por 50,000 MWCO membrane (Spectrum Labs, Rancho Dominquez, CA) against a 50% (vol/vol) glycerol solution with two changes in dialysis solution every 12 hours. The purified extract was diluted in half with a storage buffer solution containing 50 mM Tris-HCl [pH 7.9] 50 mM KCl, 0.1 mM EDTA, 1 mM dithiothreitol, 0.5 mM PMSF (phenylmethylsulfonylfluoride, Sigma, St. Louis, CA), and 50% (vol/vol) glycerol. The diluted extract was stored at –20° C. The Taq polymerase contained in the purified extract was evaluated against purchased Taq polymerase (Promega, Madison, WI) to estimate the activity units per micro-liter.

Calibration curves for real-time PCR. Common practice using real-time PCR includes a standard curve on each plate for each unknown amplicon. This was not economically feasible for the 170 genes selected for this research. Rather, it was assumed that SYBR™ green I could not discriminate between amplicons of different genes, but could discriminate between amplicon length, i.e., the longer the amplicon the more dye intercalates between the two DNA strands resulting in greater fluorescence per amplicon. Two approaches were used for development of standard curves. In the first approach, primers were used to generate PCR products of length 101 bp, 104 bp, and 106 bp representing the smallest, average, and largest amplicon lengths for the 170 genes. The PCR products were purified using 10,000 MWCO spin columns (Millipore, Bedford, MA), and quantified by measuring absorbance at 260 nm. Purified products were then adjusted to the same absorbance. A 1:10 dilution series was created from each purified and adjusted PCR product to be used as the starting template for the standard curves. Four dilutions, 10^{0}, 10^{-2}, 10^{-4}, and 10^{-6}, from each series were evaluated for linearity and agreement in cycle threshold (Ct).

It seemed plausible that poor amplification from the short pieces of DNA, generated in the first approach might occur. Therefore, in the second approach additional primers were designed (Table 3.1) that encompassed the primers used in the first approach. Amplified PCR products using these new primers were roughly 1000 bp. These PCR products were purified using Qiagen PCR purification columns (Qiagen, Valencia, CA), quantified, then adjusted as

previously described. A 1:10 dilution series from each product was created and six dilutions, 10^{0}, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-5}, and 10^{-6}, from each series was used as starting template for the standard curves. Amplification of these templates occurred satisfactorily using the original primers designed in the first approach.

Table 3.1. Genes and their associated primers used in developing standards for relative quantification of mRNAs. Two approaches were used in developing the standard curves. In the first approach amplification was attempted from small DNA templates associated with the genes ranging from 101 bp to 106 bp. In the second approach, larger DNA templates were used encompassing the smaller amplicons to improve amplification.

Primers for First Approach

ORF	*Gene ID*	*Description*	*Forward Primer (5'-3')*	*Tm (C)*	*Reverse Primer (5'-3')*	*Tm (C)*	*Amplicon Length (bp)*
YCR012W	PGK1	phosphoglycerate kinase	CCAGAAAGGTCGATGGTCAAAA	60	CCGAAGGCATCGTTGATGTAA	60	106
YCL004W	PGS1	phosphatidylserine synthase	TTGATGGCCTTCGAGGAACA	60	TCGGCAATCCACTCTCTCTGAC	60	101
RDN18-1*		Portion of 18s rDNA	GCTGGCGATGGTTCATTCAA	58	CCGGAATCGAACCCTTATTCC	59	104

Primers for Second Approach

ORF	*Gene ID*	*Description*	*Forward Primer (5'-3')*	*Tm (C)*	*Reverse Primer (5'-3')*	*Tm (C)*	*Amplicon Length (bp)*
YCR012W	PGK1	phosphoglycerate kinase	GCCAACCATCAAGTACGTTTTG	58	CCACCAGTAGAGACATGGGAGAT	58	1050
YCL004W	PGS1	phosphatidylserine synthase	GCTCCAACTCACTCGTCCTCAT	59	AGTATCCTGCAGTGAACGTCCA	58	1007
YBR011C*	IPPI	pyrophosphate phosphohydrolase	CGCGTCGAAGTTAGGAAGGTT	59	TTGAAGCCGCATACGTGGTAG	60	1037

*RND18-1 and YBR011C W were evaluated as endogenous controls and had different final amplicon lengths of 104 and 101 bp respectively

CHAPTER 4

RESULTS

4.1 Application of the Moment of Area to Calculate Strain

Moment of Area results for the environmental stress response are described using two comparisons. Both comparisons are used to demonstrate that the Moment of Area is relatively insensitive to the method used to calculate it. In the first comparison, Moments of Area resulting from the relative cumulative response equation (Equation 2.7), and the relative displacement equation (Equation 2.8) are compared. These equations were used to describe the aggregate response of the ESR. Their comparison is made to determine which equation is better suited for this purpose and to fulfill the first objective of development of tools for describing strain.

The second comparison looks at the cluster approach used to calculate the Moment of Area. Recall that Gasch, et al., (2000), observed that the ESR was comprised of two clusters, one representing induction and the other repression. To demonstrate the relative insensitivity of the Moment of Area, the Moment of Area resulting from combining both clusters into one (single cluster), then calculating the aggregate response envelope using either equations 2.7 or 2.8 is compared to the summed aggregate response envelopes of each cluster multiplied by their respective moment arms (set of two clusters).

Comparison of equations used to calculate the aggregate response envelope. Moments of Area for the ESR perturbed by the eight stresses (described in Table 1.1) are shown in Figure 4.1. Figure 4.1a compares the Moments of Area associated with the aggregate response envelopes calculated using the cumulative response equation (i.e., Equation 2.7) and Figure 4.1b compares the Moments of Area resulting from the displacement equation (i.e., Equations 2.8). Error bars in Figures 4.1a and 4.1b represent residual error propagated through the Moment of Area calculation. Both aggregate methods resulted in qualitative agreement of the ESR to the different perturbations, but differed by an order of magnitude in their calculated values. Smaller values associated with the displacement equation were expected due to the square root being taken after summation of relative expression values rather than before summation of relative expression values as with the case of the cumulative equation.

Table 4.1 lists the coefficients of variation calculated from the residual standard deviation for each perturbation divided by their respective Moments of Area. This was done to evaluate which aggregate response equation was more suited to the regression model used for linear regression. For all perturbations, except osmotic shock and temperature shock from 25° C to 37° C, the contribution of the residual error to the Moment of Area was greater for the aggregate response envelope described by the displacement equation than the aggregate response envelope described by the cumulative equation. In the case of the two exceptions, the contribution of the residual error to the calculated

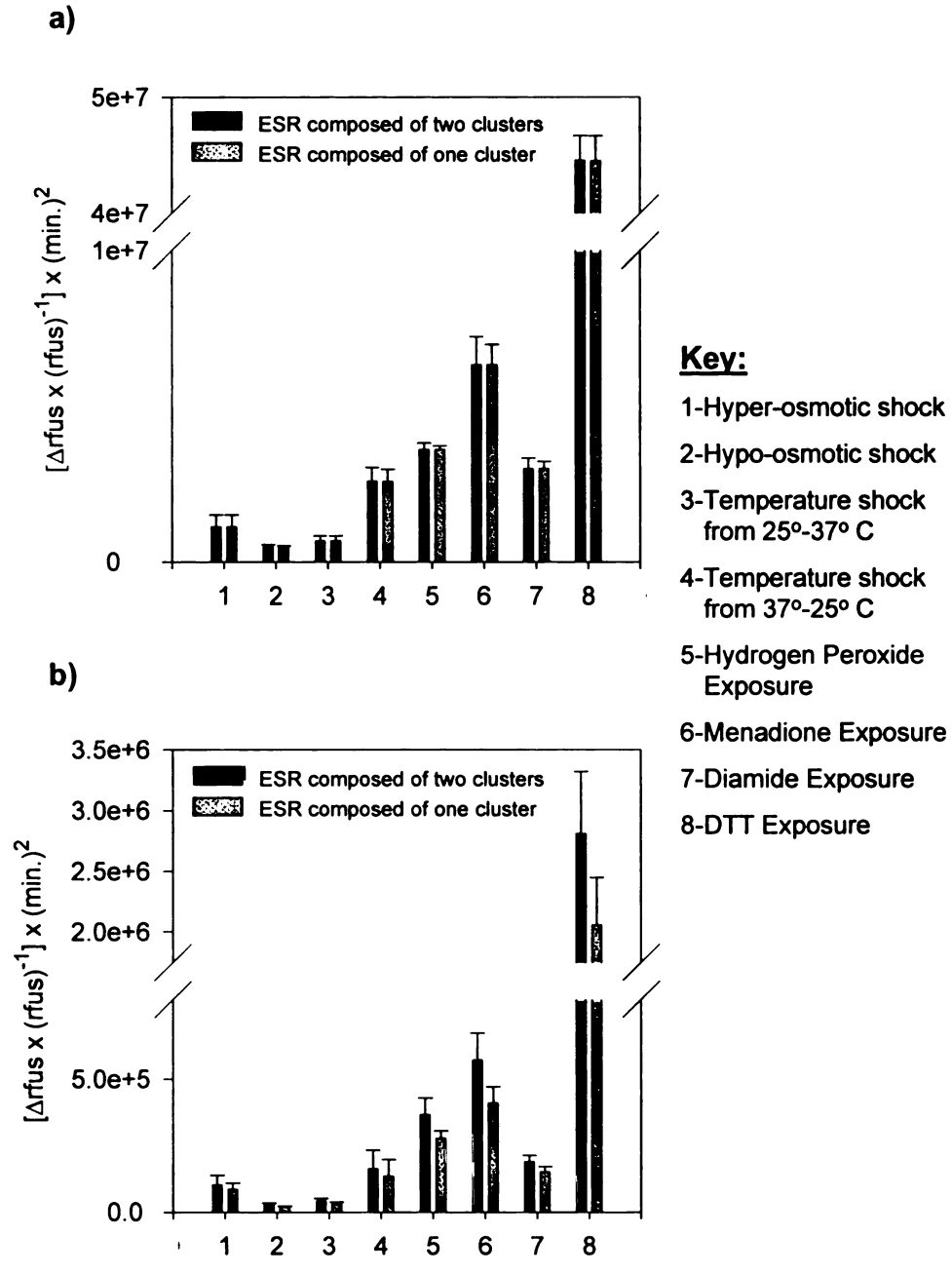

Figure 4.1. Comparison of Moments of Area for the eight perturbations. **(a)** Using the cumulative equation to describe aggregate response of the ESR associated with the set of two-clusters and single cluster approach. **(b)** Using the displacement equation to describe the aggregate response of the ESR associated with the set of two-clusters and single cluster approach.

Table 4.1. Comparison of residuals relative to their respective Moments of Area for the eight perturbations. Relative residuals for the cumulative equation were identical between the cluster approaches. This was not observed for the displacement equation.

Cluster Approach	*Aggregate Response*	*Osmotic*	*Hypo-osmotic*	*Temperature 25°-37° C*	*Temperature 37°-25° C*	*Hydrogen Peroxide*	*Menadione*	*Diamide*	*Dithiothreitol*
Set of two-clusters	Cumulative equation	0.35	0.02	0.23	0.15	0.04	0.10	0.08	0.05
	Displacement equation	0.34	0.14	0.21	0.43	0.17	0.18	0.12	0.18
Single cluster	Cumulative equation	0.35	0.02	0.23	0.15	0.04	0.10	0.08	0.05
	Displacement equation	0.26	0.01	0.22	0.47	0.10	0.15	0.13	0.19

Moment of Area was approximately the same between the two aggregate response equations. Therefore, the regression model was more suitable for the cumulative equation than the displacement equation.

Effect of clusters. Within Figures 4.1a and 4.1b a comparison between the two-cluster approach to represent the ESR versus the single cluster approach (i.e., both clusters combined into one) to represent the ESR is made. In Figure 4.1a the Moments of Area and the residuals associated with the two approaches exhibited 99% agreement. This demonstrated that when the cumulative equation was used to calculate the aggregate response envelopes of the ESR, residual error from linear regression had little impact on the Moments of Area.

However, a different result was observed for Moments of Area for the two cluster approaches in Figure 4.1b. Here, the use of the displacement equation to

calculate the aggregate response envelopes resulted in worse agreement, ranging from 71% to 84%, between the Moments of Area. The lack of agreement can be explained by observing the residual error associated with linear regression propagated to the area of the aggregate response envelopes (i.e., the A_i term in the Moment of Area equation). Table 4.2 compares this propagated residual error (residual area) for the single cluster and set of two clusters. Once again, the summed residual areas of the set of two clusters and the residual area of the single cluster are equal, confirming that the Moment of Area is less subject to error associated with linear regression when using the relative cumulative response equation. However, the summed residual areas of the set of two clusters do not agree with the residual area of the single cluster using the displacement equation. In fact, the summed residual areas of the set of two clusters were consistently larger than the area of the single cluster. This helps explain why the Moments of Area associated with the set of two clusters in Figure 4.1b were consistently larger than the Moments of Area associated with the single cluster.

Hence, linear regression of the aggregate response envelopes calculated from the displacement equation resulted in a worse fit than linear regression of the aggregate response envelopes calculated using the cumulative response equation. This difference in fit may result from taking the square root after summation of the relative expression values required by the displacement

Table 4.2. Residual areas associated with the aggregate response equations and cluster approaches. The residual areas are additive when applying the cumulative equation. This property was not associated with the displacement equation.

	Cumulative Equation		Displacement Equation	
Approach	Residual Area for Osmotic (min)	Residual Area for Diamide (min.)	Residual Area for Osmotic (min.)	Residual Area for Diamide (min.)
Cluster I (Repressed)	3010	1064	480	37
Cluster 2 (Induced)	5479	5300	367	380
Cluster 1 & 2 Summed	8489	4384	847	417
Single Cluster Approach	8489	4384	482	361

equation, which tends to "flatten" the parabolic shape of the aggregate response envelopes. This "flattening" could result in larger differences between the predicted responses using the regression model and measured responses. Thus, a different regression model may be more suitable for the fitting the aggregate response envelopes calculated using the displacement equation. This was not investigated because Moments of Area in Figures 4.1a and 4.1b exhibited qualitative agreement, regardless of the aggregate response equation. However, because the regression model in combination with the cumulative response equation was better suited for Moments of Area, these results were carried on to the next analysis.

Comparison of strain for various perturbations. For the magnitudes of the perturbations tested, the stress imposed on the ESR by exposure to DTT resulted in the largest Moment of Area and hence least stability compared to the Moments of Area for the other perturbations. This comparison cannot easily be made in a quantitative manner without the approach being developed here. Hypo-osmotic shock resulted in the smallest Moment of Area, and hence the

greatest stability of the ESR. Normalization by the Moment of Area of hypo-osmotic shock showed that the stability of the ESR was roughly the same for hypo-osmotic shock and temperature shock of 25° to 37° C, over twice as stable compared to osmotic shock. Additionally, the normalization showed that the ESR for hypo-osmotic shock was five times as stable compared to temperature shock 37°-25° C, six times as stable compared to diamide exposure, approximately seven times as stable compared to hydrogen peroxide exposure, and twelve times as stable compared to menadione exposure. The stability of the ESR to hypo-osmotic shock was ninety times more stable than to DTT exposure.

Such large differences in stability can be evaluated in terms of the stability parameters. Calculated values for the stability parameters are displayed in Table 4.3. These values were found by taking the natural logarithm and fitting the data to the equation $y(t)=\beta_0 + \beta_1 x$, where β_1 is either the value for resilience or reactivity. In terms of resistance, the ESR was most resistant to the hypo-osmotic shock and least resistant to diamide exposure, which is also consistent with their Moments of Area. In terms of reactivity and resilience, the ESR was most reactive and resilient to osmotic shock, but least reactive to DTT exposure and least resilient to menadione exposure. It is interesting to note that although the ESR proved most stable to hypo-osmotic shock it was neither the most reactive nor most resilient. However, it was the most resistant suggesting that this attribute of stability contributed most to the overall stability. Likewise, the ESR showed the smallest stability to DTT exposure even though it was neither the

Table 4.3. Comparison of calculated stability parameters and Moments of Area for the eight perturbations. The differences in Moments of Area can be evaluated in terms of the individual stability parameters.

	Hyper-osmotic	*Hypo-osmotic*	*Temperature 25°-37° C*	*Temperature 37°-25° C*	*Hydrogen Peroxide*	*Menadione*	*Diamide*	*Dithiothreitol*
Reactivity (1/min)	0.163	0.085	0.137	0.074	0.066	0.035	0.086	0.005
Resilience (1/min)	-0.034	-0.015	-0.015	-0.012	-0.015	-0.003	-0.024	-0.005
Resistance (unitless)	69.4	13.9	27.7	63.7	78.1	52.6	118.7	32.9
Moment of Area (min^2)	$1.5x10^6$	$5.1x10^5$	$8.3x10^5$	$3.0x10^6$	$3.7x10^6$	$7.0x10^6$	$3.2x10^6$	$4.7x10^7$

least resilient nor resistant. In this case however, the ESR exhibited the smallest reactivity, by an order of magnitude compared to the other perturbations suggesting it contributed most to the overall stability.

