Dynamic Regulation of Metabolism in Archaea

by

Horia Todor

Department of Biology
Duke University

Date:_______________________

Approved:

___________________________
Amy Schmid, Supervisor

___________________________
Nicolas Buchler

___________________________
Fred Nijhout

___________________________
Paul Magwene

___________________________
Jen Tsan Chi

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2015
Dynamic Regulation of Metabolism in Archaea

by

Horia Todor

Department of Biology
Duke University

Date:_______________________

Approved:

________________________________________
Amy Schmid, Supervisor

________________________________________
Nicolas Buchler

________________________________________
Fred Nijhout

________________________________________
Paul Magwene

________________________________________
Jen Tsan Chi

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2015
Abstract

The regulation of metabolism is one of the key challenges faced by organisms across all domains of life. Despite fluctuating environments, cells must produce the same metabolic outputs to thrive. Although much is known about the regulation of metabolism in the bacteria and the eukaryotes, relatively little is known about the regulation of metabolism in archaea. Previous work identified the winged helix-turn-helix transcription factor TrmB as a major regulator of metabolism in the model archaeon *Halobacterium salinarum*. TrmB was found to bind to the promoter of 113 genes in the absence of glucose. Many of these genes encode enzymes involved in metabolic processes, including central carbon metabolism, purine synthesis, and amino acid degradation. Although much is known about TrmB, it remains unclear how it dynamically regulates its ~100 metabolic enzyme-coding gene targets, what the effect of transcriptional regulation is on metabolite levels, and why TrmB regulates so many metabolic processes in response to glucose. Using dynamic gene expression and TrmB-DNA binding assays, we found that that TrmB functions alone to regulate central metabolic enzyme-coding genes, but cooperates with various regulators to control peripheral metabolic pathways. After determining the temporal pattern of gene expression changes and their dependence on TrmB, we used dynamic metabolite profiling to investigate the effects of transcriptional changes on metabolite levels and
phenotypes. We found that TrmB-mediated transcriptional changes resulted in substantial changes in metabolite levels. Additionally, we showed that mis-regulation of genes encoding enzymes involved in gluconeogenesis in the ΔtrmB mutant strain in the absence of glucose results in low PRPP levels, which cause a metabolic block in de novo purine synthesis that is partially responsible for the growth defect of the ΔtrmB mutant strain. Finally, using a series of quantitative phenotyping experiments, we showed that TrmB regulates the gluconeogenic production of sugars incorporated into the cell surface S-layer glycoprotein. Because S-layer glycosylation is proportional to growth, we hypothesize that TrmB transduces a growth rate signal to co-regulated metabolic pathways including amino acid, purine, and cobalamin biosynthesis. Taken together, our results suggest that TrmB is a global regulator of archaean metabolism that works in concert with other transcription factors to regulate diverse metabolic pathways in response to nutrients and growth rate.
Dedication

To my parents, for showing me the way.
2.3.7 Modeling approach and fitting................................................................................ 24
2.4 Results .............................................................................................................................. 26
2.4.1 Gene expression dynamics in response to nutrients are dependent on TrmB . 26
2.4.2 NanoString measurement of temporal gene expression reveals metabolic
modularity governed by TrmB................................................................................................. 31
2.4.3 TrmB promoter occupancy can explain gene expression dynamics for ppsA...37
2.4.4 Prediction of gene expression based on TrmB-promoter enrichment ............... 40
2.4.5 Inferring regulatory network topology from dynamic gene expression output.
............................................................................................................................................... 44
2.5 Discussion....................................................................................................................... 48
2.6 Supplementary Data ...................................................................................................... 52
2.7 Acknowledgement ......................................................................................................... 52

3. Dynamic metabolite profiling in an archaeon connects transcriptional regulation to
metabolic consequences........................................................................................................... 54
3.1 Abstract............................................................................................................................ 54
3.2 Introduction..................................................................................................................... 55
3.3 Methods ........................................................................................................................... 58
3.3.1 Strains........................................................................................................................... 58
3.3.2 Growth conditions and growth rate calculation................................................... 58
3.3.3 Collection and quantification of metabolomics time course samples.............. 59
3.3.4 Normalization, outlier removal, and clustering of Metabolon data ............... 60
3.3.5 Analysis of purine metabolites and purine genes ........................................ 61
3.4 Results and Discussion ................................................................................................. 62
3.4.1 TrmB-driven transcriptional regulation directly and specifically affects the levels of certain metabolites........................................................................................................................................62

3.4.2 Temporal dynamics of growth recovery.........................................................................................................................67

3.4.3 Purine synthesis is negatively regulated by purine levels .........................................................................................................................................................................................................70

3.4.4 Defect in purine synthesis may be due to lack of PRPP ..................................................................................71

3.4.5 PRPP is found at a low level in ΔtrmB mutant strain in the absence of glucose ........................................................................................................................................74

3.4.6 The oxidative pentose phosphate pathway proceeds through gluconate......76

3.6 Conclusions .................................................................................................................................................................78

3.7 Supplementary Data .........................................................................................................................................................78

4. A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon Halobacterium salinarum. ........................................................................................................................................79

4.1 Abstract..............................................................................................................................................................................79

4.2 Introduction.......................................................................................................................................................................80

4.3 Methods ...........................................................................................................................................................................83

4.3.1 Strains and plasmids..................................................................................................................................................83

4.3.2 Growth Conditions.......................................................................................................................................................84

4.3.3 Cell lysis, S-layer enrichment, and staining protocol..................................................................................85

4.3.4 Microscopy and image quantitation.......................................................................................................................86

4.3.5 Collection and analysis of GC-MS samples ........................................................................................................87

4.3.6 Analysis of microarray and growth data ...........................................................................................................88

4.4 Results...............................................................................................................................................................................89

4.4.1 The ΔtrmB mutant is round in the absence of glucose.......................................................................................89
4.4.2 Carbohydrate staining suggests that $\Delta trmB$ morphological defects are associated with reduced S-layer glycosylation. ............................................................. 92

4.4.3 The $\Delta trmB$ mutant glycosylation defect is associated with an inadequate supply of carbohydrate precursors.......................................................... 95

4.4.4 Direct detection of intracellular metabolite concentrations reveals inadequate supply of carbohydrates in the $\Delta trmB$ mutant strain................................................. 98

4.4.5 TrmB activity is correlated with instantaneous growth rate.............................. 99

4.5 Discussion ...................................................................................................................... 102

4.6 Supplementary Data .................................................................................................... 108

4.7 Acknowledgements ...................................................................................................... 108

5. Conclusion and Future Directions ...................................................................................... 110

5.1 Conclusion ..................................................................................................................... 110

5.2 Future directions .......................................................................................................... 112

5.2.1 Biochemical characterization of TrmB binding mechanism............................... 113

5.2.2 Evolutionary perspectives ...................................................................................... 113

5.3 Funding .......................................................................................................................... 114

Appendix A................................................................................................................................ 115

References .................................................................................................................................. 123

Biography ................................................................................................................................... 135
List of Tables

Table 1: P-values of enrichment of COG terms in k-means clusters of dynamic expression profiles of the Δura3 and Δura3ΔtrmB strains after glucose stimulus. ...........33

Table 2: P-values of enrichment of COG terms in k-means clusters of Δura3 model residuals .......................................................................................................................................44

Table 3: Summary of metabolomics experiments.............................................................................................................................................60

Table 4: Number of Compounds detected in different metabolite classes. ..................63

Table 5: p-values of enrichment of functional categories in clusters. ..........................67

Table 6: Metabolites significantly correlated with growth rate in the ΔtrmB mutant strain and the Δura3 isogenic parent strain. .............................................................................................................68

Table 7: Bonferroni corrected p-values of the T-test of difference from CDM.............73
List of Figures

Figure 1: *ppsA* and *pykA* gene expression exhibits state-change dynamics in response to glucose perturbation when TrmB is present ................................................................. 28