Contribution of clusters to the Moment of Area. The magnitudes of the temperature perturbations and the magnitudes of the osmotic perturbations were the same, but in opposite directions. The stability of the ESR resulting from a decrease in 12° C was less than the stability of increasing the temperature by 12° C. With the osmotic perturbations, the ESR was less stable with increased osmolarity due to the addition of 1 M sorbitol and more stable upon removal of 1 M sorbitol. The difference in stability can be explained by different contributions of the clusters originally observed by Gasch et al., (2000), to the overall stability. The calculated percent contribution of the two clusters to the overall Moment of Area for hyper-osmotic shock showed that the induced cluster contributed roughly 67% compared to the 33% contributed by the repressed cluster. The opposite was observed for the hypo-osmotic shock. The induced cluster

contributed roughly 32% and the repressed cluster contributed 68%. This pattern was not observed for the repressed and induced clusters associated with the temperature perturbations. Here, the repressed clusters contributed most to the overall Moments of Area for both shifts in temperatures, i.e., roughly 58% for the temperature shift from 25° C to 37° C, and roughly 80% for the temperature shift from 37° C to 35° C.

4.2 Development of Calibration Curves from cDNA Targets

Calibration curves were developed using standards composed of three target lengths, 101 bp, 104 bp, and 106 bp. Two approaches were employed in generating amplicons from the targets using RT-PCR. The first approach involved amplification directly from the short targets described above. The other involved amplification of the short targets from larger DNA fragments roughly, 1000 bp, which encompassed the sequence of short targets. Figure 4.2a shows calibration curves resulting from first approach and Figure 4.2b shows the calibration curves resulting from the second approach. Approximate copy number is shown above each dilution. Both approaches showed good reproducibility per size of template, except for the 10^{-4} dilution of RND18-1, which encodes the 18s rRNA. Differences in the Ct values for this dilution ranged from roughly 15 to 25. No apparent explanation for this poor reproducibility was discovered. For the first

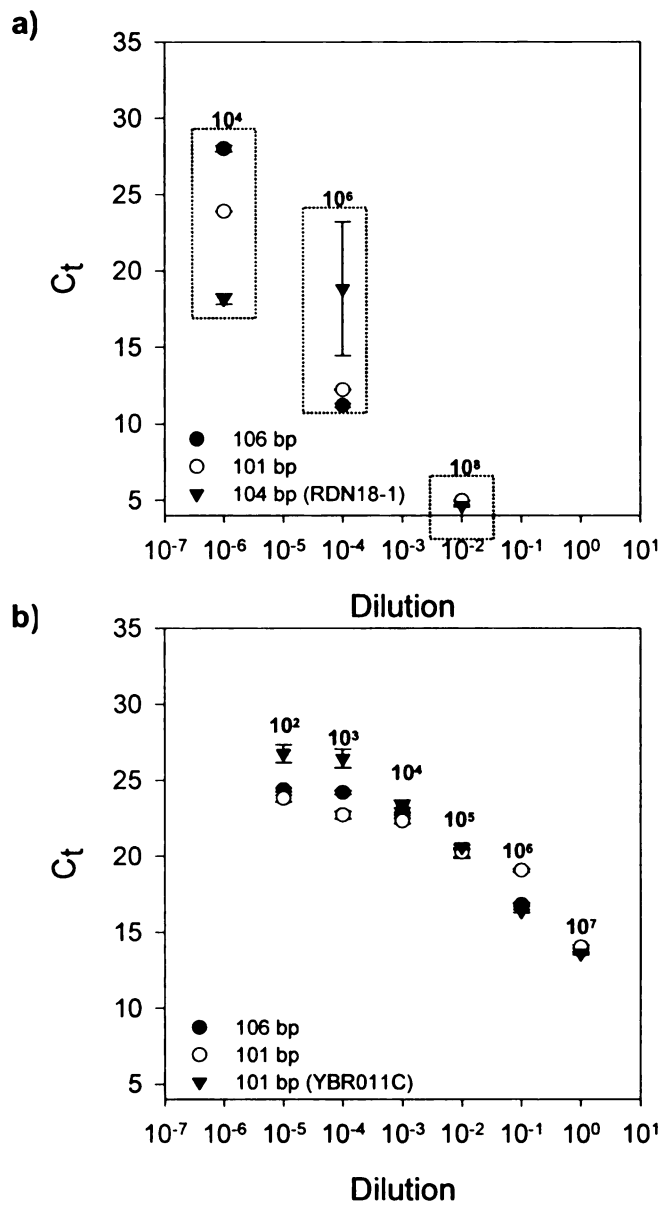

Figure 4.2. Standard curve development results. **(a)** Results from the first approach using small lengths of DNA ranging from 101 bp to 106 bp as template for amplification. Good reproducibility was observed, but large separation between lengths occurred at the 10^{-4} and 10^{-6} dilutions. **(b)** Results from the second approach using longer lengths of DNA encompassing the shorter amplicons. Amplicons were grouped closer together for all dilutions compared to the first approach.

approach (Figure 4.2a), the 10^{-3} dilution containing 10^{8} copies represented the upper bound of amplification for the given PCR components and conditions. No amplification occurred for the 10^{-1} dilution containing 10^{10} copies. The upper bound of amplification for the second approach (Figure 4.2b) was not reached within the range of dilutions tested. However, it appears from visual extrapolation that amplification could have occurred with template copies greater than 10^{7}.

Disagreement between the amplicon sizes increased with dilution for both approaches used to generate template. However, comparison of the two approaches showed that a better agreement between amplicon sizes existed for the second approach. For the 10^{4} and 10^{6} template copy numbers of the first approach, roughly 10 cycle thresholds separated the 106 bp and 104 bp amplicons. In comparison, the number of cycle thresholds of the second approach separating the same amplicon sizes for the 10^{4} copy number was insignificant and less than 4 for the 10^{6} copy number. Significant difference in cycle thresholds of the second approach was not observed until 10^{3} and 10^{2} copy numbers. A statistical F-test comparing the slopes and y-intercepts (calculated for the 10^{0} through 10^{-5} dilutions), between the different amplicon sizes, failed to rejected the null hypothesis (p-value ≤ 0.05) indicating that there was insufficient evidence that one of the slopes and intercepts was statistically different from the others. Thus, it was concluded that the slight difference in amplicon lengths would not statistically influence the resulting cycle threshold values of the unkowns.

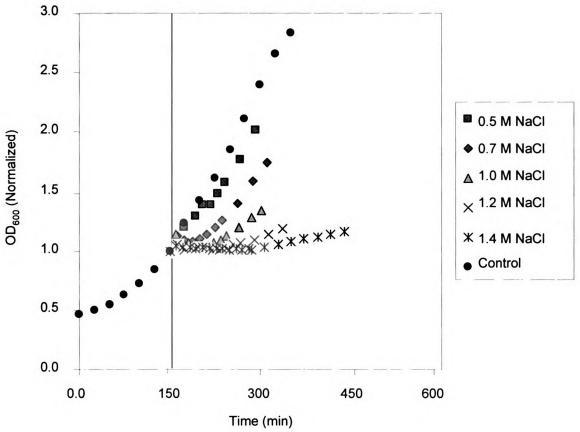

Figure 4.3. OD_{600} measurements normalized by absorbance at perturbation onset for *S. cerevisiae*. The vertical line represents the onset of the perturbation. Retarded growth was seen followed by decreased growth rates with increasing perturbation magnitude.

4.3 Growth of *S. cerevisiae* under the Applied Osmotic Stresses

Growth of *S. cerevisiae* following the five osmotic stresses is shown in Figure 4.3. Absorbance values presented for each perturbation were normalized by their readings at the time of perturbation. Each perturbation occurred at roughly 180 minutes indicated by the vertical line. The effects of the perturbations were demonstrated by an arrest of growth following the perturbation, and a resumption of growth but at a retarded rate. The time of arrested growth

increased with perturbation magnitude and the resumed growth rate decreased. At 0.5 M NaCl growth arrest either did not occur or was not detected by the measurement time scale. The rate of resumed growth was $1.47 \times 10^{-2} \pm 4.5 \times 10^{-3}$ min^{-1}, slightly less than the unperturbed growth rate (1.63×10^{-2} min^{-1}). At 1.4 M NaCl, resumption of growth was not observed for 3 ½ hrs. following onset of the perturbation. The new growth rate, $2.58 \times 10^{-3} \pm 5.6 \times 10^{-4}$ min^{-1}, was an order of magnitude less than the growth rate of the control.

4.4 Reactivity, Resilience, and Resistance in Response to Osmotic Stress

Using the developed tools described in Sections 4.1, gene expression data was analyzed to determine the amount of strain for the five levels of applied stresses (0.5, 0.7, 1.0, 1.2, and 1.4 M NaCl). Only 153 genes out of the original 169 were used in describing the aggregate response envelope. Cycle threshold values could not be determined by real-time PCR for the remaining sixteen genes and therefore they were not included in the analysis. The aggregate response envelopes shown in Figure 4.4 were calculated for each perturbation using the cumulative response equation. Asymptotic stability, i.e., a return to pre-perturbed expression following perturbation, was displayed in response to each perturbation. It was interesting to note that the time associated with the occurrence of maximum displacement increased with an increase in the magnitude of stress. Maximum displacement for the 1.4 M NaCl stress occurred at a time that was approximately 6-fold longer than the maximum displacement for the lowest stress tested (0.5 M NaCl). To see if the stability parameters also

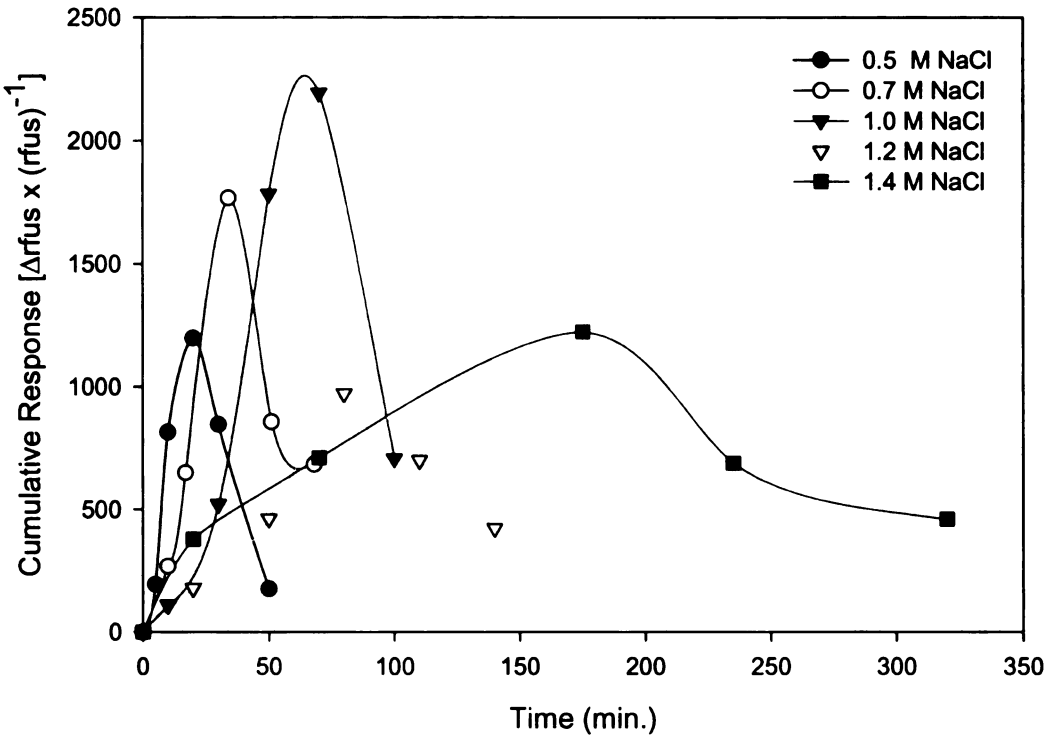

Figure 4.4. Aggregate response envelopes of 153 out of 169 genes showing expression to osmotic shock. The aggregate response was calculated using the cumulative response equation described in Chapter 3. Resistance decreases until the 1.2 M NaCl perturbation. For this perturbation and the 1.4 M NaCl perturbation resistance increases.

Table 4.4. Calculated stability parameters corresponding to the aggregate cumulative response envelopes shown in Figure 4.4.

[NaCl]	*Reactivity (min^{-1})*	*Std. Dev.*	*Resilience (min^{-1})*	*Std. Dev.*	*Resistance*	*Std. Dev.*
0.5	$1.2x10^{-1}$	$2.2x10^{-2}$	$-6.4x10^{-2}$	$1.9x10^{-2}$	1198	513
0.7	$7.2x10^{-2}$	$1.7x10^{-2}$	$-5.0x10^{-2}$	$2.5x10^{-2}$	1800	679
1.0	$5.2x10^{-2}$	$2.6x10^{-3}$	$-3.8x10^{-2}$	$6.7x10^{-3}$	2194	210
1.2	$3.0x10^{-2}$	$6.0x10^{-3}$	$-1.2x10^{-2}$	$1.5x10^{-4}$	972	272
1.4	$7.9x10^{-3}$	$3.4x10^{-3}$	$-7.0x10^{-3}$	$3.8x10^{-3}$	1224	109

reflected this trend, reactivity, resilience, and resistance values were also calculated. These values are listed in Table 4.4. Reactivity associated with the 0.5 M NaCl perturbation was two orders of magnitude greater than the reactivity associated with the 1.4 M NaCl. Resilience also decreased with increased perturbation magnitude. Resilience for the 0.5 M NaCl was roughly one order of magnitude larger than resilience for the 1.4 M NaCl stress. Recall that a comparatively larger negative value for resilience and a comparatively larger positive value for reactivity denote greater stability associated with these aspects. Thus, the aggregate response envelope was more stable in terms of reactivity and resilience for the 0.5 M NaCl stress than the other stresses. Recall that for resistance a comparatively larger value denotes lesser stability. Thus, stability in terms of resistance, decreased for the 0.5, 0.7, and 1.0 M NaCl stresses but not for the 1.2 and 1.4 M NaCl stresses. For these last two stresses, the resistance suddenly increased and was similar in magnitude to the resistance displayed by the 0.5 M NaCl stress.

The increased resistance stepping from 1.0 M to 1.2 M NaCl corresponded to a 55% net reduction in the aggregate response at the time of maximum displacement indicating a decrease in contribution of genes and

possibly of a shift in the make-up of significantly expressed genes. Forty nine genes exhibited at least a 2-fold decrease in expression at the time of maximum displacement from 1.0 M to 1.2 M NaCl. These genes accounted for roughly 83% of the reduction in aggregate response. Within this percentage, genes making up the MAPK signaling pathway showed the largest reduction, ~45%, followed by the glycolysis pathway, ~44%, cell cycle pathway, ~8%, glycerolipid metabolism pathway, ~3% (Figure 4.5).

Of interest is the number of genes in each pathway and the pathway's contribution in the observed reduction in aggregate response. The MAPK signaling pathway included 11 of the 49 genes. However, the cell cycle pathway, which included 25 of the 49 genes, contributed least among the pathways. Even more dramatic are the 6 genes of the glycolysis pathway, which accounted for the second highest contribution in aggregate response reduction indicating the significant role of these genes. One of which was identified as ALD3 (YMR169C) encoding an aldehyde dehydrogenase known to respond to osmotic shock and the general stress response mechanism of *S. cerevisiae* (Norbeck and Blomberg, 2000). Another gene within this pathway was identified as PDC6 (YGR087C), an isozyme of pyruvate decarboxylase (Eberhart et al., 1999). From the MAPK signaling pathway was identified CTT1 (YGR088W), which exhibited the second largest reduction in resposne and is known to encode a catalase that also plays a role in the general stress response (Wiesser et al., 1991).

The gross reduction in aggregate response at the point of maximum displacement was offset by an increase in relative expression. Twenty five genes

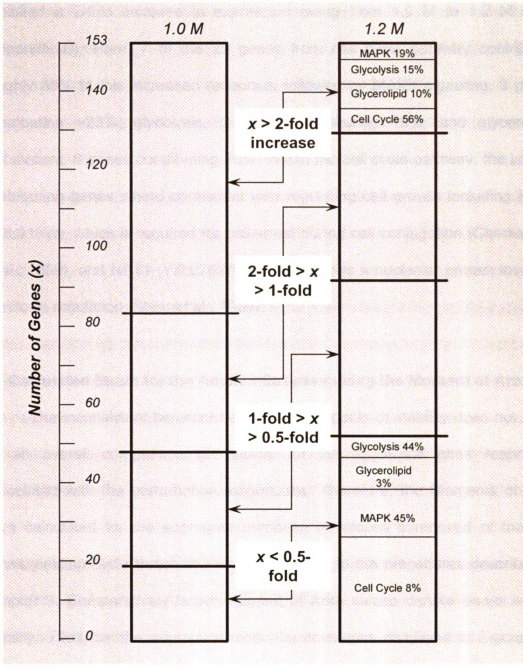

Figure 4.5. Comparison of fold changes in relative gene expression stepping from 1.0 M to 1.2 M NaCl. The largest fraction of genes at 1.0 M NaCl showed at least a 2-fold increase in expression. Whereas, the largest fraction of genes at 1.2 M NaCl showed at least a 2-fold decrease in expression. Percentages associated with the pathways indicate percent contribution in terms of response for the given fraction of genes.

exhibited a 2-fold increase in expression going from 1.0 M to 1.2 M NaCl (Appendix D). Here, 7 of the 25 genes from cell cycle pathway contributed roughly 56% to this increased response, followed by MAPK signaling, 3 genes contributing ~23%, glycolysis, 5 genes contributing 15%, and glycerolipid metabolism, 6 genes contributing 10%. Within the cell cycle pathway, the largest contributing genes where connected with regulating cell growth including FUS3 (YBL016W), which is required for cell arrest during cell conjugation (Cherkasova et al., 1999), and NET1 (YJL076W), which encodes a nucleolar protein involved in mitosis regulation (Shou et al., 1999).

4.5 Calculated Strain for the Applied Stresses using the Moment of Area

The inconsistent behavior between the aspects of stability does not allow for an overall comparison of stability for the aggregate gene responses associated with the perturbation magnitudes. Therefore, the Moments of Area were calculated for the aggregate response envelopes composed of the 153 genes yielding cycle threshold values according to the procedures described in Chapter 3. Comparatively larger Moment of Area values denote lesser overall stability. Thus, for the aggregate response envelopes displayed in Figure 4.5, overall stability decreased with an increase in the magnitude of stress. This was not surprising for perturbation magnitudes less than 1.2 M NaCl. However, the decreased overall stability despite the increased resistances for the 1.2 and 1.4 M NaCl perturbations was surprising. Evaluation of the components making up the Moment of Area revealed the influence of the moment arm on the overall

stability, which gives additional weight to response envelopes occurring over a larger time frame.

The 153 genes were divided according to biochemical pathways, i.e., glycolysis, glycerolipid metabolism, MAPK signaling, and cell cycle pathways, and the contribution of each pathway on the strain was evaluated. For the 0.5 M NaCl perturbation, glycolysis and the MAPK signaling pathways were roughly equal in their contribution to the Moment of Area and represented the largest contribution followed by the glycerolipid and cell cycle pathways. The largest contribution to the Moments of Area for the 1.2 M NaCl and 1.4 M NaCl stresses came from the glycerolipid pathway. A transition from glycolysis as the largest contributor to the glycerolipid metabolism pathway becoming the largest contributor was observed for the 0.7 M NaCl and 1.0 M NaCl. Glycolysis contributed least to the Moment of Area associated with the 1.4 M NaCl perturbation.

4.6 Relationship Between Stress and Strain for Gene Expression Patterns

The Moment of Area calculated above for each magnitude of stress is plotted in Figure 4.6. It showed an exponential increase with increase in the magnitude of stress. This relationship was also observed on an individual gene basis (Figure 4.6a), and for each group of genes making up the four pathways (Figure 4.6b). Data in Figure 4.6c was linearized by taking the natural logarithm of the Moments of Area. The slope of the linearized data was found by linear regression. This slope is defined as the modulus of stability, which predicts the

strain associated with changes in stress. A modulus of stability of 4.14 ± 0.12 [(min^2) x M^{-1}] was calculated for the aggregate response composed of the 153 genes. This predicts that for every unit change in osmotic stress a change in 4.14 units of strain occurs. Hence, a comparatively larger modulus of stability value denotes a greater change in strain per unit change in stress. Comparison of modulus of stability values to a defined standard conveys a degree of sensitivity to stress relative to the standard. The modulus of stability is of course specific to many things including the microorganisms, the type of stress, the set of genes included in the analysis, and the manner in which it was calculated.

The modulus of stability for genes grouped according to biochemical pathway ranged from 5.31 ± 0.33 [(min^2) x M^{-1}] for the glycerolipid pathway to 3.92 ± 0.21 [(min^2) x M^{-1}] for glycolysis. This suggests that the stability associated with the transcriptional response of genes making up the glycerolipid metabolism

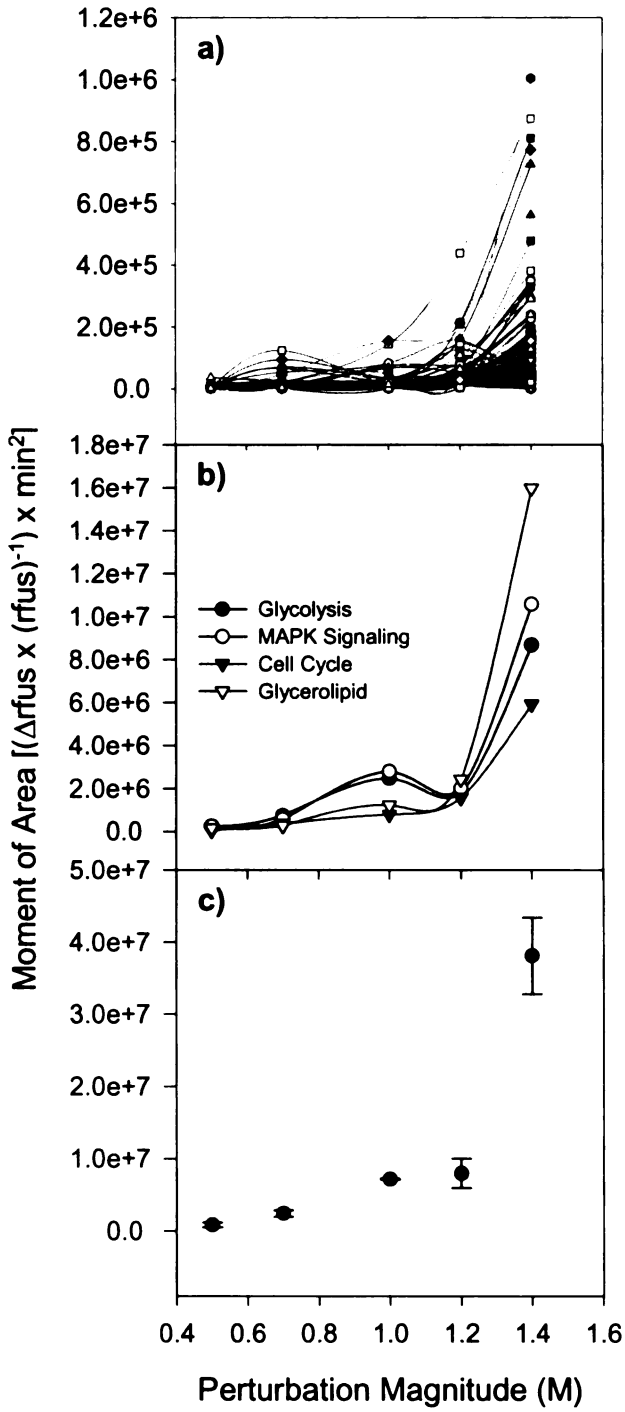

Figure 4.6. Comparison of Moments of area plotted against perturbation magnitude for **(a)** individual genes, **(b)** gene associated pathways, and **(c)** the aggregate response of selected genes described in this research.

pathway was more sensitive to changes in osmotic stress compared to the aggregate modulus of stability. Whereas, the stability associated with the transcriptional response of genes making up the glycolysis pathway was roughly equal in sensitivity to the modulus of stability of the aggregate. Of interest is the cell cycle pathway which had a modulus of stability (4.81 ± 0.51 [$(min^2) \times M^{-1}$]) greater than the aggregate set of genes and almost equal to the glycerolipid pathway, yet contributed the least to the aggregate Moment of Area for each perturbation. This suggests that while the cell cycle demonstrated greater overall stability per perturbation, it was more sensitive to osmotic shock compared to the other glycolysis and MAPK signaling pathways.

Modulus of stability values for individual genes are shown in Figure 4.7 normalized by the aggregate modulus of stability. Only those genes that exhibited a r^2 value greater than 0.90 were included in this figure (roughly 110 out of 153 genes). Genes that were less sensitive to stress relative to the aggregate modulus of stability are plotted below 1.0. Whereas, genes that were more sensitive to stress relative to the aggregate modulus of stability are plotted above 1.0. A larger number of genes exhibited greater sensitivity to osmotic shock than lesser sensitivity. Of those genes that showed lesser sensitivity most were part of the glycolysis pathway. The gene, PGM2, showed the least sensitivity to osmotic stress among the genes plotted. However, no single gene stood out among those exhibiting greater relative sensitivity to osmotic stress. A few genes per pathway falling below 1.0 are labeled in Figure 4.7. Among these, ALD3, which encodes an aldehyde dehydrogenase, proved more sensitive to

osmotic stress compared to the aggregate for both the glycolysis and glycerolipid pathways. CDC24, which encodes a GTP/GDP exchange factor showed slightly more sensitivity than STE7, which encodes a MAP kinase kinase. CDC20, a cell division control protein showed slightly more sensitivity than its neighbor, MAD2. A description of all the expressed genes indicated in Figure 4.7 can be found in Appendix C.

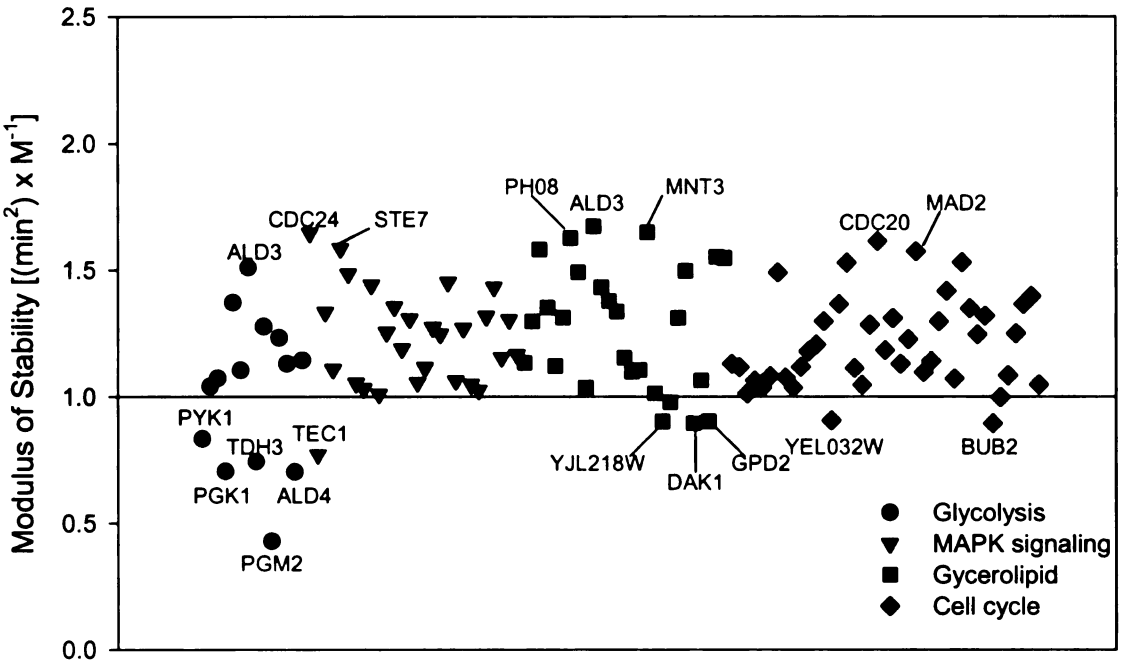

Figure 4.7. Modulus of stability values for individual genes normalized by the aggregate modulus of stability value. Genes greater than 1.0 are relatively more sensitive to stress than the aggregate modulus of stability. Genes below 1.0 are less sensitive to stress than the aggregate modulus of stability. More genes showed greater sensitivity to the stress imposed by osmotic shock than less sensitivity compared to the aggregate.

CHAPTER 5

DISCUSSION

5.1 Strain as an Aggregate Response of the Transcriptome to Stress

This research is the first to demonstrate the use of the cumulative response and relative displacement equations as a means of describing the aggregate response of a set of genes to stress. This response could be from a group of genes defined either as a cluster, or a specific functional set such as the genes making up a biochemical pathway. Describing the aggregate response of a group of genes has previously been reported in terms of a mean expression ratio. Gasch et al., (2000), demonstrated the use of the mean expression ratio to show the reciprocal responses of the ESR cluster to the osmotic and temperature perturbations. Additionally, hierarchical clustering algorithms often rely on this mean to build similarity trees from top to bottom (Eisen et al., 1998). However, using the mean response to represent the group of genes relies on the assumption that each expression profile within the group is a representative of the group rather than an individual component.

In comparison, the cumulative response and relative displacement equations used here to describe the aggregate response of the genes making up the ESR and four biochemical pathways, assume that each expression profile is an equally weighted contributor to the group of genes. Rather than measuring the central tendency, these equations describe the total response of the cluster or set of genes. Treating each expression profile as a contributor rather than a

representative has the disadvantage of being more sensitive to large erroneous expression profiles. The mean ratio of the group, in such a case, is pulled toward the erroneous profile, whereas the erroneous profile is propagated to the total response with the cumulative or displacement equations.

Because of this sensitivity, the demonstration of experimental and instrument reproducibility takes precedence. Good instrument reproducibility associated with real time PCR has been demonstrated in the past (Bustin, 2000), and is often used to validate gene expression measured using dense microarrays (Talaat et al., 2002; Yuen et al., 2002). The research approach used in this dissertation was similar to those studies in the sense that previously published microarray data was used to select a specified number of genes, then real-time PCR was used to measure their relative mRNA abundance. Because the Taq polymerase used in this research was produced in house, and PCR conditions were altered from those suggested by the equipment manufacturer, instrument reproducibility was tested and found to be excellent, as shown in Chapter 4. Rather than validating expression profiles using microarrays, experimental replication was deemed a more suitable use of resources in order to give statistical significance expression profiles used to generate the aggregate responses.

5.2 Stability Parameters: Resilience, Reactivity, and Resistance

The biological significance associated with the aggregate responses to the environmental stresses was demonstrated by comparison of their stability

parameters for eight different types of environmental perturbations (Figure 4.1). Because the magnitudes associated with these perturbations varied, it is entirely possible that under different perturbation magnitudes, stability in terms of the stability parameters and Moments of Area could be different than those estimated in Table 1.1. Normalization by the perturbation magnitudes introduces non-equivalent units, i.e., molar concentration and temperature, reducing the number of possible comparisons among the perturbations. However, if the perturbations could be represented in terms of equivalents of stress, then a more accurate picture of stability among the different perturbations for the ESR could be formulated using the analysis and developed tools demonstrated in the Results Chapter. This opens the possibility of addressing questions related to controlling gene expression by interchanging different perturbations, i.e., can the same aggregate response be generated or observed by very different types of perturbations.

Because the method of calculating strain presented here has never been used in genomics, no direct comparison to similar studies can be made. Indirect comparisons are possible, however, between the results obtained from this study and published research describing the expression of individual genes such as GPD1 to stress. Rep et al. (1999), observed that the time of occurrence of maximum relative expression of GPD1 increased, with an increase in NaCl concentration. It was also accompanied by a decrease in the rate of induction and rate of feedback, terms describing reactivity and resilience in this dissertation. In contrast, Rep at al., (1999), did not report any estimation of the

kinetics associated with these rates. A lag time following the onset of the perturbation and the start of induction with an increase in the magnitude of osmotic shock is also known to occur (Rep et al., 1999, Wuytswinkel et al., 2000). The reported lag time prior to induction or repression was not directly observed in this research, perhaps due to the schedule of temporal sampling.

The decrease in reactivity and resilience with an increase in the magnitude of perturbation observed for a single gene as well as for a larger set of genes suggests that it is a general property of relative gene expression to osmotic shock. In the case of GPD1, explanations for decreased stability in terms of reactivity and resilience could stem from direct inhibition of the transcriptional factors; Hot1p and Msn2/Msn4p, or indirect inhibition by an impeded high-osmolarity glycerol signaling pathway (HOG pathway). Evidence of the latter was presented by Wuytswinkel et al., (2000). Hog1p with a green fluorescent fusion protein (GFP) was observed to accumulate in the cytoplasm with increased osmolarity indicating hampered translocation across the nuclear membrane. It was further demonstrated that the phosphatases responsible for inactivating Hog1p were inhibited, retarding the reintroduction of Hog1p back to the cytoplasm. The nuclear accumulation of activated Hog1p in mutants unable to dephosphorylate this protein proved lethal under osmotic shock (Wuytswinkel et al., 2000).

Hindered translocation and enzymatic activity resulting in impeded transcription of GPD1 can possibly be extrapolated to the transcription of many genes suggested by the decreased reactivity and resilience of the aggregate

response. Indeed, Wuytswinkel et al. (2000), hypothesized that this translocation and hindered phosphatase problem might extend beyond the transcriptional response of GPD1 to the general stress response. Supporting this hypothesis, Yancey et al. (1982), in an earlier review on the molecular effects of osmotic shock discussed the proportional increase in the Michaelis constant (K_m) with increased NaCl concentration on enzyme activity.

The three main stability parameters (resistance, resilience, and reactivity) can be helpful in identifying the effect of such stresses on the general stress response. These parameters may also be useful in identifying candidate genes that are most relevant to nullify the effect of a given stress.

5.3 Moment of Area as a Measure of Strain

Relative gene expression is often compared by measuring the fold-change in their expression (Causton et al., 2001; Rep et al., 2000; Yuen et al., 2002). This is an appropriate comparison for relative gene expression at a single time point. In this case, such an approach is less appropriate for time course experiments because the relative expression of genes is dynamic. Hence, a fold change comparison describes only one aspect of the response, in this case resistance, and does not adequately describe the total response. This was demonstrated by comparing the observed resistances of the aggregate responses in Figure 4.5 and Table 1.1 and their respective Moments of Area with increased osmolarity. An increase in resistance (i.e., decrease in fold-change) occurred between the 1.0 M NaCl and 1.4 M NaCl perturbations, yet the

Moments of Area increased because they encompassed the total response. The advantage of measuring the overall response using the Moment of Area also has a disadvantage. The disadvantage in using the Moment of Area can be the cumbersome calculation (at least at present) and the effect of residual error associated with regression techniques can make precise calculations and comparisons of overall stability difficult.

The amount of residual error was critical for deciding whether the relative cumulative or displacement equations were better suited for calculating the aggregate response envelopes and is directly related to the regression model selected to describe the envelope. Because residual error results from the difference in the predicted value and the measured value, it can be minimized by the choice of an appropriate regression model. In the case of the linear regression model used to describe the aggregate response envelopes calculated from the displacement equation, hypothesis testing for lack of fit, i.e., whether the linear model was appropriate, was not possible with the single experimental observations made for response of the ESR. It is possible that different regression models, even non-linear ones, would have more adequately described the aggregate response envelopes. This would have been especially true if the aggregate response envelopes exhibited neighborhood stability, which was not observed for the ESR or genes associated with the different biochemical pathways.