Figure 2: *ppsA* (A) and *pykA* (B) do not respond significantly to 5% sucrose in the Δura3 strain (black lines) ........................................................................................................... 30

Figure 3: NanoString gene expression measurements are tightly correlated with qPCR measurements ............................................................................................................................. 32

Figure 4: TrmB is required for metabolic modularity in *H. salinarum* metabolism .......... 36

Figure 5: (A) TrmB-binding enrichment at the *ppsA* promoter is correlated with (B) predicted mRNA synthesis rate inferred from the gene expression data (‘Methods’ section) at the TrmB-dependent *ppsA* promoter after the addition of 5% glucose .......... 39

Figure 6: ODE model fit to NanoString gene expression data in response to glucose is predictive of gene expression in the Δura3 (parent strain) and Δura3ΔtrmB mutant background for some but not all genes ................................................................. 41

Figure 7: The difference between model-predicted and actual gene synthesis rates in Δura3 (black) and Δura3ΔtrmB (gray) in clusters over time in response to glucose ......... 43

Figure 8: (A) An FFL is the most likely topology for the regulation of cobalamin synthesis. Model of proposed FFL involved in cobalamin biosynthesis is shown .......... 46

Figure 9: (A) Metabolite levels measured by Metabolon are well-correlated with targeted measurements of amino acids and organic acids. Log-log linear regression of metabolites and time points in common is shown in red. (B) Metabolon’s untargeted assay can measure metabolites with levels in the millimolar intracellular range in both the Δura3 parent strain (dark bars) and the ΔtrmB mutant strain (red bars) ......................................................... 64

Figure 10: Figure depicts hierarchical clustering using six clusters on the mean-scaled and normalized combined Δura3 (black lines) and ΔtrmB (red lines) data. Each cluster graph depicts metabolite data for individual metabolites (thinner lines) and the mean expression profile for the cluster (thicker lines). ................................................................. 66
Figure 11: Instantaneous growth rate of the Δura3 parent strain (A) and ΔtrmB mutant strain (B) during the glucose addition (red lines) and control (black lines) time course. Error bars represent the standard error of 3 biological replicates.

Figure 12: Genes encoding enzymes involved in de novo purine synthesis are negatively correlated to intracellular purine levels.

Figure 13: The average growth potential (Methods) of the Δura3 parent strain (dark bars) and ΔtrmB mutant strain (light bars) in CDM, CDM supplemented with 5% glucose (CDM+ Gluc), and CDM supplemented 100 µM of adenine (CDM+Ade), adenosine (CDM+AdeN), guanine (CDM+Gua), and guanosine (CDM+GuaN). Error bars represent the standard error of 7 biological replicates.

Figure 14: PRPP levels in the Δura3 parent strain (dark bars) and the ΔtrmB mutant strain (light bars). Glucose complements the ΔtrmB mutant’s PRPP deficiency. Error bars represent the standard error of 8 biological replicates. Detection limit calculated from standard curve is shown as a dashed grey line.

Figure 15: Log-transformed glucose (black line), gluconate (blue line), and adenosine (red line) levels in the ΔtrmB mutant strain are temporally separated suggesting these metabolites are sequential steps of the OPPP and purine synthesis pathways.

Figure 16: The ΔtrmB mutant strain exhibits altered cell morphology in the absence of glucose.

Figure 17: S-layer protein glycosylation is deficient in the ΔtrmB mutant strain in the absence of glucose.

Figure 18: The ΔtrmB mutant (light grey line) exhibits impaired growth at lower 2DG (A) concentrations than the Δura3 parent strain (dark grey line). In contrast, the growth inhibition due to bacitracin (B) is constant between the ΔtrmB mutant (light line) and the Δura3 parent strain (dark line).

Figure 19: The ΔtrmB mutant strain (light bars) is deficient in sugars in the absence of glucose.

Figure 20: Inferred TrmB–DNA binding (black line) correlates with growth rate (grey line) across growth in batch culture (A), during the shift from anoxic to oxic conditions in a turbidostat (B), during a sublethal challenge with H₂O₂ (C), and during mid-log phase growth in media of varying salinity (D).
Acknowledgements

I would like to thank my advisor, Amy Schmid, who has been an amazing mentor and has given me the structure, resources, and ideas that resulted in this dissertation.

The rest of the Schmid Lab has also been a great source of support, friendship, and comraderie. I’d especially like to thank Adrianne Pittman and Kriti Sharma for keeping the lights on, the chemicals stocked, and the lab “Danger Officer” out of danger. I’d also like to thank Keely Dulmage for putting up with my constant questions, puns, and “OMG you have to look at this totally funny article I found online”. Also, the dress was black and blue.

In addition, I would like to acknowledge the great community of Biograds as well as the departmental support staff. In particular, Anne Lacey for keeping me paid and up to date with grad school requeriments.

Finally, I would like to thank my family and Erica for their constant support and advice. Ski weekends, long dinners, and neverending bottles of wine undoubtedly helped keep me sane. Thanks!
1. Introduction

1.1 Motivation and Perspective

In order to survive and reproduce, all organisms must be able to assimilate organic compounds from the environment and process them into energy and building blocks for the next generation. These chemical reactions, referred to as metabolism, are as varied as the organisms and the environments in which they exist. Some organisms metabolize glucose for energy using oxygen as a terminal electron acceptor, while others reduce iron in anoxic conditions at temperatures of over 121°C (Kashefi et al., 2003). In addition to having the necessary enzymes and metabolic processes to thrive in a certain environment, an organism must also be able to regulate its metabolism to ensure a proper supply of building blocks and to avoid the buildup of toxic intermediates.

Metabolic regulation has been thoroughly studied in eukaryotes such as humans and yeast, and in bacteria such as *Escherichia coli* and *Bacillus subtilis*. However, little is known about how metabolism is regulated in the archaeal domain. Many archaea are extremophiles, living in extreme environments such as undersea volcanic vents, saturated brines, and extreme cold. In these environments they use non-canonical energy generating pathways and enzymes to thrive. Understanding how their metabolism is regulated can provide a unique perspective on the evolution of metabolic regulation in diverse environments. Previous work in the hyperthermophilic species *Thermococcus*
*kodakarenensis* and *Pyrococcus furiosus* identified TrmB as a transcriptional regulator of maltose, trehalose, and maltotriose metabolism and gluconeogenesis (Lee *et al.*, 2003, Kanai *et al.*, 2007). In contrast to the limited regulatory scope of TrmB from these hyperthermophilic species, the TrmB homologue from the model archaeon *Halobacterium salinarum* is a major regulator of metabolism. In the absence of glucose, TrmB binds to a conserved cis-regulatory sequence in the promoters of 113 genes, many of which encode enzymes involved in metabolic processes. When bound to the DNA, TrmB can act as both an activator, by helping to recruit the transcriptional apparatus (Ouhammouch *et al.*, 2003), or as a repressor, by binding the TATA box and occluding the transcriptional apparatus (Bell *et al.*, 1999). Glucose abolishes TrmB DNA binding, leading to de-activation or de-repression of regulated promoters (Schmid *et al.*, 2009). Although TrmB’s binding sites in *H. salinarum* have been identified, much remains unclear about how and why it regulates metabolism in response to glucose. In order to address these questions, we undertook the following study.

### 1.2 Archaea

The Archaea represent one of the three primary domains of cellular life, together with the Bacteria and the Eukaryotes. Archaea are found all over the world in both extreme and mesophilic environments. Despite their preponderance in the environment and their unique lineage, the first archaeon was isolated only ~130 years ago (Farlow,
1878), and Archaea was not recognized as a unique domain until the pioneering work of Carl Woese (Woese & Fox, 1977). Initially mischaracterized as bacteria, archaea are morphologically similar to bacteria. Like bacteria, archaea grow as relatively small, anucleate cells. Most archaea are motile. Despite these morphological similarities, molecular techniques revealed that archaea are a unique blend of bacteria, eukaryotic, and archaeal biology. For example, the N-glycosylation pathway in archaea is similar to the eukaryotic pathway (Eichler, 2013), while some metabolic pathways are similar to those in bacteria. The archaeal cell membrane lipids and flagellar composition, on the other hand are not similar to anything found in either the eukaryotic or bacterial domain (Sprott, 2011, Jarrell & Albers, 2012).

One of the best examples of how archaea include bacterial, eukaryotic, and uniquely archaeal biology is in their transcriptional apparatus and regulation. The transcriptional apparatus of archaea is similar to that of eukaryotes, consisting of a RNA polymerase homologous to eukaryotic RNA polymerase II, a TFIIB homologue, and a TATA-binding protein (TBP) homologue (Micorescu et al., 2008). On the other hand, many transcriptional regulators in archaea are similar to those of bacteria. Specifically, both domains use transcriptional regulators with a helix-turn-helix or winged-helix-turn-helix motif that often bind DNA and effector molecules directly (Bonneau et al., 2007, Karr, 2014).
There are several major phyla within the archaeal domain (Brochier-Armanet et al., 2011). The best characterized of these are the crenarchaea and the euryarchaea (Woese et al., 1990). These phyla have some major differences. For example, most euryarchaea are highly polyploid, while most crenarchaea are monoploid (Soppa, 2011). Additionally, the cell division machinery in euryarchaea is FtsZ based, while crenarchaea use a system similar to the endosomal sorting complex required for transport (ESCRT) that eukaryotes use for vesiculation to divide (Ettema & Bernander, 2009). The class Halobacteria belongs to the euryarchaeal domain, and therefore is closely related to the methanogens, the thermococci, the pyrococci, and the archaeoglobi.

1.3 Halobacterium salinarum

*H. salinarum* is a hypersaline adapted euryarchaeon. It was initially isolated in 1922 by Harrison and Kennedy as a red discoloration on a salted codfish (Harrison & Kennedy, 1922). *H. salinarum* grows optimally aerobically, at 42°C, and requires nearly saturated sodium chloride concentrations to survive and grow (~4.5 M; Gochnauer & Kushner, 1969). Like all obligate halophiles, *H. salinarum* has a specific requirement for sodium and will lyse in the presence of less than 1.5M Na⁺ (Mohr & Larsen, 1963). *H. salinarum* was among the first archaeal species to have its genome sequence completely sequenced. It was found that *H. salinarum* contained 2,400 ORFs, approximately 130 of which were putatively annotated as transcription factors (Ng et al., 2000, Bonneau et al.,
After the adaptation of homologous recombination based cloning (Peck et al., 2000) to *H. salinarum*, genetic research on the unique regulatory mechanisms required to survive in the extreme and varied environment of the saltern began in earnest. Using gene expression arrays and transcription factor knockouts, the gamma radiation, metal, oxygen, and many other stress responses of *H. salinarum* have been characterized (Whitehead et al., 2006, Kaur et al., 2006, Schmid et al., 2007).

In routine culture, *H. salinarum* generates energy using oxidative phosphorylation with oxygen acting as the terminal electron acceptor (Gonzales et al., 2008). However, *H. salinarum* is also able to generate energy anaerobically in three ways: (a) ferment arginine (Hartmann et al., 1980), (b) respire using DMSO or TMAO as a terminal electron acceptor (Muller & DasSarma, 2005), and (c) phototrophy using the light-driven proton pump, bacteriorhodopsin (Baliga et al., 2002; Hartmann et al., 1980).

Although *H. salinarum* can generate energy using four different kinds of metabolic pathways, by far the most efficient of these is the tricarboxylic acid (TCA) cycle followed by oxidative phosphorylation, as evidenced by having the highest growth rate under aerobic conditions (Hartmann et al., 1980). Interestingly, unlike many bacteria and other archaea, it does not appear that *H. salinarum* can metabolize glucose for energy. Specifically, both the encoding gene and the enzyme activity for phosphofructokinase are undetectable (Sonawat et al., 1990, Ng et al., 2000); *H. salinarum* also cannot use glucose as the sole source of carbon and energy (Gochnauer & Kushner, 1969), does not
actively transport glucose from the media (L.O. Severina, 1990), and does not convert \textsuperscript{13}C-labeled glucose to pyruvate (Sonawat et al., 1990). Since it cannot catabolize glucose or other sugars when growing aerobically, \textit{H. salinarum} catabolizes amino acids as a source of carbon, nitrogen, and energy. The pathways and enzymes involved in the degradation and synthesis of these amino acids have been well characterized, and are similar to bacterial pathways (Falb et al., 2008, Gonzales et al., 2008). Because it uses amino acids for both growth and for energy, \textit{H. salinarum} is routinely grown in either synthetic or complex media rich in amino acids. The defined medium contains 17 amino acids, as well as a vitamin and trace metal mixture (complete defined medium, CDM, Todor et al., 2013). Neither medium contains glucose or other sugars, since these compounds do not improve growth (Schmid et al., 2009). Although \textit{H. salinarum} has complex nutritional requirements for amino acids, it is able to make a number of secondary metabolites. Specifically, the \textit{H. salinarum} genome encodes pathways for the \textit{de novo} synthesis of purines, pyrimidines, cobalamin, thiamin, folate, and other cofactors (Ng et al., 2000, Falb et al., 2008). Many of these peripheral pathways appear to be relatively recent horizontal gene transfer (HGT) events (Koonin, 2003). Additionally, there is many pathways show evidence of homologous replacement events at the enzyme level, whereby the replacing enzyme performs the same reaction, but may use a different cofactor and mechanism (Galperin & Koonin, 1999, Siebers & Schonheit, 2005).
1.4 Metabolic regulation in \textit{Halobacterium salinarum}

The regulation of central metabolism and peripheral pathways to produce the correct outputs required for the cell to thrive under varying nutrient conditions is an important task for the cell. In many eukaryotic and bacterial cells, allosteric and posttranslational regulation plays an important role in helping cells maintain internal metabolic homeostasis. Purification and analysis of many enzymes from \textit{H. salinarum} has found only limited allosteric regulation (Siebers \textit{et al.}, 2005, Leicht \textit{et al.}, 1978, Hecht \textit{et al.}, 1990, de Medicis \textit{et al.}, 1982). Additionally, although recent studies suggest posttranslational regulation via phosphorylation may be important in the regulation of metabolic enzymes (Aivaliotis \textit{et al.}, 2009), definitive proof that phosphorylation affects enzyme function in \textit{H. salinarum} remains elusive. Together, the limited allosteric and posttranslational regulation observed in \textit{H. salinarum} suggests that transcriptional regulation may play an important role in regulating metabolic flux. Several transcriptional regulators of metabolism have been identified in \textit{H. salinarum}: the Lrp family and TrmB.

\textit{H. salinarum} has been found to possess eight leucine-responsive regulatory protein (Lrp) homologues. These proteins possess a helix-turn-helix DNA binding domain connected by a flexible linker region to a RAM (regulation of amino acid metabolism) domain (Schwaiger \textit{et al.}, 2010). In \textit{E. coli}, Lrp is a central metabolic
regulator directly responsible for the expression of over 130 genes (Cho et al., 2008). In contrast to the central role played by E. coli Lrp, homologues in H. salinarum were found to play a more nuanced role. Recent studies using gene expression profiling and ChIP-chip genome-wide binding arrays found that these transcription factors coordinately activate and repress an interconnected web of genes in response to specific growth, stress, and amino acid effectors (Plaisier et al., 2014, Ashworth et al., 2014). For example, VNG1237C, which has been putatively implicated in the regulation aromatic amino acid synthesis (Gulko et al., 2014) was found to activate and repress genes involved in a variety of processes. Deletion of VNG1237C, unlike deletion of the other lrps, resulted in a growth defect under normal growth conditions. In contrast to the role played by VNG1237C, AsnC was found to be involved in responding to sub-inhibitory concentrations of paraquat (Plaisier et al., 2014, Ashworth et al., 2014). Despite the large number of Lrp family members in H. salinarum, it appears that their role is more specialized than global.