Intuitively, it seems that increased activity of the glycolysis, glycerolipid, MAPK signaling, and cell cycle pathways would be required with osmotic stress,

and any stress for that matter, in order to meet additional cell maintenance and energy needs. In terms of gene expression, this would be implied by an increase in transcriptional response of associated genes. Hirayma et al. (1995), observed an increased transcriptional response of genes making up glycolysis to increased osmotic shock, supporting this observation. In terms of stability, the decreased overall stability corresponding to progressively larger Moments of Area seems logical considering the taxation on resources and energy requirements to repair cell wall damage and produce glycerol. However, the observed change in contribution of individual genes as well as genes associated pathways to the overall stability is not easy to explain without a better understanding of their transcriptional regulation.

It is interesting to note that on a pathway level, the observed shift from glycolysis as the largest contributor to the Moment of Area to the glycerolipid pathway indirectly correlates to the three categories of osmotic shock experienced by *S. cerevisiae* W303. Mild osmotic shock in terms of NaCl concentration, occurs from 0.4 M to 0.7 M, while hyper-osmotic shock occurs from 0.7 M to 1.2 M. Finally, severe osmotic shock occurs at NaCl concentrations greater than 1.2 M with an upper limit of possibly 1.7 M. Growth of *S. cerevisiae* W303 has been observed even at 1.7 M NaCl concentration (Hohmann and Mager, 1997; Wuytswinkel et al., 2000).

5.4 Relationship between Stress and Strain: The Modulus of Stability

The exponential relationship between the Moment of Area and perturbation magnitude observed for the aggregate response of genes supports the division of mild, hyper, and severe osmotic shocks. Up to 0.7 M NaCl the amount of stress per unit strain (slope) remains close to zero. Between 0.7 M NaCl and 1.2 M NaCl the slope transitions from close to zero to rapidly increasing. At concentrations greater than 1.2 M NaCl the slope approaches infinity. The range of NaCl concentrations for which the observed change in slope transitions from zero to approaching infinity represents a domain of strain that can be defined by upper and lower boundaries.

An upper boundary must exist for which an additional increase in perturbation results in complete cellular failure either from complete dehydration or cell wall failure. Blomberg, (2000), estimated this concentration at roughly 2.0 M NaCl for most *S. cerevisiae* strains using plating techniques. Although no gene expression studies for concentrations greater than 1.4 M NaCl have occurred, it is logical that the aggregate gene response would occur over a longer time period and its maximum displacement would become increasingly smaller with magnitudes of stress greater than 1.4 M NaCl. Theoretically, the aggregate response becomes non-existent right before 2.0 M NaCl, which roughly coincides with the concentration at which cellular failure happens. These theories are based on the observed trend in aggregate response envelopes described in Figure 4.5. Figure 5.1 theoretically depicts this upper boundary as a vertical line

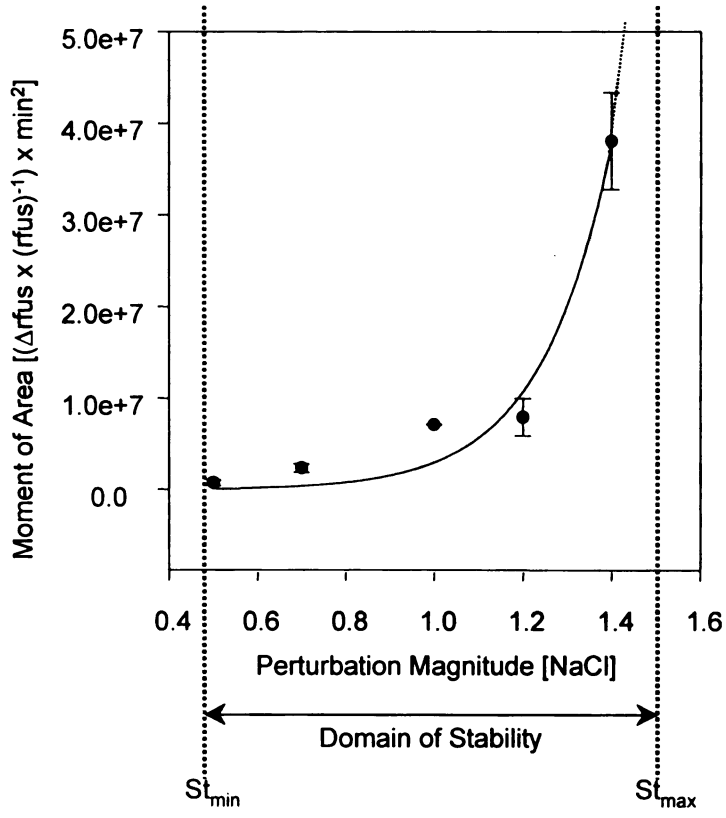

Figure 5.1. Plot of moment of area versus perturbation magnitude in terms of NaCl molar concentration. The domain of stability is depicted between the lower bound of strain, St_{min}, and upper bound of strain, St_{max}.

for the strain of the aggregate set of 153 genes, which is labeled as St_{max} (maximum strain)

Two scenarios are possible for defining this upper boundary. In the first scenario, a collapse in the Moment of Area, from "transcriptional arrest", at a concentration greater than 1.4 M NaCl would result in an undefined modulus of stability. The concentration at which this occurs would represent the upper boundary of stability for the gene response of *S. cerevisiae* to osmotic shock. It seems logical that "transcriptional arrest" would proceed the threshold of cell

death. However, whether this "transcriptional arrest" is an instantaneous occurrence or gradual is uncertain on a single cell level. It seems more likely that on a population level, transcriptional arrest would not be instantaneous reflecting the different fitnesses of cells within the population as evidenced by the dramatic reduction in, yet remaining cell viability following severe osmotic shock.

In the second scenario, the upper boundary would be represented as an asymptote in which the amount of strain occurring at this magnitude of osmotic shock is undefined. The amount of strain would infinitely increase as this asymptote is approached. A weakness of this scenario is the apparent disconnect between the strain becoming undefined and the need for increasing transcriptional response as the asymptote is approached. This weakness seen from a kinetic viewpoint suggests that the rates of transcription and translocation responsible for the required transcriptional response would soon be unable to meet the demands for maintaining cell viability.

A lower boundary of strain must also exist for which physiological conditions are altered just enough so that a transcriptional response is required altering, at least temporarily, transcriptional equilibrium. This lower boundary represents the transition from perfect stability to either asymptotic or neighborhood stability. The lower boundary is theoretically depicted in Figure 5.1, and labeled as St_{min} (minimum strain). The experimental determination of the concentration of NaCl causing this initial transcriptional response is currently dependent on the capability of equipment to resolve small changes in mRNA abundance.

CHAPTER 6

CONCLUSIONS AND FUTURE RESEARCH

6.1 Conclusions

The objectives of this study were to develop mathematical tools to quantify strain at a transcriptome level and to demonstrate these tools in a model microorganism. Developing new tools to increase our understanding of gene expression in response to environmental perturbation plays an important role in genomics. It also indirectly plays an important role in areas that are utilizing the information learned through genomics. For example, the ability not only to quantitatively monitor, but also quantitatively describe the response of environmentally important genes to stress has applications in the area of biotechnology, biological process engineering, drug development, pathogen disinfection, and remediation. Because many of the approaches used to fulfill these objectives are new, or have been adapted from other fields, it was necessary to demonstrate them in a well characterized model microorganism such as *S. cerevisiae*. With this organism in mind, combined with the defined objectives, the major findings are summarized below:

1. Gene expression response to environmental stresses can be described by established definitions of stability.
2. Overall stability calculated from the Moments of Area of individual response envelopes of genes are additive and statistically equate with

the overall stability of the aggregate response made up of the individual genes.

3. The defined environmental stress response of *S. cerevisiae* exhibited the most stability to hypo-osmotic shock, and the least stability to DTT exposure for the magnitudes of the applied perturbations.
4. Stability in terms of the parameters reactivity, resilience, and resistance decreased with increasing osmotic shock. Overall stability also decreased with increasing osmotic shock.
5. Genes respnsible for glycolysis in *S. cerevisiae* contributed most to the Moment of Area at a perturbation magnitude of 0.5 M NaCl. However, a transition was observed as perturbation magnitude increased. At 1.4 M NaCl, genes belonging to the glycerolipid pathway contributed most to the Moment of Area.
6. An exponential relationship between overall stability and osmotic shock was observed on an individual gene response, pathway response, and aggregate response for *S. cerevisiae.*
7. The modulus of stability was defined as the slope of the relationship between stability and perturbation magnitude, which describes the sensitivity of gene response to stress.
8. Genes belonging to the glycerolipid pathway for *S. cerevisiae* were most sensitive to stress from osmotic shock. Whereas, genes belonging to the glycolysis pathway were least sensitive to stress from osmotic shock.

The research presented in this dissertation combines concepts from different fields including ecology, engineering, and mathematics and apply it to genomics. An aim has been to preserve these concepts in their original form. However, because of the compexity of the genome some concepts have been slightly adapted to better fit our current understanding. For example, the displacement equation was origanally intended to describe the resilience of nutrient cycles not an aggregate response of a set of expressed genes. Also, the stability parameters, particularly resilience, vary throughout the literature in definition and calculation method. Indeed, one can become easily and rapidly confused from reading theoretical papers on ecological stability. For the stability parameters, the most adaptable definition was applied to observed gene expression. Application of concepts from other disciplines to genomics is not new, however. Clustering algorithms so often used to sort related genes originated from systematic biology. It is expected that the presented concepts, will serve as a basis for expanding our understanding of the biological significance of gene expression to environmental perturbation.

6.2 Suggested Future Research

Future research directions utilizing the concepts presented in this dissertation are potentially many. A few are briefly described below.

Two types of perturbations were defined as press and pulse and the stability of gene response associated with press perturbations investigated. However, the stability of gene response associated with pulse perturbations was

not investigated. This lack of investigation stemmed from a deficiency of published gene expression data to this type of perturbation. However, such perturbations are common. For example, the stability of genes important to waste water treatment could be investigated in response to a simulated shock loading of the reactor. Further, modeling the relationship between the pulse perturbation and aggregate gene response could be adapted and tested from previously proposed ecological models. An exactly similar experiment conducted with a pathogen (e.g., *Escherichia coli* O157:H7) and one of the many possible disinfectants will be of great value in providing information and modeling the response of the same organism in the water distribution system to varying concentrations of the disinfectant. Similar arguments can be advanced for remediation.

Intuitively, gene response associated with the establishment of a new equilibrium expression should occur, especially with constitutively expressed genes. Quantification of this neighborhood stability was proposed but not demonstrated in this research. This is because the aggregate responses that were observed demonstrated or tended to return to pre-perturbed expression despite the environmental change imposed by the press perturbation. However, it has been observed with GPD1 and other related genes responding to osmotic shock that following a return to pre-perturbed expression, expression slightly increases possibly due to adaptation. The overall stability of genes previously exposed to osmotic shock compared to the stability of genes never having been exposed to osmotic shock could be investigated.

Finally, the application of high throughput genomic technologies to mass microorganism identification is currently being evaluated. The Moment of Area and modulus of stability concepts could easily be applied at the organism scale to describe the stability of microbial communities. If the challenges of validating organism specificity can be overcome then the stability of individual organisms within the community as well as the stability of the aggregate community could be evaluated to different environmental perturbations. Until specificity issues are resolved the stability of a community "fingerprint" could be observed to different perturbations.

APPENDIX A

COMMON METRICS USED FOR MEASURING SIMILARITY (NEARNESS), DISSIMILARITY, AND CONFIDENCE

Metrics and associated equations for defining similarity, dissimilarity, or confidence between expressed genes are displayed in the following table. Where *G* represent the primary data of a gene for condition or time point *i* over a series of *m* conditions and time points, with *X* and *Y* representing gene expression vectors for comparison. For time series, in which expression profiles are compared by rows in the expression matrix, the selected metric is normalized by the number of time points in the series.

Table A.1. Additional metrics used for relating similarity and dissimilarity of relative gene expression data obtained from microarray studies.

Metric	*Equation*	*Category*	*Source*
Euclidean	$D(X,Y)=\sqrt{\sum_{i=1}^{m}(X_i-Y_i)^2}$	Distance	Legrendre and Legendre, 1998; Wen *et al.*, 1998
Manhattan	$D(X,Y)=\sum_{i=1}^{m}\lvert X_i-Y_i\rvert$	Distance	Pielou, 1984; Xia and Xie, 2001
Chord	$D(X,Y)=\sum_{i=1}^{m}\left(\sqrt{X_i}-\sqrt{Y_i}\right)^2$	Distance	Pielou, 1984; Xia and Xie, 2001
Pearson or Modified Pearson (See footnote a)	$S(X,Y)=\sum_{i=1}^{m}\left(\frac{X_i-\overline{X}}{\Phi_X}\right)\left(\frac{Y_i-\overline{Y}}{\Phi_Y}\right)$ $\Phi_G=\sqrt{\sum_{i=1}^{m}\frac{(G_i-\overline{G})^2}{m}}$	Nearness	Eisen *et al.*, 1998
Spearman	Similar to Pearson's with the exception that objects are first ranked according to their measured values.	Nearness	Xia and Xie, 2001
Jackknife (See footnote b)	$J_{X,Y}=\min\left(S^{(1)}_{X,Y}\ldots S^{(2)}_{X,Y},\ldots S^{(m)}_{X,Y},\ldots S_{X,Y}\right)$	Nearness	Heyer *et al.*, 1999
Percent Remoteness	$C(X,Y)=100-100\left(\frac{\sum_{i=1}^{m}\min(X_i,Y_i)}{\sum_{i=1}^{m}\max(X_i,Y_i)}\right)$	Confidence	Pielou, 1984

[a]Eisen et al., (1998) uses G_{offset} instead of the mean of G over m observations. G_{offset} represents a reference or standard value in which to compare with G_i.
[b]The Jackknife correlation is really just the minimum value from a correlation set obtained by deleting the ith observation and calculating a new value, $S^{(1)}_{x,y}$.

APPENDIX B

MATHCAD PROGRAM USED FOR CALCULATING THE MOMENT OF AREA FOR LARGE DATA SETS

The following program was used to calculate the Moment of Area for large data sets. The program can be divided into four areas: 1) linear regression analysis, 2) statistical analysis, 3) Moment of Area calculation, and 4) generation of and output table containing results. A detailed explanation of commands used for the program can be found in the software manuals.

MOA :=

A1 ← READPRN("C:/test/test.txt")

"Regression Analysis: finding coefficients"

for $a1 \in 1..(\text{cols}(A1) - 1)$

$b1_{a1} \leftarrow \text{linfit}(A1^{\langle 0 \rangle}, A1^{\langle a1 \rangle}, A)$

$c1_{a1} \leftarrow (b1_{a1,0})_{0,0}$

$d1_{a1} \leftarrow (b1_{a1,0})_{1,0}$

$e1_{a1} \leftarrow (b1_{a1,0})_{2,0}$

$f1_{a1} \leftarrow (b1_{a1,0})_{3,0}$

$g1_{a1} \leftarrow (b1_{a1,0})_{4,0}$

C1 ← augment(c1, d1)

D1 ← augment(e1, f1)

E1 ← augment(C1, D1)

F1 ← augment(E1, g1)

$G1 \leftarrow (F1)^T$

WRITEPRN("C:/test/coefficients.txt") := G1

"Regression Analysis: multiple regression statistics"

H1 ← submatrix(G1, 0, (rows(G1) − 1), 1, (cols(G1) − 1))

for $h1 \in 0..(\text{cols}(H1) - 1)$

for $i1 \in 0..(\text{rows}(A1) - 1)$

$j1_{h1} \leftarrow H1_{0,h1} + H1_{1,h1} \cdot A1_{i1,0} + H1_{2,h1} \cdot A1_{i1,0}^{\ 2} + H1_{3,h1} \cdot A1_{i1,0}^{\ 3} + H1_{4,h1} \cdot A1_{i1,0}^{\ 4}$

$k1_{i1,h1} \leftarrow j1_{h1}$

I1 ← submatrix(A1, 0, (rows(A1) − 1), 1, (cols(A1) − 1))

"Regression Analysis: multiple regression statistics:SS(Residual)"

$\overrightarrow{J1 \leftarrow (k1 - I1)^2}$

for $l1 \in 0..(\text{cols}(J1) - 1)$

$m1_{l1} \leftarrow \sum J1^{\langle l1 \rangle}$

$n1 \leftarrow m1^T$

"Regression Analysis:multiple regression statistics: residual standard deviation)"

$\overrightarrow{o11 \leftarrow \sqrt{\frac{n1}{(i1 + 1) - 4}}}$

MathCad Program (continued)

"Regression Analysis:multiple regression statistics: SS(Total)"

for $111 \in 0..(\text{cols}(J1)-1)$

$$m11_{111} \leftarrow \frac{\sum I1^{\langle 111 \rangle}}{i1+1}$$

$$n11 \leftarrow m11^{T}$$

for $112 \in 0..(\text{cols}(n11)-1)$

for $113 \in 0..(\text{rows}(I1)-1)$

$$m12_{112} \leftarrow \left(I1_{113,112} - n11_{0,112}\right)^2$$

$$n12_{113,112} \leftarrow m12_{112}$$

$$n13_{112} \leftarrow \sum n12^{\langle 112 \rangle}$$

$$m13 \leftarrow n13^{T}$$

"Regression Analysis:multiple regression statistics: SS(Regression)"

$$n14 \leftarrow \overrightarrow{(m13 - n1)}$$

"Regression Analysis:multiple regression statistics: coefficient of determination R^2"

$$n15 \leftarrow \overrightarrow{\left(\frac{m13 - n1}{m13}\right)}$$

"Regression Analysis:multiple regression statistics: F test statistic calculated"

$$n16 \leftarrow \overrightarrow{\left[\frac{\frac{n14}{4}}{\frac{n1}{(i1+1)-4}}\right]}$$