TrmB was initially identified in P. furiosus as a Transcriptional Regulator of Maltose metabolism. Using gene expression measurements and genome wide promoter binding ChIP-chip assays, it was found that TrmB homologue in H. salinarum (VNG1451C) is a global regulator of many metabolic pathways, binding to the promoters of 113 genes and operons (Schmid et al., 2009). Many of the genes regulated by TrmB encode enzymes involved in diverse metabolic processes, including
gluconeogenesis, purine synthesis, cobalamin (vitamin B12) synthesis, amino acid metabolism, and thiamin synthesis. This broad regulatory reach suggests that TrmB may play a more global role in the regulation of metabolism. TrmB was found to be able to either activate or repress gene expression, depending on the location of its conserved cis-regulatory motif (TACT-N(7-8)-AGTA) relative to the transcription start site and TATA box. When the binding site overlaps the TATA box, the transcriptional apparatus may be occluded from assembling (Bell et al., 1999). In this case, TrmB can function as a repressor. When the binding site is located proximal to the TATA box, TrmB can help recruit the transcriptional complex, acting as an activator (Ouhammouch et al., 2003). The addition of glucose, but not other sugars, to the media abolishes TrmB-DNA binding, leading to depression of repressed genes and deactivation of activated genes (Schmid et al., 2009, Todor et al., 2013). The genes that show the largest change in expression upon the addition of glucose (and the unbinding of TrmB) are those encoding enzymes involved in gluconeogenesis. As expected, these genes are activated by TrmB-DNA binding in the absence of glucose and are deactivated upon the addition of glucose. Using the homologous recombination based cloning method previously described (Peck, 2000), an in frame deletion mutant of trmB (ΔtrmB) was generated (Schmid et al., 2009). When grown in the absence of glucose, this mutant was found to have a severe growth defect (~1/5 the wildtype growth rate), a morphological defect, and a defect in NAD+/NADH ratio. These phenotypes were specifically and completely
complemented by the addition of glucose to the media. Although much is known about the role of TrmB in *H. salinarum*, much remains to be learned. Specifically, the important role of the glucose responsive transcription factor TrmB, as evidenced by the plethora of phenotypes in the *trmB* knockout, stands at odds with the inability of *H. salinarum* to metabolize glucose for energy. Additionally, TrmB-DNA binding is controlled only by glucose availability. Given the conserved *cis*-regulatory site of TrmB, it remains to be determined if TrmB regulates all of its target genes in the same way. If not, TrmB may work with other regulators to appropriately regulate various pathways. How TrmB regulates its target genes, and any additional regulators involved remain to be discovered.

1.5 TrmB in other archaea

TrmB is widely conserved across the archaeal domain, having been identified and characterized in *Pyrococcus furiosus*, *Thermococcus litoralis*, *Thermococcus kodakarensis*, and *Methanosarcina acetivorans* (Lee, *et al.*, 2003, Lee, *et al.*, 2005, Kanai *et al.*, 2007, Reichlen *et al.*, 2012). TrmB was initially identified in *T. litoralis* part of a 16 kb horizontal gene transfer from *P. furiosus*. The 16 kb fragment contained genes for the uptake and metabolism of trehalose. Interestingly, despite the gene transfer event, these genes remained inducible, suggesting that the regulation of these genes was in the 16 kb region transferred from *P. furiosus*. TrmB was identified as the transcriptional regulator
responsible for repression of these genes in the absence of sugar. Using electrophoretic mobility shift assays (EMSA), it was found that both maltose and trehalose bound TrmB. Maltose abolished binding, while trehalose strengthened binding (Lee et al., 2003). Further analysis in both P. furiosus and T. litoralis revealed that TrmB bound and differentially regulated an additional promoter – the maltodextrin (MD) promoter. This operon is regulated by TrmB binding to sucrose and malotriose, revealing wide sugar specificity for TrmB. In addition, using promoter scanning mutagenesis, it was shown that the inverted palindrome TACTNNNAGTA was important for TrmB binding function (Lee et al., 2005). Interestingly, despite large differences in morphology, habitat, and metabolism between anaerobic hyperthermophiles and hypersaline aerobes, TrmB in both P. furiosus and H. salinarum binds to a similar inverted palindromic sequence.

Additional research in the hyperthermophiles identified several TrmB homologues, including PF0124 and Thermococcus kodakarensis TK1769. These homologues appear to encode a protein which functions as a global sugar sensing transcriptional regulator, rather than a specific sugar metabolism and transport operon. Specifically, an inverted repeat site (TATCAC-N(5)-GTGATA) was identified in the promoter region of glycolytic and starch utilization genes in P. furiosus and named the Thermococcales Glycolitic Motif (TGM). In contrast, in T. kodakarensis, the TrmB homologue TK1769, also known as Tgr (Thermococcales glycolytic regulator), was shown to function as both an activator and as a repressor of glycolytic and gluconeogenic genes. A Δtgr strain was
generated, which was found to be defective in growth in gluconeogenic, but not glycolytic conditions. Using EMSA, maltotriose was identified as the physiological effector. Binding of this sugar to Tgr resulted in dissociation from the DNA (Kanai et al., 2007).

The most diverged TrmB homologue was discovered in *M. acetivorans*. A protein, MreA, was found to be a central regulator of distinct methanogenic pathways. A Δ*mreA* strain was generated, and found to be defective in growth on acetate but not on methylotrophic substrates. Expression profiling of the Δ*mreA* strain found 280 genes with significantly altered expression, suggesting that MreA is a global regulator of metabolism in *M. acetivorans* (Reichlen et al., 2012).

Taken together, the TrmB family of transcription factors appears to play several distinct but related roles in the archaea. Some TrmB homologs appear to be specific repressors of sugar utilization pathways, such as the first TrmB family member identified in *T. litoralis*. Other TrmB homologues, such as TK1769 and VNG1451C function as global regulators of gluconeogenic and glycolytic flux. Regardless, all homologues of TrmB share a conserved mechanism. DNA binding occurs at a specific *cis*-regulatory sequence, and this binding is abolished by sugar binding to the TrmB protein. Although much is known about the function of TrmB as a global regulator in *H. salinarum*, much remains to be determined about how TrmB regulates its 113 targets, the effects of TrmB transcriptional regulation on metabolite levels, and why TrmB responds...
to glucose in a non-glucose metabolizing archaeon. In this work, we address these questions using gene expression measurement, metabolite assays, and targeted phenotypic tests. Using dynamic gene expression and TrmB- promoter occupancy profiling, we find that TrmB directly regulated genes encoding enzymes involved in central carbon metabolism, while working with other regulators to regulate purine metabolism and cobalamin synthesis (Todor et al., 2013). Using a combination of targeted and untargeted metabolomics techniques, we show that transcriptional regulation through TrmB causes major changes in metabolite levels. Using these data, we demonstrate that the pentose phosphate pathway proceeds through gluconate. Further, as a result of deactivation of gluconeogenesis caused by the deletion of trmB, purine levels are affected. Supplementation of growth media with purines reveals that the defect in purine biosynthesis is partially responsible for the growth defect of the ΔtrmB mutant strain in the absence of glucose. Finally, we show that a primary use of glucose in H. salinarum is the glycosylation of the S-layer protein. Since the amount of new S-layer is proportional to the growth rate of H. salinarum, TrmB is able to sense growth rate and therefore transduces a growth rate signal to co-regulated pathways (Todor et al., 2014).
1.6 Statement of collaboration and independent effort

Each of the projects that I describe in this dissertation involved collaborations with other labs, researchers, and undergraduate students. In each case, I conceived, designed and performed experiments, analyzed data, and wrote the resulting manuscript. I designed the project reported in Chapter 2, but co-authors Adrianne Pittman and Kriti Sharma were instrumental in performing the ChIP-qPCR experiments. The project reported in Chapter 3 is the result of collaboration with Chris Newgard, Olga Ilkayeva, and Jessica Gooding at the Sarah Stedman Center for Nutrition and Metabolism, Duke Molecular Physiology Institute. Their metabolomics expertise was crucial to metabolic profiling in *H. salinarum*. Finally, James Bain and Mike Muehlbauer, also at the Sarah Stedman Center for Nutrition and Metabolism, performed the metabolomics assays reported in Chapter 4. Additionally, Keely Dulmage and Nick Gillum provided data for several of the supplementary figures.
2. Protein-DNA binding dynamics predict transcriptional response to nutrients in archaea

2.1 Abstract

1 Organisms across all three domains of life use gene regulatory networks (GRNs) to integrate varied stimuli into coherent transcriptional responses to environmental pressures. However, inferring GRN topology and regulatory causality remains a central challenge in systems biology. Previous work characterized TrmB as a global metabolic transcription factor in archaeal extremophiles. However, it remains unclear how TrmB dynamically regulates its ~100 metabolic enzyme-coding gene targets. Using a dynamic perturbation approach, we elucidate the topology of the TrmB metabolic GRN in the model archaeon *Halobacterium salinarum*. Clustering of dynamic gene expression patterns reveals that TrmB functions alone to regulate central metabolic enzyme-coding genes, but cooperates with various regulators to control peripheral metabolic pathways. Using a dynamical model, we predict gene expression patterns for some TrmB-dependent promoters and infer secondary regulators for others. Our data suggest feed-forward gene regulatory topology for cobalamin biosynthesis. In contrast, purine biosynthesis appears to require TrmB-independent regulators. We conclude that

TrmB is an important component for mediating metabolic modularity, integrating nutrient status and regulating gene expression dynamics alone and in concert with secondary regulators.

2.2 Introduction

Diverse metabolic processes must be differentially regulated in order to maintain homeostasis and optimize growth in changing environmental and intracellular conditions. Environmental fluctuations occur at several temporal scales. This is mirrored in the integrated transcriptional and metabolic regulatory networks at the enzymatic, transcriptional, and post-transcriptional levels in organisms responding to fluctuating conditions (Seshasayee et al., 2009). For example, transient nutrient changes may result in microsecond-scale regulation of enzyme activity, whereas prolonged exposure over the course of minutes or hours may trigger changes in the levels of enzyme-coding transcripts (Alon, 2007a). Because of the potential for the buildup of toxic intermediates and futile cycles, temporal dynamics of metabolic enzyme-coding transcripts can be as important as overall levels (Goyal et al., 2010). Over evolutionary time scales, the constant presence of a given nutrient may lead to gene loss in competing metabolic pathways. Such streamlining of the genome is thought to enable faster replication and long-term adaptations in network structure or topology (Koonin & Wolf, 2010).
Discerning the dynamic function of the underlying network resulting from the interaction of many different levels of regulation remains a central challenge.

In archaea, evidence is mounting that transcription may be an important mechanism for regulating metabolism. Unlike eukaryotes and bacteria, in which regulation of flux through central carbon metabolism and other pathways appears to be allosteric (Johnsen et al., 2003), archaea seem to lack many classic allosteric regulatory control points (Siebers & Schonheit, 2005). For example, in hypersaline-adapted archaea, glutamate dehydrogenase was found to be unresponsive to ADP and GDP (Leicht et al., 1978), while D-Lactate dehydrogenase was not regulated by fructose 1,6 bisphosphate (Hecht et al., 1990). Additionally, pyruvate kinase from H. salinarum was found to possess only weak allosteric regulation (de Medicis et al., 1982). Other potential regulatory mechanisms such as protein phosphorylation have been proposed in two different archaeal species (Aivaliotis et al., 2009, Esser et al., 2012); however, in the hypersaline-adapted archaeal model system H. salinarum, much of the missing regulation seems to take place at the level of transcription (Baliga et al., 2002, Schmid et al., 2007, Bonneau et al., 2007, Muller & DasSarma, 2005). This property of haloarchaea provides a simplified model system for understanding the underlying logic of the integrated transcriptional and metabolic network. The largely transcriptional nature of archaeal responses has already been used to infer regulatory networks with a remarkable degree of accuracy (Bonneau et al., 2007). H. salinarum survives in an extreme
and fluctuating environment, where daily and seasonal changes in salinity, oxygen, and nutrients require constant adjustment of metabolism (Oren, 1999). Current evidence suggests that many of these metabolic adjustments are regulated transcriptionally.

For example, in *H. salinarum*, TrmB has been characterized as the central transcriptional regulator of carbon metabolic pathways under aerobic conditions (Schmid *et al.*, 2009). Conserved throughout the archaea, TrmB is a winged helix-turn-helix transcription factor (TF) that binds a palindromic inverted repeat cis-regulatory sequence in *Pyrococcus furiosus* (Lee *et al.*, 2007), *Thermococcus litoralis* (Lee *et al.*, 2003) and *H. salinarum* (Schmid *et al.*, 2009). TrmB acts as a repressor at some promoters and as an activator at others. In the obligately anaerobic hyperthermophilic archaea, TrmB has been characterized as a regulator specifically involved in oligosaccharide transport and catabolism (Lee *et al.*, 2008). In contrast, TrmB in *H. salinarum* has been found to bind 113 promoters in the genome to regulate genes involved in diverse processes including central carbon metabolism, TCA cycle, amino acid metabolism, and cofactor metabolism (Schmid *et al.*, 2009).

Yet even in these simplified systems, where metabolic flux should largely be predictable from transcriptional data alone, regulatory causality is complicated by metabolic feedback. For example, TrmB appears to be regulated at the level of TF activity (Schmid *et al.*, 2009, Bonneau *et al.*, 2007), binding sugars (glucose, trehalose, maltose) with varying affinities, which in turn decrease TrmB affinity for DNA (Lee *et
al., 2007, Lee et al., 2003, Schmid et al., 2009). However, by de-activating transcription of ppsA (encoding phosphoenolpyruvate synthase) and other gluconeogenic genes while de-repressing transcription of pykA (encoding pyruvate kinase) and other glycolytic genes, the amount of glucose is decreased, causing increased TrmB-promoter binding and reversing the effect (Schmid et al., 2009). In bacteria, the enzymes involved in glycolysis and gluconeogenesis are regulated allosterically (Johnsen et al., 2003) in response to transient changes, and transcriptionally through cAMP Responsive protein (Crp in gram negative bacteria) in response to more permanent changes (Shimada et al., 2011). In contrast, we hypothesize that TrmB regulates metabolic enzyme-coding genes rapidly in response to stimuli, while also adapting the equilibrium levels of these genes to longer-term conditions.

Regulation of diverse pathways in response to environmental fluctuations of varying temporal scales is also likely to be mediated by several TFs working together in various combinations to produce appropriate transcriptional dynamics for each pathway (Bonneau et al., 2007, Harbison et al., 2004). Such network topology is difficult to infer on the basis of steady state measurements of gene expression and promoter occupancy. In order to unravel cause and effect relationships in the gene regulatory network (GRN) controlling metabolism, we measured gene expression using NanoString® (Geiss et al., 2008) and TrmB binding dynamics using ChIP-qPCR over time in response to a glucose stimulus. Integration of these data in the context of a
dynamical model enabled prediction of how TrmB and its partners enact several transcriptional programs in different ways across metabolic pathways.