"Center of Mass"

for $o1 \in 0..(\text{rows}(A1)-1)$

$$p1 \leftarrow A1_{o1,0}$$

for $q1 \in 0..(\text{cols}(G1)-1)$

$$r1_{q1} \leftarrow \int_0^{p1} G1_{0,q1} + G1_{1,q1}\cdot x + G1_{2,q1}\cdot x^2 + G1_{3,q1}\cdot x^3 + G1_{4,q1}\cdot x^4 \, dx$$

$$s1_{q1} \leftarrow \int_0^{p1} x\cdot G1_{0,q1} + G1_{1,q1}\cdot x + G1_{2,q1}\cdot x^2 + G1_{3,q1}\cdot x^3 + G1_{4,q1}\cdot x^4 \, dx$$

$$t1_{q1} \leftarrow \int_0^{p1} G1_{0,q1} + G1_{1,q1}\cdot x + G1_{2,q1}\cdot x^2 + G1_{3,q1}\cdot x^3 + G1_{4,q1}\cdot x^4 \, dx$$

MathCad Program (continued)

"'Center of Mass: x-coordinate"

$u1_{q1} \leftarrow \frac{s1_{q1}}{r1_{q1}}$

"Center of Mass: y-coordinate"

$v1_{q1} \leftarrow 0.5 \cdot \frac{t1_{q1}}{r1_{q1}}$

r11 ← submatrix(r1, 1, (rows(r1) − 1), 0, cols(r1) − 1)

u11 ← submatrix(u1, 1, (rows(u1) − 1), 0, cols(u1) − 1)

v11 ← submatrix(v1, 1, (rows(v1) − 1), 0, cols(v1) − 1)

"End"

"Moment Arm"

for $w1 \in 0..(\text{rows}(u1) - 1)$

$x1_{w1} \leftarrow \sqrt{(u1_{w1,0} - 0)^2}$

x11 ← submatrix(x1, 1, (rows(x1) − 1), 0, cols(x1) − 1)

"End"

"Moment of Area"

for $y1 \in 0..(\text{rows}(x1) - 1)$

$z1_{y1} \leftarrow r1_{y1} \cdot x1_{y1}$

z11 ← submatrix(z1, 1, (rows(z1) − 1), 0, cols(z1) − 1)

"End"

"Organizing Results"

K1 ← augment(m1, r11)

L1 ← augment(K1, u11)

M1 ← augment(L1, v11)

N1 ← augment(M1, x11)

O1 ← augment(N1, z11)

P1 ← ("Residual Standard Deviation" "Area" "x" "y" "Moment Arm" "Moment of Area")

Q1 ← stack(P1, O1)

o11

APPENDIX C

DESCRIPTION OF GENES GROUPED ACCORDING TO BIOCHEMICAL PATHWAY USED TO STUDY RELATIVE GENE EXPRESSION TO OSMOTIC SHOCK.

Genes studied in response to osmotic shock are listed in the following Tables. Description of genes and related information was obtained from the *Saccharomyces cerevisiae* genome database (SGD) located at www.yeastgenome.org. Diagrams for each pathway can be obtained from the Kyoto encyclopedia of genomes and genes (KEGG) located at www.genome.ad.jp/kegg/kegg2.html.

Table C.1. Descriptions of genes, and their accompanying primer sets, encoding proteins for the Glycolysis/Gluconeogenesis pathway.

Systematic Name	*Standard Name*	*Gene Description*	*Enzyme Commision*	*Forward Sequence (5'-3')*	*Tm*	*Reverse Sequence (5'-3')*	*Tm*
YAL038W	CDC19, PYK1	pyruvate kinase 1	EC:2.7.1.40	TACTCCAAAGCCAACCTCCACC	62	CCGGAAGTGGACAAGACAATGA	60
YAL054C	ACS1	acetyl-CoA synthetase	EC:6.2.1.1	TTCACAGCTGGAGACATTGGC	60	TACGCAGGAGTCCCTTCAAAGA	60
YBR019C	GAL10	UDP-glucose 4-epimerase / aldose 1-epimerase	EC:5.1.3.2 5.1.3.3	GCCAGCATTGTTGACCTGAAA	58	TCGTGGCGCCTATATAAGCACT	60
YBR196C	PGI1	glucose-6-phosphate isomerase	EC:5.3.1.9	TCACTAACGCTAACACTGCCAA	58	ACTTCGGTTTCGTTAGTGGACA	58
YCL040W	GLK1	glucokinase	EC:2.7.1.2	TGTACCAGGAGCAATTAAGCGC	60	GGACGTGTAGCAATGCGATAGG	62
YCR012W	PGK1	phosphoglycerate kinase	EC:2.7.2.3	CCAGAAAGGTCGATGGTCAAAA	58	CCGAAGGCATCGTTGATGTAA	58
YDL021W	GPM2	phosphoglycerate mutase	EC:5.4.2.1	ACCCCAGATTGGTTACACCTCA	60	CGTAAACAACCTGCAGTTGTGG	60
YDL168W	SFA1	formaldehyde dehydrogenase (glutathione) / long-chain alcohol dehydrogenase	EC:1.2.1.1 1.1.1.1	TTCTCAATTTGGTGCCACGG	57	AATCCAGACCCCCATCAGTCAT	60
YDR050C	TPI1	triosephosphate isomerase (TIM)	EC:5.3.1.1	AGGCCGGTAAGACTTTGGATGT	60	ATGGCCCAGACTGGTTCGTAA	60
YDR380W	ARO10	pyruvate decarboxylase	EC:4.1.1.1	GTTTCCAACCGATCAGCAACA	58	TTGAAGTCACCAGGAACACCG	60
YER073W	ALD2, ALD5	aldehyde dehydrogenase	EC:1.2.1.3	AGGAGGAAGTGTTTGGTCCCAT	60	ACCTGCGGCTAACCCATATTG	60
YER178W	PDA1	pyruvate dehydrogenase E1 component, alpha subunit	EC:1.2.4.1	AGAATGGAGATGGCTTGTGACG	60	TGGCATTCTCGATACCGACAG	60
YFR053C	HXK1, HKA	hexokinase I	EC:2.7.1.1	ACATTTGGTCTTGCCAAGAACC	58	CCCAAGTAGTAACCGGAGGTCA	62
YGR087C	PDC6	pyruvate decarboxylase isozyme 3	EC:4.1.1.1	CGTAGGTGAATTATCCGCCTTG	60	TGCTTAGCCTGAGCGGAGATAG	62
YGR192C	TDH3, GPD3	glyceraldehyde 3-phosphate dehydrogenase	EC:1.2.1.12	TTCCGCTGGTATCCAATTGTCT	58	CGTGTTCAACCAAGTCGACAAC	60
YGR240C	PFK1	6-phosphofructokinase	EC:2.7.1.11	GTTGGCCTTGATGGCTGGTAT	60	GCACACTTCCTTCAATTCGTCC	60
YKL060C	FBA1	fructose-bisphosphate aldolase	EC:4.1.2.13	CAGCAAGTCCCCAATCATTTTG	58	CACCCTTGATGGAAGCATTTTG	58
YLR377C	FBP1	fructose-1,6-bisphophatase	EC:3.1.3.11	GCTTGCTATGCCATGTACGGAT	60	GAGTCAAGATGAATTCGCCCAA	58
YMR105C	PGM2	phosphoglucomutase	EC:5.4.2.2	GGCCAGCATGGTCTTCTGTCTA	62	CCTGGATTATGTGAGGCGGTTA	60
YMR169C	ALD3	aldehyde dehydrogenase (NAD(P)+)	EC:1.2.1.5	GGCTGCTTTTGATAACGTTTGG	58	GCAAGTGTGTCTTGCTCCTCCT	62
YMR170C				AGCAGCCGGTAACACGGTTAT	60	CACCAGGTGGAAAACCTGCTT	60
YMR303C	FKS3	1,3-beta-glucan synthase	EC:2.4.1.34	GCGAACGGTACTGTTGTCTTGG	62	CACGTAAGAGCCGACAATGGAG	62
YMR323W	YMR323W	enolase related protein 1	EC:4.2.1.11	TATTGGCGAAGGAGCAATATGG	58	AATCATGTCCAAGGCGTCCTC	60
YNL071W	LAT1, PDA2	pyruvate dehydrogenase E2 component (dihydrolipoamide acetyltransferase)	EC:2.3.1.12	GGCTTGTCGCAAATCTCTAACG	60	GGATATGCAAATGGTCCCACC	60
YOL056W	GPM3	phosphoglycerate mutase	EC:5.4.2.1	GAGGAGCTGCAGGACACAAGAT	62	GGCTTTCTTTGTCCTTGCCAT	58
YOR374W	ALD4	aldehyde dehydrogenase	EC:1.2.1.3	ATCTTTGGCCCTGTTGTCACTG	60	GTGTGAATACCAGCAGCCAACC	62
YPL017C	YPL017C	dihydrolipoamide dehydrogenase	EC:1.8.1.4	TTTGCGGCAGTTTACGATGTC	58	TCAACACATGCTGTGAGCAGG	60

Table C.2. Descriptions of genes, and their accompanying primer sets, encoding proteins for the Glycerolipid Metabolism pathway.

Systematic Name	*Standard Name*	*Gene Description*	*Enzyme Commission*	*Forward Sequence (5'-3')*	*Tm*	*Reverse Sequence (5'-3')*	*Tm*
YBR029C	CDS1, CDG1	phosphatidate cytidylyltransferase	EC:2.7.7.41	GATATCTTCGCCTACCTGTGCG	62	CAGTGAAAAACCAGGCACCAA	58
YBR184W	YBR184W	alpha-galactosidase	EC:3.2.1.22	AAGGTCCGAGGAAGAGGTTGTT	60	CTCAAAACCTTCTGCGCTATCC	60
YCL004W	PGS1	phosphatidylserine synthase	EC:2.7.8.8	TTGATGGCCTTCGAGGAACA	57	TCGGCAATCCACTCTCTCTGAC	62
YDL022W	GPD1, OSG1, DAR1, HOR1	glycerol-3-phosphate dehydrogenase (NAD+)	EC:1.1.1.8	AATGGTTGGAAACATGTGGCTC	58	ATCATGTCCGGCAGGTTCTTC	60
YDL052C	SLC1	1-acyl-sn-glycerol-3-phosphate acyltransferase	EC:2.3.1.51	CGCCTCTATCCTTTGCACGTT	60	ACGTCAAGGCCAAGCATCAA	57
YDL142C	CRD1, PGS1	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase	EC:2.7.8.5	GCTGTGAAAAAACCGAGGAGG	60	CCGCTCATTATTGTGGTCGTG	60
YDL168W	SFA1	formaldehyde dehydrogenase (glutathione) / long-chain alcohol dehydrogenase	EC:1.2.1.1 1.1.1.1	TTCTCAATTTGGTGCCACGG	57	AATCCAGACCCCCATCAGTCAT	60
YDL243C	AAD4	putative aryl-alcohol dehydrogenase	EC:1.1.1.-	TTCTGCGGCAAACTACTACGC	60	GGAATGATATCGCGCTCAAAGT	58
YDR058C	TGL2	triacylglycerol lipase	EC:3.1.1.3	AAGGACATGGATCATCTGGACG	60	CAGGATGTCAACCTTCTCTCCG	62
YDR147W	EKI1	ethanolamine kinase	EC:2.7.1.82	AAATTGTGACAAGCCAACGAGG	58	CATTCGTGCATGTTATTGGCAG	58
YDR481C	PHO8	repressible alkaline phosphatase	EC:3.1.3.1	GCGCAGTCGACGTCAACATAT	60	GCCAACTTCCGTATTTTCGTGA	58
YER001W	MNN1	alpha-1,3-mannosyltransferase	EC:2.4.1.-	CCGAAATGATGCGCTTATACG	58	GGAGAGCCGAAGGCTTCAATA	60
YER026C	CHO1, PSS1, PSS	phosphatidylserine synthase	EC:2.7.8.8	AATTCCCACACACGGACACAG	60	AGGCCCTTCTGCTTAATGTGC	60
YER062C	GPP2	glycerol-3-phosphatase	EC:5.3.1.1	GGTTGGAGAACGTTTGATGCC	60	TTTTCACCGTACTTGACCGGAA	59
YER073W	ALD2, ALD5	aldehyde dehydrogenase	EC:1.2.1.3	AGGAGGAAGTGTTTGGTCCCAT	60	ACCTGCGGCTAACCCATATTG	60
YFL056C	AAD6	putative aryl-alcohol dehydrogenase (AAD)	EC:1.1.1.-	GGAGCGAGCTTTTGAGTTGCT	60	CCGATCCAAGCTTCTGACTGTT	60
YGL038C	OCH1	alpha-1,6-mannosyltransferase	EC:2.4.1.-	CGTCAAATTCAATGCCCTGG	57	GCCCCACATCTGGTGAAAAA	57
YGL257C	MNT2	alpha-1,3-mannosyltransferase	EC:2.4.1.-	GCGAGGTTTACCATGCGTCTA	60	TTCTCCCAAAAAGATAGCCCTG	58
YGR007W	MUQ1	choline-phosphate cytidylyltransferase	EC:2.7.7.14	ACTGCTGCCGGAAAATTCAGT	58	CCCTTTTTTTTGGTTCCTCGC	57
YGR170W	PSD2	phosphatidylserine decarboxylase	EC:4.1.1.65	GCCACAAGATTATCACCGGTTT	58	GAACGGCCATTGGATTTACAGT	58
YGR202C	PCT1, CCT1, CCT	choline-phosphate cytidylyltransferase	EC:2.7.7.15	TGCGCAAGTATCGTCCTAAAGG	60	GTGGCCAAGATGGAAAAGATCA	58
YHL032C	GUT1	glycerol kinase	EC:2.7.1.30	TGGTCCAATGGCTACGTGATAA	58	GACGAAAACTACGCCACCAGA	60
YHR123W	EPT1	ethanolaminephosphotransferase (ETHPT)	EC:2.7.8.1	TCCCTAGCTGTTTTCGGTTTGG	60	TTGCGTGATATTGTTGGCCG	57
YIL014W	MNT3	alpha-1,3-mannosyltransferases	EC:2.4.1.-	AATGAAGTGGAACCCTCTGCAC	60	GATGCTTCTTCTGAGGTTTCGG	60
YIL155C	GUT2	glycerol-3-phosphate dehydrogenase	EC:1.1.99.5	AATCAAATCGCCGTCATGGA	55	TCCAACAAACCCATATCCTTCG	58

Table C.2. (Continued)

Systematic Name	Standard Name	Gene Description	Enzyme Commission	Forward Sequence (5'-3')	Tm	Reverse Sequence (5'-3')	Tm
YJL218W	YJL218W	putative acetyltransferase in HXT11-HXT8 intergenic region	EC:2.3.1.-	ATACGGCAGGTCATCCAATTGA	58	TGACACTCCACCTCCAATCCA	60
YKR009C	FOX2	peroxisomal hydratase-dehydrogenase-epimerase (HDE) containing 2-enoyl-CoA hydratase / D-3-hydroxyacyl CoA dehydrogenase	EC:4.2.1.- 1.1.1.-	AAAAGCCGGAGGTATAGCTGTG	60	AGTACATCAACCCTGCCGAATT	58
YKR031C	SPO14	phospholipase D	EC:3.1.4.4	TGCATTGCCCAAATCCTTTG	55	CCTCGACGAGGATCATGTTCAT	60
YLR133W	CKI1	choline kinase	EC:2.7.1.32	TAATCAACCCGCCAAAGCAG	57	GATTCGCTAAATCATATGCGGC	58
YML070W	DAK1	dihydroxyacetone kinase	EC:2.7.1.29	CGCTGGTCCCTGAAGAAAAAA	58	ATGAAACCGGCGTGTGTAGGT	60
YMR303C	FKS3	1,3-beta-glucan synthase	EC:2.4.1.34	GCGAACGGTACTGTTGTCTTGG	62	CACGTAAGAGCCGACAATGGAG	62
YNL130C	CPT1	diacylglycerol cholinephosphotransferase	EC:2.7.8.2	GTATGCATTTTGCACAGGAGCA	58	TGCTTGGAGTTGCAGATTTGG	58
YOL059W	GPD2, GPD3	glycerol-3-phosphate dehydrogenase (NAD+)	EC:1.1.1.8	ACCATCGCCAAAGTCATTGC	57	TCAGATTTTCGTCGCCGATC	57
YOR120W	GCY1, GCY	galactose-induced protein of aldo/keto reductase family	EC:1.1.1.-	AAGTCGGTCAAGCCATCAAGG	60	TGATCCAGCGCTACTTCAGGTT	60
YOR374W	ALD4	aldehyde dehydrogenase	EC:1.2.1.3	ATCTTTGGCCCTGTTGTCACTG	60	GTGTGAATACCAGCAGCCAACC	62
YPL206C		similar to glycerophosphoryl diester phosphodiesterases	EC:3.1.4.46	ATCGCCTCGCAATTTGTGAA	55	AATTGGCTTGTCCACGAGGAG	60
YPR113W	PIS1	CDP-diacylglycerol--Inositol 3-phosphatidyltransferase	EC:2.7.8.11	ACGGAACCATGGCAAGAAAGT	58	AAAGGAAACACATCAAGCCAGC	58

Table C.3. Descriptions of genes, and their accompanying primer sets, encoding proteins for the Mitogen Activated Kinase (MAPK) Signaling pathway.

Systematic Name	*Standard Name*	*Gene Description*	*Enzyme Commission*	*Forward Sequence (5'-3')*	*Tm*	*Reverse Sequence (5'-3')*	*Tm*
YAL041W	CDC24	GTP/GDP exchange factor for CDC42P		TTCCATCTCCATACCCATGCA	58	ACACCTCCAGTCTTTTCCGGA	60
YBL016W	FUS3	mitogen-activated protein kinase (MAP kinase)	EC:2.7.1.-	CAATTCAGAGCCCACAGGTCAG	62	GGCCCTTGAGTATTTGGCAGA	60
YBL105C	PKC1	protein kinase C-like 1 (PKC 1)	EC:2.7.1.-	GAGATATTCGAACAGCATGGCC	60	TGGAACCCAGTATACCGAAGGA	60
YBR083W	TEC1	transcriptional regulator of Ty1 expression		TGAAATCTCCATCAGCAAAGGC	58	ACGTTTCGCTACTTTGTCCGG	60
YBR200W	BEM1, SRO1	bud emergence mediator		TTTGTTTATTTGCGCCCACC	55	TGATGCTAACAAACCCAACAGG	58
YCL027W	FUS1	nuclear fusion protein		CTTCGCAAGCTAGTTCCTCGAC	62	CTGAGCCGCCACATTAGAAAA	58
YDL159W	STE7	ser/thr/tyr protein kinase of MAP kinase kinase family	EC:2.7.1.-	CAGCAACCGCAAAATGAACCT	58	GCATCTATAATGCATGGCGTGG	60
YDL235C	YPD1	two-component phosphorelay intermediate	EC:2.7.3.-	TGGATGACGATGATTCCGAT	55	TTTTTCACCGTCCAGCTGTC	57
YDR103W	STE5	pheromone signal transduction pathway protein		GGTACCTTGTCGAGAACTCCCA	62	CCTGGCTAACTTTTCCGTCCA	60
YER111C	SWI4, ART1	transcription factor		GTAGCATCCAATGGCTCCTCCA	62	ATTTCCTGCGCTGAACGGTGT	60
YER118C	SHO1	involved in the HOG1 high-osmolarity signal transduction pathway		GTGATAATGCTTCCCCAACCAA	58	GCCTCTTCGACGAGCTCTATGA	62
YFL026W	STE2	pheromone alpha factor receptor		TCCTCGCATACAGTTTGGAACC	60	CGTGGCCCACATTGATGATAAT	58
YGR032W	GSC2, GLS2, FKS2	1,3-beta-glucan synthase	EC:2.4.1.34	AACTTGGTTCAACCAAGGAACG	58	ACCAGGTCTTTAAGGATGGCG	60
YGR040W	KSS1	serine/threonine protein kinase of the MEKK family	EC:2.7.1.-	CTGGCAATGTACCATGACCCA	60	TCCTCTTCCGGACGCATTATC	60
YGR088W	CTT1	catalase	EC:1.11.1.6	TTCCACGTCTTGTCGGATACTG	60	TGAGTGAACAGCTTTGCCTGAT	58
YHL007C	STE20	serine/threonine-protein kinase	EC:2.7.1.-	CGATTACCAAGGACCACTGGAA	60	TTCTTCGCAGGATCCATGGA	57
YHR005C	GPA1, SCG1, CDC70, DAC1	guanine nucleotide-binding protein alpha-1 subunit		GAGTACCGGACGAGCAAAAGCT	62	CCTCATTTTGCGTCACCGTTT	58
YHR030C	SLT2	serine/threonine MAP kinase	EC:2.7.1.-	CGCCACAACAAGAATCATTTGG	58	TGGCAGGTCTCATCTCCATCAT	60
YHR084W	STE12	transcription factor		TACCCAAACGGAATGGTTCCA	58	AAATCGTCCGCGCCATAAAA	55
YIL147C	SLN1	osomolarity two-component system protein	EC:2.7.3.-	ACTAGCAAAAA TCATCACCGG	56	GACGTACAATTGGTTGCACTGC	60
YJL095W	BCK1	serine/threonine protein kinase BCK1/SLK1/SSP31	EC:2.7.-.-	CACCACCAATTCCTGAGGACAC	62	TAGCGGTTGGCCTTTTCTCTG	60
YJL128C	PBS2	tyrosine protein kinase of the MAP kinase kinase family	EC:2.7.1.-	TTGAAAAGGTCTGCCAGCGT	57	AGGCGGCAATGGCTTATTAAC	58
YJL157C	FAR1	cyclin-dependent kinase inhibitor (CKI)		CATTTGGCACCTTCATCGGA	57	TTCATCGCGGAAGACGTTTG	57
YKL178C	STE3	pheromone A factor receptor		GACGACGATTTCCTCACGAGAT	60	TTGGTAACGGCACATGATATGC	58

Table C.3. (Continued)

Systematic Name	Standard Name	Gene Description	Enzyme Commission	Forward Sequence (5'-3')	Tm	Reverse Sequence (5'-3')	Tm
YKR095W	MLP1	myosin-like protein		AACGAAGATGCGTCATACAGCA	58	TGCGAGATACGAGTTTGCAAGT	58
YLR006C	SSK1	osomolarity two-component system protein		ATAACGCAACCCATGTCGGA	57	AGCGAGTTTGAGTAATGCGGC	
YLR113W	HOG1	mitogen-activated protein kinase	EC:2.7.1.-	CGCTGGTTGTATTTTTGCCG	57	CGGAGATCCCAACAAGTCAGTG	62
YLR229C	CDC42	belongs to the Rho subfamily of Ras-like proteins		TTCCCCACCCTCTTTTGAAAAC	58	GATCAATCTGCGTACCGACGA	60
YLR342W	FKS1, CND1	1,3-beta-glucan synthase	EC:2.4.1.34	ATTCAGAGGCTGAGCGTCGTA	60	GTCAATACTGTGAACGTTGGCA	58
YLR362W	STE11	Ser/Thr protein kinase involved in the mating signalling pathway	EC:2.7.1.-	AACGTTGAGTGCATTGAGCCAG	60	TGTCGATCCGGAGTTTCCCTT	60
YML004C	GLO1	lactoylglutathione lyase	EC:4.4.1.5	GAGCCTGATGTTTTTAGCGCAC	60	GGTTCCTCATTCCCGTTGTTG	60
YMR037C	MSN2	stress responsive regulatory protein		TTATCACGCCGAAGATTTAGCG	58	TTTCGGATGGTTCCAAAGGTC	58
YMR043W	MCM1	putative transcriptional activator of alpha-specific genes		CACGCAAATAGCTTAGGCCATC	60	TTGATTAGGATTGGCACCGC	57
YNL053W	MSG5	dual-specificity protein tyrosine phosphatase	EC:3.1.3.48	CGTCAAGGACTTATCCCGATTG	60	GACGCCGATCTTGATACTCCAC	62
YNL098C	RAS2, GLC5, CTN5	GTP-binding protein		GCAAATGAACGCTCCTTTCTTG	58	TTGCCGCCTTCGTCTCTAACT	60
YNL271C	BNI1	BNI1 protein (synthetic lethal 39)		TTAGCTAGGAATTACGCCCCGT	60	CAGCCCTCTGTAAATCGTTTGG	60
YNR031C	SSK2	serine/threonine protein kinase	EC:2.7.-.-	TCGGATGGCTGAAGTTCTTAGC	60	GCTTACCATATGCATCGCGAA	58
YOR008C	SLG1	SLG1 protein precursor		CTTCGGATACTTTGGCAACGAG	60	AAAATTGTGGAACCGCCCTC	57
YOR212W	STE4	guanine nucleotide-binding protein beta subunit		GATCTCCGTCCGCTGTACAAAG	62	CCATTGTCACTTCCTGCAACAA	58
YOR231W	MKK1	serine/threonine protein kinase	EC:2.7.1.-	GAATCTGCGAAATGCAACCC	57	GAACTCCCATTCGACGTCAAGT	60
YPL049C	DIG1	MAP kinase-associated protein		AGCGTAGGATCAAGTGCGAATG	60	TGCATCCCGTTCAAAGTGGTA	58
YPL089C	RLM1	serum response factor-like protein		AGATGCACCCGTAAATGGAGCT	60	ATTATATGCGGTGGCAGCAGC	60
YPR165W	RHO1	GTP-binding protein of the rho subfamily of ras-like proteins		CGACCCACAAACCATTGAACAA	58	AGTATCCGGTTGCGCCAATCT	60

Table C.4. Descriptions of genes, and their accompanying primer sets, encoding proteins for the Cell Cycle pathway.

Systematic Name	*Standard Name*	*Gene Description*	*Enzyme Commission*	*Forward Sequence (5'-3')*	*Tm*	*Reverse Sequence (5'-3')*	*Tm*
YAL024C	LTE1	GDP/GTP exchange factor		AAGCGGAACGCTTCAAGGTA	57	AGCGCTGCGATGTAACACAGAC	62
YAL040C	CLN3, DAF1, WHI1	cyclin, G1/S-specific		CAAAGAGCGCTACGGTTTCATC	60	GAAGGCAAATGGTGGAGAAATG	58
YAR019C	CDC15	cell division control protein 15	EC:2.7.1.-	TAAGCCTTTGACGGACTTGCA	58	GTTTATCAACAGTTGTGCGCGA	58
YBL016W	FUS3	mitogen-activated protein kinase (MAP kinase)	EC:2.7.1.-	CAATTCAGAGCCCACAGGTCAG	62	GGCCCTTGAGTATTTGGCAGA	60
YBR093C	PHO5	repressible acid phosphatase precursor (P60)	EC:3.1.3.2	TGCCAACTCGGACGATGTTT	57	TGGTTTTCGACCATGTAACCGT	58
YBR112C	CYC8	glucose repression mediator protein		TCC CCAACCTTTAATCCAGCA	58	TGGATACGGCGGCATCTACTAA	60
YBR133C	HSL7	histone synthetic lethality		CCTGCAAAATCGCATCGTCTT	58	ATTCCACAGTTCCCAAGTGGC	60
YBR135W	CKS1	cyclin-dependent kinase regulatory subunit		ACCATCACTATCACGCCTTCCA	60	CGTCTGAGTACCGCGGAGAATA	62
YBR136W	MEC1, SAD3, ESR1	required for DNA damage induced checkpoint responses in G1, S/M, intra S, and G2/M in mitosis, similar to PI3-kinases		AATTAGGCAGCTGGTGCAAAG	58	CAGTCAAGGCCTGTTCGTAAAG	60
YBR160W	CDC28, SRM5	cyclin-dependent protein kinase catalytic subunit	EC:2.7.1.-	CACACCGTATTCTGCATCGTGA	60	AACACCAAAAGCACGCGCTA	57
YBR274W	CHK1	probable serine/threonine-protein kinase	EC:2.7.1.-	GTTGGAGTTGATTCCGACGTG	60	GGCTTGATGTCTCTGTGAGCAA	60
YCR069W	SCC3	peptidyl-prolyl cis-trans isomerase precursor	EC:5.2.1.8	CTGAAAGACTCGCAATGGCCT	60	CGACACTTTTGCCATCCAATTC	58
YCR084C	TUP1	glucose repression regulatory protein		GCAGAATTGGCAAAAGATGTGG	58	GCCAAAAATTTCCCATCTGGAG	58
YDL003W	MCD1	mitotic chromosome determinant		CATGTCTGAGGAGCCTACGGAT	62	TGTGATTCTGTCATGGCATCG	58
YDL017W	CDC7	cell division control protein 7	EC:2.7.1.-	GAACCTCAATTACCTGACCCCA	60	AAGCACCAATAATGATCGGACC	58
YDL028C	MPS1	serine/threonine/tyrosine protein kinase (regulatory cell proliferation kinase 1)	EC:2.7.1.-	GGCACCAGAAGCACTAGTTGCT	62	AACCGCATGACCACATATCAGA	58
YDR052C	DBF4, DNA52	regulatory subunit for Cdc7p protein kinase		GCCGGATGATTTGTGCACTT	57	TTGCCAAAAGAGGTTGCCAC	57
YDR113C	PDS1	cell cycle regulator		TTGGAAGACGAAGATGACACCC	60	TGATTCCGAGTGATGCACTCAG	60
YDR146C	SWI5	transcriptional factor		GAGTGATAAGCCCCACGATGAA	60	TCCCCGAGGTTTCTATCGAAG	60
YDR217C	RAD9	DNA repair checkpoint protein		GGACTCCTGACCTTGACCGAAT	62	TGTCTTCCAGATCGGAGCTTTG	60
YDR507C	GIN4	serine/threonine-protein kinase	EC:2.7.1.-	TGGTAAAGTCCAGCTTGCTCGT	60	GAAGTACCGCTGACATTACCGG	62
YEL032W	MCM3	minichromosome maintenance protein 3		AGCATCGTCCACCTCATTGAAT	58	GGTTCGACTAAAATGCCCGAC	60
YER111C	SWI4, ART1	transcription factor		GTAGCATCCAATGGCTCCTCCA	62	ATTTCCTGCGCTGAACGGTGT	60
YER173W	RAD24	cell cycle checkpoint protein		CACCCCGAACGAGCATAAAGT	60	ACCGTGCGCCTTTTAAGAACTC	60
YFL008W	SMC1, CHL10	chromosome segregation protein		TACCAGCCTAGTCCCTTCTTCG	62	TCTGGATTACGGTGCCTTCTTA	58
YFL009W	CDC4	cell division control protein		AGGTGCATCTCTGCCAGTATGG	62	GGCGATGCTGAATTTGTTGC	57

Table C.4. (Continued)

Systematic Name	Standard Name	Gene Description	Enzyme Commission	Forward Sequence (5'-3')	Tm	Reverse Sequence (5'-3')	Tm
YFL029C	CAK1, CIV1	cdk-activating protein kinase	EC:2.7.1.-	CAGCACCGTATCCCCAAAGTT	60	TCCGTCACCTTGCTATCCATG	60
YFR034C	PHO4	phosphate system positive regulatory protein PHO4		ACCGTAAACACGAAACGGCA	57	TCGTTCCCATTCTCACTGCTG	60
YGL003C	CDH1	substrate-specific activator of APC-dependent proteolysis		TCTGCGGCCTCTTTGTTACAA	58	TGTCTAAATTGCTTCCCCGGT	58
YGL086W	MAD1	spindle assembly checkpoint component		CTTCAGTGCATTGAGTGGACGA	60	TCCAAGAAAGGAGAGGACCCA	60
YGL116W	CDC20	cell division control protein		CGAAGTCTGCGGTTTGAGCTA	60	TGAGGCAAGGACGTTCTGGTA	60
YGL201C	MCM6	involved in replication		TGGACGTCATTCTTCGAGGTG	60	GCCCCAATTGTGTAACATCAGG	60
YGR092W	DBF2	serine/threonine protein kinase	EC:2.7.1.-	CCACGACAAGGTCTGAATGGTT	60	ACGAAAATCACCGCCTGGTAC	60
YGR098C	ESP1	involved in regulation of spindle pole body duplication		CCAAATTAGAACCGCGACGTAA	58	GCCTCCGGAAATGAGAGATGA	60
YGR108W	CLB1, SCB1	cyclin, G2/M-specific		TTATGTGCCTCAGCGGCAAT	57	TTTGCCTTTGTGTAACCACCAC	58
YGR109C	CLB6	cyclin, B-type		ATCACTTGCCTGTTCATTGCCT	58	AATTCCTTCGACAGTGGCAGC	60
YGR113W	DAM1	essential mitotic spindle pole protein		GCGCTGATTTACCCATCGAAA	58	CCATTGACCACTCTTGCAGCA	60
YGR188C	BUB1	checkpoint serine/threonine-protein kinase	EC:2.7.1.-	TTGTCCCCAAATCTACGCGA	57	CGTTGTCGTTTGCACTTGTACC	60
YGR233C	PHO81	cyclin-dependent kinase inhibitor		TAATAATCATGACGCCGCCA	55	GAAAAGCGGTTGACACAAGTG	58
YIL046W	MET30	negatively regulates sulfur amino acids biosynthesis		GGAAATGTCCCAACTGTGGCT	60	TCTGGATATCTGCGTTGCTGC	60
YJL030W	MAD2	spindle-assembly checkpoint protein		TTCACAGGTGGCTTCTTGGTG	60	TGTTGCACATTGAAGGACCATC	58
YJL074C	SMC3	chromosome segregation protein		GGAGATTAGTGCAGCGCACAA	60	GTTTGGACTCCAATTCGGCA	57
YJL076W	NET1	establishes silent chromatin		AGATTGAAGCCCCCTCTCCAT	60	AACTGACGGAGGAGACGACATG	62
YJL157C	FAR1	cyclin-dependent kinase inhibitor (CKI)		CATTTGGCACCTTCATCGGA	57	TTCATCGCGGAAGACGTTTG	57
YJL187C	SWE1	mitosis inhibitor protein kinase	EC:2.7.1.-	GGATGAGGAAAGCGTGGATTC	60	CTGTTCGTTTTTTCCGCAGATC	58
YJL210W	PEX2	peroxisomal assembly protein - peroxin		CGCACCTTGTATGTCGTCACTG	62	GCTTCAGCCACATCATCTGGTT	60
YJR053W	BFA1	unknown		AGAAGCAGACCCTTTTGCAGG	60	AACTTGGCTTTTGGAACGCA	55
YJR090C	GRR1	required for glucose repression and for glucose and cation transport		GGCAACTTCGGAAAGAAATGC	58	TGTTGGTTGGCAAAAGCGTT	55
YKL101W	HSL1	probable serine/threonine-protein kinase	EC:2.7.1.-	TGGCCAAGAAGGCAGCTAAA	57	TGGTAGCGTCGTTCACAGATTG	60
YLL004W	ORC3	origin recognition complex protein, subunit 3		GCTGCCTTCGCTTGATAAAGAA	58	TCGCCTCACGATAAAGCTTGA	58
YLR079W	SIC1	substrate and inhibitor of the cyclin-dependent protein kinase Cdc28		CTCGCCGTTTCCAAAATCTTC	58	TTTACACGACCCAATGGTTCCT	58
YLR103C	CDC45	required for minichromosome maintenance and initiation of chromosomal DNA replication		ATTGTTGCCCATGGTGCTTG	57	TGTGTATCGTGTCTAGTCCGCG	62