2.3 Methods

2.3.1 Strains and growth conditions

*Halobacterium salinarum* NRC-1 (ATCC strain 700922) was used as the wild type strain background for all studies (Supplementary Table 1). Gene expression was assayed in a previously constructed strain containing an in-frame deletion of VNG1451C (Δura3ΔtrmB, Schmid et al., 2009) and its isogenic parent strain (Δura3). A previously constructed strain containing trmB::c-myc fusion on a low copy number plasmid was used for ChIP-qPCR to determine binding site occupancy (Schmid et al., 2009). Cells for gene expression were grown in Complete Defined Medium (CDM) containing 19 amino acids (modified from (Schmid et al., 2009, Supplementary Table 2) supplemented with 50 µg/ml uracil to complement the ura3 deletion. Strains carrying the trmB::c-myc construct for ChIP-qPCR were grown in CDM supplemented with 20 µg/ml mevinolin for plasmid maintenance. Cultures were grown at 225 r.p.m. shaking at 42°C under low ambient light.
2.3.2 mRNA preparation for gene expression time course

*H. salinarum Δura3* (parent) and *Δura3ΔtrmB* (knockout mutant) strains were grown to early logarithmic phase (OD600 ≈ 0.3). For gene expression time courses, three 4 mL aliquots were removed from the continuously shaking cultures before the addition of 5% (w/v) glucose (-240, -60, 0 minute time points) and seven afterwards (5, 10, 20, 45, 90, 180, 360 minute time points). Cells were immediately pelleted by centrifugation and snap frozen in liquid nitrogen. RNA was prepared using the Absolutely-RNA miniprep kit (Stratagene/Agilent) according to the manufacturer’s instructions. RNA quality was assessed using the Agilent 2100 BioAnalyzer using the RNA-Nano Chip and the Prokaryotic Nano RNA protocol (Agilent Technologies, Santa Clara, CA), and samples were verified to be free of DNA contamination by 25 cycles of PCR amplification on at least 200 ng RNA prior to cDNA conversion in RT-qPCR.

2.3.3 Reverse transcriptase quantitative PCR (RT-qPCR).

RNA extracted from cells was quantified using RT-qPCR. Briefly, RNA was quantified using the Power SYBR RNA-to-C\textsuperscript{T} 1-step kit (Applied Biosciences) in an Eppendorf Mastercycler ep Realplex thermocycler using Realplex software. Reaction size was 20 µL and reactions were prepared according to manufacturer’s instructions. Plate setup was performed robotically using a Corbitt Life Sciences liquid handling system. CQ threshold was determined automatically by the software. Primers were
crosschecked against the *H. salinarum* genome using BLAST to insure specificity.

Amplification efficiencies for each pair of primers (*ppsA* – VNG0330G, *pykA* – VNG0324G, *eif1a2* - VNG1756G) were determined by running three serial 10-fold dilutions of the same sample (see Supplementary Table 3 for primers). These efficiencies were used to calculate enrichment relative to the reference gene (VNG1756G) in RNA from time course sampling from the measured CQ values as previously described (Pfaffl, 2001). Each experiment consisted of 2-3 biological replicates (separate cultures), and 3 technical replicates. The significance of change in gene expression before and after nutrient addition was assessed using the Welch one-sided t-test. Using the first three and last two points of each biological replicate time course, we determined p-values for the difference in means equivalent to a 1.5-fold up- or down-regulation.

### 2.3.4 Measurement of gene expression using NanoString

Quantification of additional genes was performed on the same RNA samples using NanoString technology (Geiss *et al.*, 2008). 100 µg of each RNA sample from glucose time courses were delivered to NanoString (Seattle, WA), where samples were hybridized to a custom probe set (Supplementary Table 4) encompassing 100 genes of interest and counted using an nCounter machine. The same mRNA samples used for RT-qPCR were used for NanoString experiments. Data were normalized by total counts per sample across strains. All raw and normalized data are included in Supplementary
Table 5. For clustering, each gene was further mean centered and normalized to a standard deviation of 1. Gene expression profiles were clustered using k-means with a Pearson correlation distance function on the normalized gene expression data with k=8. The same clustering parameters were used for both ∆ura3 parent and ∆ura3ΔtrmB mutant datasets. Each cluster was analyzed for enrichment of COG categories (Tatusov et al., 2003, Tatusov et al., 1997) using the hypergeometric test.

2.3.5 ChIP-qPCR protocol

*H. salinarum* cells harboring the *trmB::c-myc* construct were grown to early logarithmic phase (OD600 ≈ 0.3) as described above. DNA-transcription factor (TF) complexes were crosslinked and immunoprecipitated using the c-myc epitope tag as previously described (Schmid et al., 2009). Primers were designed according to the criteria presented in (Mukhopadhyay et al., 2008). The qPCR thermocycling reaction and thermocycling conditions were as described previously (Schmid et al., 2011), except that the SsoAdvanced SYBR Green Supermix (Bio-Rad) was used. Each of 5 biological replicates was run in 3 technical replicates. Enrichment of TrmB binding at the *ppsA* promoter was calculated as relative enrichment of the immunoprecipitated sample versus the input. The ratios given in the figures compare enrichment at the binding peak to the 3′ end of the gene of interest (Schmid et al., 2011, Supplementary Table 3).
2.3.6 Degradation constant calculation

In order to assess the validity of our data and improve the accuracy of our modeling, we decided to incorporate genome wide mRNA degradation parameters. Although these parameters have been determined for a related strain (*Halobacterium salinarum* R1 ATCC 29341/DSM 671, Hundt *et al.*, 2007), the actual half-lives were not published. We therefore reanalyzed the microarray data (ArrayExpress accession number E-MEXP-1088, http://www.ebi.ac.uk/arrayexpress/) to determine these values (Supplementary Table 6). In short, we normalized each gene on each array to the t=0 time point, and for each gene, fit an exponential decay curve through the points that were annotated as valid in the microarray file using the nls() function in the statistical package R. Standard error was also calculated using nls().

2.3.7 Modeling approach and fitting

**Synthesis rate determination:** To better understand how binding events at the promoter influence gene expression, we calculated the derivative with respect to time for every gene at each time point by the weighted average of the slope of the segments before and after each time point. We used the $K_{deg}$ values from (Hundt *et al.*, 2007) to calculate the synthesis rate according to Formula 1.

$$\frac{d[mRNA]}{dt} + K_{deg}[mRNA] = K_{synthesis}$$
Gene Expression Prediction: In order to determine the importance of TrmB regulation to specific genes, we used an ordinary differential equation (ODE) model to predict gene expression dynamics in both WT and $\Delta$ura3$\Delta$trmB strains as a function of ChIP-qPCR enrichment at the $ppsA$ promoter. The first and last three points of the $\Delta$ura3 gene expression and TrmB-binding time courses were used to determine the values for $K_{basal}$ and $K_{eff}$ by setting the right hand side of Formula 2 to 0. The Hill coefficient (n) was set to either negative or positive one to indicate repression or activation, respectively. The level of gene expression during the nutrient stimulus was predicted by calculating the change in gene expression every minute according to Formula 2.

$$\frac{d[mRNA]}{dt} = -K_{deg}[mRNA] + K_{basal} + K_{eff} \cdot TrmB_{enrichment}$$

Calculating model residuals: In order to elucidate the topology of nutritional control of central and peripheral metabolism, we compared our predicted and actual gene expression data. Each gene was normalized to a maximum of 1. The normalized values were used as input to the TrmB-enrichment based predictive ODE model of gene expression (above). For each gene in each of the $\Delta$ura3$\Delta$trmB and $\Delta$ura3 strain backgrounds at each mRNA time point, we calculated the difference between the predicted synthesis rate and the actual synthesis rate as calculated above (see synthesis rate determination). We clustered resulting traces on the basis of the $\Delta$ura3 data using k-means clustering with k=5 and Euclidean distance.
Feed forward loop logic approximation: In order to assess whether a feed-forward loop might be responsible for the dynamics we observed in the cobalamin biosynthesis cluster, we simulated the network using a logic approximation (Figure 8A, Alon, 2007b). Genes involved in cobalamin biosynthesis were normalized so that the lowest value was 0 and the highest value was 1. A regulator was presumed active when its value was > 0.5. We used the TrmB ChIP-qPCR binding data as an input and fit the degradation parameter ($K_{deg\ FFL\ Effector}$, $K_{deg\ cob/cbi}$) for both of the other genes to the average scaled expression profile of the cob / cbi cluster using least squares. The fit was evaluated at 5, 10, 20, and 45 minutes following glucose addition. This is the period in which feed-forward dynamics would be expected to be most apparent.