Table C.4. (Continued)

Systematic Name	Standard Name	Gene Description	Enzyme Commission	Forward Sequence (5'-3')	Tm	Reverse Sequence (5'-3')	Tm
YLR210W	CLB4	cyclin, G2/M-specific		TCAAGTAGAGCTTACATGGCCG	60	TCGTCAGGTATAGCGTTTCTGG	60
YLR288C	MEC3	G2-specific checkpoint protein		TGAGGATAGCTCGATCGCAGAT	60	CAGCCTTTTCCACCATAACGG	60
YML064C	TEM1	GTP-binding protein of the ras superfamily		TGTTTGACCTGACACGTCCAGA	60	CTTTGTGCCCACCAAAATAGGA	58
YML065W	ORC1	origin recognition complex protein, subunit 1		GGTTGCCAACGTAATGGATGG	60	TCCCCAGTCGGCTCACAAATA	60
YMR001C	CDC5	protein kinase which functions at the G(sub)2/M boundary	EC:2.7.1.-	ATTGAGCGAAAAACCACGAGAG	58	TGCCAAACGTAGATACTCGGCT	60
YMR036C	MIH1	M-phase inducer phosphatase	EC:3.1.3.48	ATCAAGTATCCTGGCAGTCCCC	62	CACCCTCCTGCTTGAATTGCT	60
YMR055C	BUB2	cell cycle arrest protein		ATGTTACGTGTGGACGGTGCT	60	GGTAAATGGTTGTGGAAGGTGG	60
YMR199W	TPS3	alpha,alpha-trehalose-phosphate synthase (UDP-forming)	EC:2.4.1.15	GCTGTCAGGTTCTACGATCGCT	62	CAAGTTTTGGCCGCTAACCA	57
YNL289W	PCL1, HCS26	cyclin, G1/S-specific		AAACGTGTCGATAGCATCCGA	58	GCGACTGTGTTGTTCGCTATGT	60
YOR026W	BUB3, OR26.16	cell cycle arrest protein		TGATAGGTAGTCCCAGCTTCCA	60	TGACGCGGCAATGAGTTTAT	55
YOR368W	RAD17	DNA damage checkpoint control protein		GCCTCAACGGTGCACTTAGAAC	62	TGACAGCCCATCAGCATCAA	57
YPL031C	PHO85	cyclin-dependent protein kinase	EC:2.7.1.-	TTTCGGTATTCCGGTCAACAC	58	TCAATGGATGTGGAGTACGTCC	60
YPL153C	RAD53	ser/thr/tyr protein kinase	EC:2.7.1.-	TTTTGTTGCTGCTCATGGTGC	58	TGGCTGATGCCCATAGAGTGA	60
YPL194W	DDC1	DNA damage checkpoint protein		AGAGGCCGGACTGAAAGACAA	60	GTTTCGAGTTCCAGCCACAGAG	62
YPL256C	CLN2	cyclin, G1/S-specific		TTGGTCAGCAGCAGTCAAACAG	60	GGACCCGTGATTTGTTCGTAGA	60
YPR111W	DBF20	cell cycle protein kinase related to DBF2P	EC:2.7.1.-	ATGTTTTGTGCCGTGAACGC	57	AATCGGTCAGCTTGATATGCCC	60

APPENDIX D

GENES EXHIBITING A SIGNIFICANT INCREASE OR DECREASE IN EXPRESSION STEPPING FROM 1.0 M TO 1.2 M NACL

Genes exhibiting a decrease in relative expression of at least 0.5 and those genes exhibiting an increase in at least 2-fold expression stepping from 1.0 M to 1.2 M NaCl are displayed in the table below. Relative expression values were obtained from the ratio of the 1.2 M NaCl data and the 1.0 M NaCl data. Negative values indicate a decrease in expression stepping from 1.0 M to 1.2 M NaCl. Positive values indicate an increase in expression stepping from 1.0 M to 1.2 M NaCl. Highlighted genes are discussed in the results section.

Table D.1. Genes exhibiting a significant increases or decreases in relative expression at the time of maximum displacement. Comparison between the 1.0 M and 1.2 M NaCl perturbation magnitudes.

Systematic Name	*Standard Name*	*Gene Description*	*Pathway*	**Change in Response from 1.0 M to 1.2 M NaCl*
***Genes exhibiting a decrease in relative expression of at least 0.5 stepping from 1.0 M to 1.2 M NaCl.*				
YBR136W	MEC1	Required for DNA damage induced checkpoint responses	Cell Cycle	-0.77
YER173W	RAD24	Cell cycle checkpoint protein	Cell Cycle	-5.04
YGL003C	CDH1	Substrate-specific activator of APC-dependent proteolysis	Cell Cycle	-5.05
YGL086W	MAD1	Spindle assembly checkpoint component	Cell Cycle	-9.77
YGR092W	DBF2	Serine/threonine protein kinase	Cell Cycle	-0.77
YGR113W	DAM1	Essential mitotic spindle pole protein	Cell Cycle	-2.13
YGR188C	BUB1	Checkpoint serine/threonine-protein kinase	Cell Cycle	-2.64
YIL046W	MET30	Negatively regulates sulfur amino acids biosynthesis	Cell Cycle	-1.35
YJL030W	MAD2	Spindle-assembly checkpoint protein	Cell Cycle	-3.94
YJL074C	SMC3	Chromosome segregation protein	Cell Cycle	-0.99
YJL210W	PEX2	Peroxisomal assembly protein - peroxin	Cell Cycle	-1.15
YJR053W	BFA1	Unknown	Cell Cycle	-1.58
YJR090C	GRR1	Required for glucose repression and for glucose and cation transport	Cell Cycle	-2.38
YLR079W	SIC1	Substrate and inhibitor of the cyclin-dependent protein kinase Cdc28	Cell Cycle	-0.18
YLR210W	CLB4	Cyclin, G2/M-specific	Cell Cycle	-1.27
YLR288C	MEC3	G2-specific checkpoint protein	Cell Cycle	-3.83
YML064C	TEM1	GTP-binding protein of the ras superfamily	Cell Cycle	-5.81
YML065W	ORC1	Origin recognition complex protein, subunit 1	Cell Cycle	-2.07
YMR001C	CDC5	Protein kinase which functions at the G(sub)2/M boundary	Cell Cycle	-1.27
YMR036C	MIH1	M-phase inducer phosphatase	Cell Cycle	-3.72
YNL289W	PCL1	Cyclin, G1/S-specific	Cell Cycle	-1.77
YOR368W	RAD17	DNA damage checkpoint control protein	Cell Cycle	-5.94
YPL153C	RAD53	ser/thr/tyr protein kinase	Cell Cycle	-2.90
YPL194W	DDC1	DNA damage checkpoint protein	Cell Cycle	-37.48
YPR111W	DBF20	Cell cycle protein kinase related to DBF2P	Cell Cycle	-9.62
YAL054C	ACS1	Acetyl-CoA synthetase	Glycolysis	-5.31
YGR087C	PDC6	Pyruvate decarboxylase isozyme 3	Glycolysis	-99.94
YLR377C	FBP1	Fructose-1,6-bisphophatase	Glycolysis	-8.79
YMR169C	ALD3	Aldehyde dehydrogenase (NAD(P)+)	Glycolysis	-479.33
YMR303C	FKS3	1,3-beta-glucan synthase	Glycolysis	-1.89
YPL017C	YPL017C	Dihydrolipoamide dehydrogenase	Glycolysis	-38.20
YDR147W	EKI1	Ethanolamine kinase	Glycerolipid	-5.46
YGR170W	PSD2	Phosphatidylserine decarboxylase	Glycerolipid	-2.36

Table D.1. (Continued)

Systematic Name	*Standard Name*	*Gene Description*	*Pathway*	**Change in Response from 1.0 M to 1.2 M NaCl*
YHL032C	GUT1	Glycerol kinase	Glycerolipid	-0.69
YIL155C	GUT2	Glycerol-3-phosphate dehydrogenase	Glycerolipid	-2.74
YKR031C	SPO14	Phospholipase D	Glycerolipid	-3.71
YML070W	DAK1	Dihydroxyacetone kinase	Glycerolipid	-28.28
YOR374W	ALD4	Aldehyde dehydrogenase	Glycerolipid	-5.54
YBR200W	BEM1	Bud emergence mediator	MAPK	-1.33
YDL235C	YPD1	Two-component phosphorelay intermediate	MAPK	-6.46
YGR040W	KSS1	serine/threonine protein kinase of the MEKK family	MAPK	-1.41
YGR088W	CTT1	Catalase	MAPK	-626.33
YHR005C	GPA1	Cuanine nucleotide-binding protein alpha-1 subunit	MAPK	-2.13
YJL128C	PBS2	Tyrosine protein kinase of the MAP kinase kinase family	MAPK	-0.53
YKR095W	MLP1	Myosin-like protein	MAPK	-1.19
YLR006C	SSK1	Osomolarity two-component system protein	MAPK	-0.73
YLR362W	STE11	Ser/Thr protein kinase involved in the mating signalling pathway	MAPK	-1.60
YMR037C	MSN2	Stress responsive regulatory protein	MAPK	-2.32
YMR043W	MCM1	Putative transcriptional activator of alpha-specific genes	MAPK	-0.76
***Genes exhibiting an increase in relative expression of at least 2-fold stepping from 1.0 M to 1.2 M NaCl.*				
YBL016W	FUS3	Mitogen-activated protein kinase (MAP kinase)	Cell Cycle	32.46
YBR093C	PHO5	Repressible acid phosphatase precursor (P60)	Cell Cycle	1.05
YBR135W	CKS1	Cyclin-dependent kinase regulatory subunit	Cell Cycle	10.33
YDR507C	GIN4	serine/threonine-protein kinase	Cell Cycle	0.43
YFR034C	PHO4	Phosphate system positive regulatory protein PHO4	Cell Cycle	2.78
YGL201C	MCM6	Involved in replication	Cell Cycle	0.83
YJL076W	NET1	Establishes silent chromatin	Cell Cycle	25.82
YBR196C	CDS1, CDG1	Phosphatidate cytidylyltransferase	Glycolysis	0.58
YDR050C	SLC1	1-acyl-sn-glycerol-3-phosphate acyltransferase	Glycolysis	2.35
YDR380W	PHO8	Repressible alkaline phosphatase	Glycolysis	12.34
YER178W	MNT2	Alpha-1,3-mannosyltransferase	Glycolysis	0.58
YGR240C	MUQ1	Choline-phosphate cytidylyltransferase	Glycolysis	1.20
YKL060C	EPT1	Ethanolaminephosphotransferase (ETHPT)	Glycolysis	2.43
YBR029C	GPP1	Glycerol biosynthesis	Glycerolipid	0.35
YDL052C	PGI1	Glucose-6-phosphate isomerase	Glycerolipid	0.30
YDR481C	TPI1	Triosephosphate isomerase (TIM)	Glycerolipid	1.44
YGL257C	ARO10	Pyruvate decarboxylase	Glycerolipid	0.54
YGR007W	PDA1	Pyruvate dehydrogenase E1 component, alpha subunit	Glycerolipid	7.87

Table D.1. (Continued)

Systematic Name	*Standard Name*	*Gene Description*	*Pathway*	**Change in Response from 1.0 M to 1.2 M NaCl*
YHR123W	PFK1	6-phosphofructokinase	Glycerolipid	1.69
YIL053W	FBA1	Fructose-bisphosphate aldolase	Glycerolipid	1.57
YER118C	SHO1	Involved in the HOG1 high-osmolarity signal transduction pathway	MAPK	0.71
YFL026W	STE2	Pheromone alpha factor receptor	MAPK	11.29
YKL178C	STE3	Pheromone A factor receptor	MAPK	1.05
YPL089C	RLM1	Serum response factor-like protein	MAPK	0.40
YPR165W	RHO1	GTP-binding protein of the rho subfamily of ras-like proteins	MAPK	11.38

*The change in response is the difference between the relative expression for the 1.0 M and 1.2 M NaCl perturbations. Negative values indicate a decrease in the response.
**Relative expression values were obtained from the ratio of the 1.2 M NaCl data and the 1.0 M NaCl data.

BIBLIOGRAPHY

Alexandre, H., V. Ansanay-Galeote, S. Dequin and B. Bondin. (2001). Global gene expression during short-term ethanol stress in *Saccharomyces cerevisiae*. FEBS Lett. 498:98-103.

Alon, U., N. Barkai, D. A. Notterman, K. Gish, S. Ybarra, D. Mack and A. J. Levin. (1999). Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc. Natl. Acad. Sci. USA 96:6745-6750.

Altman, R. B. and S. Raychaudhuri. (2001). Whole-genome expression analysis: challenges beyond clustering. Curr. Opin. Struct. Biol. 11:340-347.

Andreishcheva, E. N. and R. A. Zvyagilskaya. (1999). Adaptation of yeasts to salt stress (Review). Appl. Biochem. Microbiol. 35:217-228.

Attfield, P. V. (1997). Stress tolerance: the key to effective strains of industrial baker's yeast. Nat. Biotechnol. 15:1351-1357.

Autio, R., S. Hautaniemi, P. Kauraniemi, O. Yli-Harja, J. Astola, M. Wolf and A. Kallioniemi. (2003). CGH-Plotter: MATLAB toolbox for CGH-data analysis. Bioinformatics 19:1714-1715.

Ayala-del-Río, H. (2002). Long-term effects of phenol and phenol plus thrichloroethene application on microbial communities in aerobic sequencing batch reactors. Ph.D. dissertation. Michigan State University, East Lansing, MI.

Beliaev, A. S., D. K. Thompson, T. Khare, H. Lim, C. C. Brandt, G. Li, A. E. Murray, J. F. Heidelberg, C. S. Giometti, J. Yates 3rd, K. H. Nealson, J. M. Tiedje and J. Zhoui. (2002). Gene and protein expression profiles of *Shewanella oneidensis* during anaerobic growth with different electron acceptors. OMICS 6:39-60.

Bender, E. A., T. J. Case and M. E. Gilpin. (1984). Perturbation experiments in community ecology: theory and practice. Ecology 65:1-13.

Ben-Dor, A., R. Shamir and Z. Yakhini. (1999). Clustering gene expression patterns. J. Comput. Biol. 6:281-297.

Benitez, T., J. M. Gasent-Ramireź, F. Castrejón and A. C. Codón. (1996). Development of new strains for the food industry. Biotechnol. Prog. 12:149-163.

Blomberg, A. (1997). The osmotic hypersensitivity of yeast *Saccharomyces cerevisiae* is strain and growth media dependent: quantitative aspects of the phenomenon. Yeast 13:529-539.

Blomberg, A. (2000). Metabolic surprises in *Saccharomyces cerevisiae* during adaptation to saline conditions: questions, some answers and a model. FEMS Microbiol. Lett. 182:1-8.

Brazma, A. L. and J. Vilo. (2000). Gene expression analysis. FEBS Lett. 480:17-24.

Brown, C. S., P. C. Goodwin and P. K. Sorger. (2001). Image metrics in the statistical analysis of DNA microarray data. Proc. Natl. Acad. Sci. USA 98:2622-267.

Bustin, S. A. (2000). Absolute quantification of mRNA using real-time reverse transricption polymerase chain reaction assays. J. Molec. Endroc. 25:169-193.

Causton, H. C., B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L True, E. S. Lander and R. A. Young. (2001). Remodeling of yeast genome expressionin response to environmental change. Mol. Biol. Cell. 12:323-337.

Cherkasova V., D. M., Lyons and E. A. Elion. (1999). Fus3p and Kss1p control G1 arrest in *Saccharomyces cerevisiae* through a balance of distinct arrest and proliferative functions that operate in parallel with Far1p. Genetics 15:989-1004.

Cho, R. J., M. J. Campbell, E. A. Winzeler, L. Steinmetz, A. Conway, L. Wodicka, T. G. Wolfsberg, A. E. Gabrielian, D. Landsman, D. L. Lockhart and R. W. Davis. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. Mol. Cell. 2:65-73.

Cottingham, K. L. and S. R. Carpenter. (1994). Predictive indices of ecosystem resilience in models of north temperate lakes. Ecology 75:2127-2138.

DeAngelis, D. L. (1980). Energy flow, nutrient cycling, and ecosystem resilience. Ecology 61:764-771.

DeAngelis, D. L., S. M. Bartell and A. L. Brenkert. (1989). Effects of nutrient and food-chain length on resilience. Amer. Nat. 134:778-805.

DeAngelis, D. L. (1992). Dynamics of nutrient cycling and food webs. Chapman and Hall, New York.