2.4 Results

2.4.1 Gene expression dynamics in response to nutrients are dependent on TrmB

Our previous work demonstrated that TrmB binds the promoter of metabolic enzyme-coding genes throughout the genome in response to carbon source availability (Schmid et al., 2009). However, these studies were conducted at steady state. To investigate the dynamic expression response of TrmB-regulated genes to nutrients (glucose, glycerol and sucrose), we used quantitative reverse transcriptase PCR (RT-qPCR) to measure repressed (pykA) and activated (ppsA) gene levels in response to
nutrients over time (Figure 1 & Figure 2). TrmB is thought to regulate these genes by binding to the promoter either to activate or to repress expression in the absence of glucose. Addition of glucose to the medium results in TrmB dissociation from the promoter and de-activation or de-repression of the target gene (Schmid et al., 2009).

Briefly, *H. salinarum* cells were grown on amino acids as a carbon and energy source to mid-logarithmic phase. Cells were sampled thrice before and seven times after the addition of nutrients (Methods). We considered a change relevant when a 1.5-fold or greater up or down regulation of the target gene was significant (p < 0.05). Since these genes are a key control point in glycolysis (*pykA*) and gluconeogenesis (*ppsA*), their levels are highly informative of the regulation of that pathway (Tjaden et al., 2006, Siebers et al., 2004).
Figure 1: ppsA and pykA gene expression exhibits state-change dynamics in response to glucose perturbation when TrmB is present. Gene expression of ppsA (A) and pykA (B) to a 5% glucose stimulus in the Δura3 (black lines) and Δura3ΔtrmB (gray lines) strains were measured by RT-qPCR and are shown plotted on a logarithmic axis. Error bars represent the standard error from the average of at least two biological replicate experiments. Asterisks indicate significance of the difference in expression level between the beginning and the end of the time course; * significant at \( P < 0.05 \); ** significant at \( P < 0.01 \); *** significant at \( P < 0.001 \).
We observed that both *ppsA* and *pykA* mRNA exhibit a characteristic steady state expression value during mid-logarithmic phase growth prior to glucose addition that differs between the ∆ura3 parent and ∆ura3∆trmB mutant strain (Figure 1). Upon addition of glucose, *ppsA* is de-activated (Figure 1A, 3.6 fold decrease) and *pykA* is de-repressed (Figure 1B, 12.7 fold increase) by the first time point (5 minutes) and both genes reached a new equilibrium level by 45 minutes. Hereafter, we refer to such monotonic changes in gene expression as state-change dynamics. These dynamics were greatly attenuated in the ∆ura3∆trmB mutant background in response to glucose (Figure 1) and in the ∆ura3 parent strain in response to sucrose (Figure 2). Gene expression dynamics similar to those in response to glucose were observed with glycerol (Figure 2), although ultimately a different final equilibrium level is reached. These data confirm that gene expression dynamics of *ppsA* and *pykA* are specific to glucose and glycerol and dependent upon TrmB. Furthermore, they confirm the role of TrmB in regulating genes coding for enzymes in glycolysis/gluconeogenesis at both short and intermediate temporal scales.
Figure 2: *ppsA* (A) and *pykA* (B) do not respond significantly to 5% sucrose in the Δura3 strain (black lines). *ppsA* and *pykA* respond to 0.167% glycerol stimulus in the Δura3 strain (gray lines). Gene expression was measured by RT-qPCR and is shown plotted on a logarithmic axis. Asterisks indicate significance of the difference in expression level between the beginning and the end of the time course; * significant at $P < 0.05$; ** significant at $P < 0.01$; *** significant at $P < 0.001$. 
2.4.2 NanoString measurement of temporal gene expression reveals metabolic modularity governed by TrmB

In our initial experiments, both *ppsA* and *pykA* gene expression exhibited state-change TrmB-dependent dynamics. Because these genes are an important control point in glycolysis/gluconeogenesis (Siebers & Schonheit, 2005), state-change dynamics in response to glucose and glycerol were expected. However, TrmB regulates genes involved in processes across metabolism (Schmid et al., 2009). To assess the impact of TrmB on gene expression in response to glucose, we assayed mRNA levels over time following glucose stimulus for 100 genes involved in central and secondary metabolism using NanoString (Geiss et al., 2008). NanoString was used to assay gene expression because it has been shown to be sensitive, precise, and require minimal processing (Baugh et al., 2011). We selected 96 genes involved in central metabolism, including both direct and indirect TrmB targets and *trmB* itself (Supplementary Table 4). Using data from previous microarray experiments (Bonneau et al., 2007, Facciotti et al., 2010, Kaur et al., 2010, Schmid et al., 2007), we also selected 3 genes (*VNG1670C, mrp, srp19*) expressed at high, medium, and low levels with minimal variation across many conditions (Supplementary Table 4). The same glucose response time points used for RT-qPCR (Figure 1) were sampled in the NanoString experiment. Since NanoString has not been previously used for mRNA quantification in *H. salinarum*, we validated the data by comparing measurements of *ppsA* and *pykA* gene expression in the same RNA time course samples between the NanoString and RT-qPCR platforms. We found a linear
relationship (Pearson correlation = 0.962) across several orders of magnitude (Figure 3), suggesting that NanoString is a robust alternative to RT-qPCR for the quantification of mRNA in *H. salinarum*.

![Figure 3: NanoString gene expression measurements are tightly correlated with qPCR measurements. mRNA levels for *ppsA* (black points) and *pykA* (gray points) were measured by RT-qPCR and NanoString. Both RT-qPCR and NanoString are shown plotted on a logarithmic axis. Error bars represent standard error from two replicates of NanoString and at least two replicates of RT-qPCR.](image)

In order to determine the pattern of TrmB-dependent regulation across central and secondary metabolic pathways, NanoString gene expression profiles were clustered separately in the ∆ura3 and ∆ura3ΔtrmB strain backgrounds using k-means. The distribution of clusters was integrated with the TrmB-centered metabolic regulatory
network reconstruction (Schmid et al., 2009, Figure 4, Supplementary Table 5). Clusters of genes from both Δura3 and Δura3ΔtrmB expression profiles were analyzed for significant (p < 0.01) enrichment of membership in Clusters of Orthologous Groups (COGs) (Table 1, Tatusov et al., 1997). Surprisingly, we found that temporal gene expression patterns clustered according to metabolic pathway modules in the Δura3 strain and to a lesser extent in the Δura3ΔtrmB mutant, suggesting that TrmB is at least partially responsible for the maintenance of metabolic modularity (Figure 4C and 2.4F, Supplementary Table 5, Table 1).

Table 1: P-values of enrichment of COG terms in k-means clusters of dynamic expression profiles of the Δura3 and Δura3ΔtrmB strains after glucose stimulus.