DeFrancesco, L. (2003). Real-time PCR takes center stage. Anal. Chem. 75:175A-179A.

DeRisi, J. L., V. R. Iyer and P. O. Brown. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278:680-686.

Eberhardt I, H. Cederberg, H. Li, S. Konig, F. Jordan and S. Hohmann. (1999). Autoregulation of yeast pyruvate decarboxylase gene expression requires the enzyme but not its catalytic activity. Eur J Biochem 262:191-201

Eisen, M. B., P. T. Spellman, P. O. Brown and D. Botstein. (1998). Cluster analysis and display of genome-wide expression patterns. Proc. Natl. Acad. Sci. USA 95:14863-14868.

Engelke, D. R., A. Krikos, M. E. Bruck and D. Ginsburg. (1990). Purification of *Thermus aquaticus* DNA polymerase expressed in *Escherichia coli*. Anal. Biochem. 191:396-400.

Estruch, F. (2000). Stress-controlled transcription factors, stress induced genes and stress tolerance in budding yeast. FEMS Microbiol. Rev. 24:469-48.

Fedorscak, I. and L. Ehrenberg. (1966). Effects of diethyl pyrocarbonate and methyl methanesulfonate on nucleic acids and nuclease. Acta. Chem. Scand. 20:107.

Fernandes, L., C. Rodrguez-Pousada and K. Struhl. (1997). Yap, a novel family of eight bZIP proteins in *Saccharomyces cerevisiae* with distinct biological functions. Mol. Cell Biol. 17:6982-6993.

Fishbane, P. M., S. Gasiorowiz and S. T. Thornton. (1993). Physics for scientists and engineers. Prentice Hall, Englewood Cliffs, New Jersey.

Gasch, A. P., P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein and P. O. Brown. (2000). Genomic expression programs in the response of yeast cells to environmental changes. Mol. Biol. Cell. 11:4241-4257.

Gollub, J., C. A. Ball, G. Binkley, J. Demeter, D. B. Finkelstein, J. M. Hebert, T. Hernandez-Boussard, H. Jin, M. Kaloper, J. C. Matese, M. Schroeder, P. O. Brown, D. Botstein and G Sherlock. (2003). The Stanford Microarray Database: data access and quality assessment tools. NAR 31: 94-96.

Grandpre, L. and Y. Bergeron. (1997). Diversity and stability of understorey communities following disturbance in the southern boreal forest. J. Ecol. 85:777-784.

Grimm, V., E. Schmidt and C. Wissel. (1992). On the application of stability concepts in ecology. Ecol. Model. 63:143-161.

Harrison, G. W. (1979). Stability under environmental stress: resistance, resilience, persistance, and variability. Amer. Nat. 113:659-669.

Hartigan, J. A. (1975). Clustering algorithms. John Wiley and Sons, New York.

Hashsham, S. A., A. S. Fernandez, S. L. Dollhopf, F. B. Dazzo, R. F. Hickey, J. M. Tiedje and C. S. Criddle. (2000). Parallel processing of substrate correlates with greater functional stability in methanogenic bioreactor communities perturbed by glucose. Appl. Environ. Microbiol. 66:4050-4057.

Hecker, M and S. Engelmann. (2000). Proteomics, DNA arrays and the analysis of still unknown regulons and unknown proteins of *Bacillus Subtilis* and pathogenic gram-positive bacteria. J. Med. Microbiol. 290:123-134.

Hernandez, J. A. 2003. Stability properties of elementary dynamic models of membrane transport. Bulletin of mathematical biology 65:175-197.

Hernandez, A., A. Figueroso, L. A. Rivas, V. Parro and R. P. Mellado. (2000). RT-PCR as a tool for systematic transcriptional analysis of large regions of the *Bacillus subtilis* genome. Microbiology 146:823-828.

Heyer, L. J., S. Kruglyak and S. Yooseph. (1999). Exploring expression data: identification and anlysis of coexpressed genes. Genome Res. 9:1106-1115.

Hinchliffe, S. J., K. E. Isherwood, R. A. Stabler, M. B. Prentice, A. Rakin, R. A. Nichols, P. C. Oyston, J. Hinds, R. W. Titball and B. W. Wren. Application of DNA microarrays to study the evolutionary genomics of *Yersinia pestis* and *Yersinia pseudotuberculosis*. Genome Res. 9:2018-29.

Hirayama, T., T. Maeda, H. Saito and K. Shinozaki. (1995). Cloning and characterization of seven cDNAs for hyperosmolarity-responsive (HOR) genes of *Saccharomyces cerevisiae*. Mol. Gen. Genet. 249:127-138.

Hohmann, S and W. H. Mager. (1997). Yeast stress response. Chapman and Hall, New York, New York.

Holling, C. S. (1973). Resilience and stability of ecological systems. Annu. Rev. Ecol. Syst. 4:1-23.

Inchausti, P. (1995). Competition between perennial grassed in neotropical savanna: the effects of fire any hydric-nutritional stress. J. Ecol. 83:231-243.

Ives, A. R. (1995). Measuring resilience in stochastic systems. Ecological Monographs 65:217-233.

Kanehisa, M. and S. Goto. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28:27-30.

Karsai, A., S. Müller, S. Platz and M. T. Hauser. (2001). Evaluation of a home-made SYBR green I reaction mixture for real-time PCR quantification of gene expression. Biotechniques 32:790-796.

Klis F. M., M. Pieternella, K. Hellingwerf and B. Stanley. (2002). Dynamics of cell wall structure in *Saccharomyces cerevisiae*. FEMS Microbiol. Rev. 26:239-256.

Kobayashi N. and K. McEntee. (1993). Identification of cis an trans components of a novel heat shock stress regulatory pathway in *Saccharomyces cerevisiae*. Mol. Cell Biol. 13:248-256.

Kurilova, S. A., N. N. Vorobjeva, T. I. Nazarova, and S. M. Avaeva. (1993). Expression of *Saccharomyces cerevisiae* inorganic pyrophosphatase in *Escherichia coli*. FEBS Lett. 333:280-282.

Larsson K., P. Eriksson, R. Ansell and L. Alder. (1993). A gene encoding *sn*-glycerol 3-phosphate dehydrogenase (NAD^+) complements an osmosensitive mutant of *Saccharomyces cerevisiae*. Mol. Microbiol. 10:1101-1111.

Laws, R. J., T. L. Bergemann, F. Quiaoit and L. P. Zhao. (2003). SignalViewer: analyzing microarray images. Bioinformatics 19:1716-1717.

Legendre, P. and Legendre L. (1998). Numerical Ecology. 2nd Edition. Elsevier Science, The Netherlands.

Lewontin, R. C. (1969). The meaning of stability. Brookhaven Symp. Biol. 22:13-24.

Livak, K. J. and T. D. Schmittgen. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. Methods 25:402-408.

Lockhart, D. J., H. Dong, M. C. Byrne, M. T. Follettie, M. V. Gallo, M. S. Chee, M. Mittmann, C. Wang, M. Kobayashi, H. Horton and E. L. Brown. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat. Biotech. 14:1675-1680.

Long, A. D., H. J. Mangalam, B. Y. P. Chan, L. Tolleri, G. W. Hatfield and P. Baldi. (2001). Improved statistical inference from DNA microarray data using analysis of variance and a bayesian statistical framework. J. Biol. Chem. 276:19937-19944.

Longnecker, M. and R. L. Ott. (2001). An introduction to statistical methods and data analysis. Duxbury Press, Pacific Grove, CA.

Maeda, T., M. Takekawa and H. Saito. (1995). Activation of yeast PBS2 MAPKK by MAPKKKS or by binding of an SH3-containing osmo-sensor. Science 269:554-558.

Mager, W. H. and M. Siderius. (2002). Novel insights into osmotic stress response of yeast. FEMS Yeast Res. 2:251-257

Marchler G., C. Schuller, G. Adam and H. Ruis. (1993). A *Saccharomyces cerevisiae* UAS element controlled by protein kinase A activates transcription in response to a variety of stress conditions. EMBO J. 12:1997-2003.

Martinez-Pastor, M. T., G. Marchler, C. Schuller, A. Marchler-Bauer, H. Ruis and F. Estruch. (1996). The *Saccharomyces cerevisiae* zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through stress response element (S.T.R.E.). EMBO J. 15:2227-2235.

Merkin, D. R. (1997). Introduction to the theory of stability. Springer-Verlag, New York.

Mittelbach, G. G., A. M. Turner, D. J. Hall and J. E. Rettig. (1995). Perturbation and resilience: a long-term whole-lake study of predator extinction and reintroduction. Ecology 76:2347-2360.

Murray, J. D. (1993). Mathematical biology. Springer-verlag, New York, New York.

Nakajima, H. (1992). Sensitivity and stability of flow networks. Ecol. Mod. 62:123-133.

Neubert, M. G and H. Caswell. (1997). Aternatives to resilience for measuring the response of ecological systems to perturbations. Ecology 78:653-665.

Norbeck A., A. K. Påhlmann, N. Akhtar, A. Blomberg and L., Alder. (1996). Purification and characterization of two isoenzymes of DL-glycerol 3-phosphatase from *Saccharomyces cerevisiae*. Identification of the corresponding GPP1 and GPP2 genes and evidence for osmostic regulation of Gpp2p expression by the osmosensing MAP kinase signal transduction pathway. J. Biol. Chem. 271:13875-13881.

Norbeck, J. and A. Blomberg. (1997). Two dimensional electrophoretic separation of yeast proteins using a non-linear wide range (pH 3-10) immobilized pH gradient in the first dimension: reproducibility and evidence for isoelectric focusing of alkaline (pI>7) proteins. Yeast 13:1519-1534.

Norbeck, J. and A. Blomberg. (2000). The level of cAMP-dependent protein kinase A activity strongly affects osmotolerance and osmo-instigated gene expression changes in *Saccharomyces cerevisiae*. Yeast 16:121-37.

O'Neill, R. V. (1976). Ecosystem persistence and heterotrophic regulation. Ecology 57:1244-1253.

Perou, C. M., T. Sorlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lonning, A. L. Borresen-Dale, P. O. Brown and D. Botstein. (2000). Molecular portraits of human breast tumours. Nature 406:747-752.

Pimm, S. L. (1982). Food Webs. Chapman and Hall, New York, NY.

Pimm. S. L. (1984). The complexity and stability of ecosystems. Nature 307:321-326.

Planet, P. J., R. DeSalle, M. Siddall, T. Bael, I. N. Sarkar and S. E. Stanley. (2001). Systematic analysis of DNA microarray Data: ordering and interpreting patterns of gene expression. Genome Res. 11:1149-1155.

Pluthero, G. (1993). Rapid purification of high-activity Taq DNA polymerase. Nucleic Acids Res. 21:4850-4851.

Rantokokko-Javala, K. and J. Javala. (2001). Development of conventional and real-time PCR assays for detection of *Legionella* DNA in respiratory specimens. J. Clin. Microbiol. 39:2904-2910.

Rep, M., J. A. Albertyn, J. M Thelvelein, B. A. Prior, and S. Hohmann. (1999). Different signalling pathways contribute to the control of gpd1 gene expression by osmotic stress in *Saccharomyces cerevisiae*. Microbiology 145:715-727.

Rep, M., M. Krantz, J. M. Thevelein and S. Hohmann. (2000). The transcriptional response of *Saccharomyces cerevisiae* to osmotic shock. J. Biol. Chem. 275:8290-8300.

Rep, M., M. Proft, J. Remize, M. Tamás, R. Serrano, J. M. Thevelein and S. Hohmann. (2001). The *Sacchromyces cerevisiae* Sko1p transcription factor mediates HOG pathway-dependent osmotic regulation of a set of gene encoding enzymes implicated in protection from oxidative damage. Mol. Microbiol. 40:1067-1083.

Ririe, K. M., R. P. Rasmussen and C. T. Wittwer. (1997). Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. Anal. Biochem. 245:154-160.

Sambrook J., E. F. Fritsch and T. Maniatis. (1989). Molecular cloning: a laboratory manual. Cold Springs Harbor, Plainview, New York.

Schena, M., D. Shalon, R. W. Davis and P. O. Brown. (1996). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467-470.

Schmitt, M. E., T. A. Brown and B. L. Trumpower. (1990). A rapid and simple method for preparation of RNA from *Saccharomyces cerevisiae*. Nucleic Acids Res. 18:3091-3092.

Sennhauser, E. B. (1991). The concept of stability in connection with the gallery forests of the chaco region. Vegetatio. 94:1-13.

Shalon D., S. J. Smith and P. O Brown. (1996). A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res. 6:639-6445.

Sherlock, G. (2000). Analysis of large-scale gene expression data. Curr. Opin. Immunol. 12:201-205.

Shou W, J. H. Seol, A. Shevchenko, C. Baskerville, D. Moazed, Z. W. Chen, J. Jang, A. Shevchenko, H. Charbonneau and R. J. Deshaies. (1999) Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. Cell 97:233-44.

Siderius, M., E. Rots and W. H. Mager. (1997). High-osmolarity signalling in *Saccharomyces cerevisiae* is modulated in a carbon-source-dependent fashion. Microbiology UK 143:3241-3250.

Spellman, P. T., G. Sherlock, M. Q. Zhang, V. R. Iyers, K. Anders, M. B. Eisen, P. O. Brown, D. Botstein and B. Futcher. (1998). Comprehensive identification of the cell cycle-regulated genes of the yeast Saccharomyces cerevisiae by microarray hybridization. Mol. Biol. Cell. 9:3273-3297.

Spiro A., M. Lowe and D. Brown. (2000). A bead-based method for multiplexed identification and quantification of DNA sequences using flow cytometry. Appl. Environ. Microbiol. 66:4258-4265.

Sutherland J. P. (1981). The fouling community at beaufort, north carolina: a study in stability. Amer. Nat. 118:499-519.

Swami S., N. Raghavachari, U. R. Muller, Y. P. Bao and D. Feldman. (2003). Vitamin D growth inhibition of breast cancer cells: gene expression patterns assessed by cDNA microarray. Breast Cancer Res. Treat. 80:49-62.

Talaat A. M., S. T. Howard, W. Hale IV, R. Lyons, H. Garner and S. A. Johnston. (2002). Genomic DNA standards for gene expression profiling in *Mycobacterium tuberculosis*. Nucleic Acids Res. 30:104-112.

Tamayo, P., D. Slonim, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. S. Lander and T. R. Golub. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc. Natl. Acad. Sci. USA 96:2907-2912.

Treger, J. M., A. P. Schmitt, J. R. Simon and K. McEntee. (1998). Transcriptional factor mutations reveal regulatory complexities of heat shock and newly identified stress genes in *Sacchromyces cerevisiae*. J. Biol. Chem. 273:26875-26879.

Toshihide S., P. J. Higgins and D. R. Crawford. (2000). Control selection for RNA quantification. Biotechniques 29:332-337.

Venema J., and D. Tollervey. (1999). Ribosome synthesis in *Saccharomyces cerevisiae*. Annu. Rev. Genet. 33:261-311.

Viragh, K. (1989). An experimental approach to the study of community stability: Resilience and Resistence. Acta Bot. Hung. 35:99-125.

Wen, X., S. Fuhrman, G. S. Michaels, D. B. Carr, S. Smith, J. L. Barker and R. Somogyi. (1998). Large-scale temporal gene expression mapping of central nervous system development. Proc. Natl. Acad. Sci. USA 95:334-339.

Wickert, L., S. Steinkrüger, M. Abiaka, U. Bolkenius, O. Purps, C. Schnabel and A. M. Gressner. (2002). Quantification monitoring of the mRNA expression pattern of the TGF-β-isoforms (β1, β2, β3) during transdifferentiation of hepatic stellate cells using a newly developed real-time SYBR green PCR. Biochem. Biophys. Res. Commun. 295:330-335

Wieser R., G. Adam, A. Wagner, C. Schuller, G. Marchler, H. Ruis, Z. Krawiec and T. Bilinski. (1991). Heat shock factor-independent heat control of transcription of the CTT1 gene encoding the cytosolic catalase T of *Saccharomyces cerevisiae*. J Biol Chem 266:12406-11.

Wodicka, L., H. Dong, M. Mittmann, M. H. Ho and D. J. Lockhart. (1997). Genome-wide expression monitoring in *Saccharomyces cerevisiae.* Nat. Biotech. 15:1359-1367.

Wuytswinkel, O. V., V. Reiser, M. Siderius, M. C. Kelders, G. Ammerer, H. Ruis, and W. H. Mager. (2000). Response of *Sacchaormyces cerevisiae* to severe osmotic stress: evidence for a novel activation mechanism of the HOG MAP kinase pathway. Mol. Microbiol. 37:382-397

Yancey P. H., M. E. Clark, S. C. Hand, R. D. Bowlus and G. N. Somero. (1982). Living with water stress: evolution of osmolyte systems. Science 217:1214-1222.

Yodzis, P. (1988). The indeterminancy of ecological interactions as perceived through perturbation experiments. Ecology 69:508-515.

Yuen, T., E. Wurmbach, R. L. Pfeffer, B. J. Ebersole and S. C. Sealfon. (2002). Accuracy and calibration of commercial oligonucleotide and custom cDNA microarrays. Nucleic Acids Res. 30:48-56.

Xia, X. and Z. Xie. (2001). AMADA: analysis of microarray data. Bioinformatics 17:569-570.

Zhang, L., Y. Zhang, Y. Zhou, S. An, Y. Zhou and J. Cheng. (2002). Response of gene expression in *Saccharomyces cerevisiae* to amphotericin B and nystatin measured by microarrays. J. Antimicrob. Chemo. 49:905-915.