<table>
<thead>
<tr>
<th>COG term</th>
<th>Description</th>
<th>Δura3</th>
<th>Δura3ΔtrmB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrate transport and metabolism</td>
<td>Glycolysis/gluconeogenesis</td>
<td>1.08E-03</td>
<td>NS</td>
</tr>
<tr>
<td>Amino acid transport and metabolism</td>
<td></td>
<td>2.95E-03</td>
<td>7.32E-04</td>
</tr>
<tr>
<td>Nucleotide transport and metabolism</td>
<td>Purine biosynthesis</td>
<td>1.35E-04</td>
<td>1.23E-09</td>
</tr>
<tr>
<td>Coenzyme transport and metabolism</td>
<td>Cobalamin biosynthesis</td>
<td>4.47E-03</td>
<td>NS</td>
</tr>
</tbody>
</table>

The gluconeogenic enzyme-coding genes exhibited the same state-change down-regulation dynamics as ppsA in the Δura3 strain, with the exception of enolase (eno), which is not a direct TrmB target (Schmid et al., 2009). These dynamics were not observed in the Δura3ΔtrmB strain (Figure 4A, Table 1). Because state-change dynamics in gluconeogenic enzyme coding genes were greatly diminished in the Δura3ΔtrmB strain, it is likely that these genes are predominantly dependent on TrmB. As expected,
the expression profiles of enzyme-coding genes in gluconeogenesis clustered in the
\( \Delta ura3 \) parent strain but not in the \( \Delta ura3 \Delta trmB \) strain (Table 1).

In contrast to the state-change dynamics (i.e. monotonic increases or decreases
and a statistically significant change in equilibrium gene expression) observed in the
glycolytic and gluconeogenic pathways, impulse-like dynamics (transient increases or
decreases in gene expression) were observed in the genes coding for enzymes in purine,
cobalamin, and amino acid biosynthesis in the \( \Delta ura3 \) strain. The variety of impulse-like
dynamics made metabolic modularity especially apparent in these clusters. The pattern
of gene expression in the purine biosynthesis cluster was distinct in both the \( \Delta ura3 \) and
\( \Delta ura3 \Delta trmB \) strains (Figure 4B and 2.4G, Table 1). Expression of these genes displayed a
transient increase in mRNA immediately upon the addition of glucose in both
backgrounds (Figure 4B and 2.4G). Although the impulse-like dynamics observed in
purine biosynthesis genes are not TrmB-dependent, the difference in overall mRNA
concentration compared to the parent strain at the start of the time-course but not at the
end suggests that TrmB is nonetheless important in controlling these genes.

Expression patterns in genes coding for enzymes involved in cobalamin (vitamin
B-12) biosynthesis also differed between the \( \Delta ura3 \Delta trmB \) mutant and \( \Delta ura3 \) parent
strains. In the \( \Delta ura3 \) parent strain, these genes were down-regulated as soon as glucose
was added. This was followed by a transient up-regulation centered 45 minutes after the
addition of glucose (Figure 4E). These distinct expression patterns were tightly
correlated and clustered together in the Δura3 strain (Figure 4E). In contrast, in the Δura3ΔtrmB background, these genes were slowly and constantly up-regulated following the addition of glucose without any rapid changes in expression level (Supplementary Figure 1). There was no significant clustering of these expression profiles in the Δura3ΔtrmB knockout (Table 1). Together, these data suggest that cobalamin biosynthesis is predominantly TrmB regulated but that other factors may be involved.

In summary, the surprising diversity of observed dynamic gene expression patterns suggests that TrmB is required for temporal coordination of gene expression across metabolism in response to glucose. These patterns implicate other, unknown regulatory factors in this process. The expression of genes coding for enzymes at the core of central metabolism appears to be predominantly TrmB regulated, while branching cofactor pathways seem to require additional, as of yet unidentified regulators.
Figure 4: TrmB is required for metabolic modularity in *H. salinarum* metabolism. Figure depicts k-means clustering and using eight clusters and Pearson correlation as a scoring metrics on the Δura3 (C) and Δura3ΔtrmB (F) data. Only clusters enriched ($P < 0.01$) for COG biological function categories are shown for each strain. Four clusters are enriched in genes involved in specific pathways in Δura3 (A,
B, D and E), whereas only two clusters are thus enriched in Δura3ΔtrmB (G and H). Metabolic maps (C and F) depict enzyme-coding genes as colored squares, with colors corresponding to cluster graphs. Each cluster graph depicts gene expression data for individual genes (thinner lines) and the mean expression profile for the cluster (thicker lines). The number of genes in every cluster that shows significant ($P < 0.05$) 1.5-fold over expression by the one-sided Welch $t$-test is listed in the lower right-hand corner of each expression graph. Metabolite abbreviations used in (C and F) are listed in Supplementary Table S7. In all, 30 of the 100 genes analyzed by NanoString are not represented on the metabolic map for the sake of clarity.

### 2.4.3 TrmB promoter occupancy can explain gene expression dynamics for ppsA

In order to determine the specific contribution of TrmB-promoter binding to gene expression dynamics, we assayed TrmB enrichment over time in response to glucose at the ppsA promoter using chromatin immunoprecipitation followed by qPCR (ChIP-qPCR, Methods). Binding was assayed from 60 minutes before glucose addition through 180 minutes after addition. The same timepoints used for gene expression measurement were sampled with 5 additional time points for increased resolution. During mid-logarithmic growth on amino acids as a sole source of carbon and energy, TrmB showed significant binding enrichment (Figure 5A). Upon the addition of glucose, TrmB dissociated from the ppsA promoter within two minutes, and binding dynamics exhibited slight but reproducible damped oscillations before reaching a new steady state level (Figure 5A). In order to quantitatively assess the role played by TrmB in the gene expression profile of ppsA, we calculated a derived mRNA synthesis rate and compared it to promoter occupancy. We reasoned that a derived mRNA synthesis rate is
representative of actual promoter activity and deconvolves the effect of differing mRNA degradation rates in different genes.

Derivation of a mRNA synthesis rate from mRNA data requires the degradation constant ($K_{\text{deg}}$) for the gene of interest. For this we used genome-wide mRNA degradation rates that have been experimentally determined previously in Halobacterium salinarum (Hundt et al., 2007). The derived mRNA synthesis rate of ppsA inferred from the gene expression data (Figure 5B) correlated strongly with the ChIP-qPCR measurement of actual enrichment (Pearson correlation = 0.94). Taken together, these calculations suggest that TrmB is the primary regulator of ppsA mRNA synthesis.
Figure 5: (A) TrmB-binding enrichment at the \textit{ppsA} promoter is correlated with (B) predicted mRNA synthesis rate inferred from the gene expression data ('Methods' section) at the TrmB-dependent \textit{ppsA} promoter after the addition of 5% glucose. Error bars represent the standard error from the average of biological replicate experiments. See also Supplementary Figure S3.
2.4.4 Prediction of gene expression based on TrmB-promoter enrichment

TrmB dynamics at the *ppsA* promoter explained the temporal gene expression pattern observed for *ppsA* mRNA (Figure 5). This led us to ask to what extent TrmB-promoter binding dynamics alone could explain the temporal expression patterns of other genes in the TrmB regulon. In order to address this question, we modeled expression across the 100 genes in the NanoString dataset as a function of TrmB enrichment at the *ppsA* promoter over time. Since TrmB binds a *cis*-regulatory motif that is conserved throughout its regulon (TACT-N(7-8)-GAGTA) and its biochemical mechanism involves reduced affinity for DNA when sugar is present in the growth medium, we expect that TrmB binding will be qualitatively similar at all regulated promoters. This is supported by previous genome-wide ChIP-chip studies showing that TrmB bound its regulon only in the absence of glucose or glycerol (Schmid *et al.*, 2009). We therefore assumed for computational modeling purposes that TrmB dynamics at the *ppsA* promoter were representative of those at other promoters in the genome. *ppsA* promoter binding dynamics (Figure 5) were used as an input to an ordinary differential equation (ODE) model of gene expression with two parameters: a basal synthesis rate ($K_{basal}$) and a scaling term ($K_{eff}$) (Methods). For 62 genes, this approach adequately explained the dynamics of gene expression and predicted the level of gene expression in the Δura3ΔtrmB mutant strain (Figure 6A and 2.6B, Supplementary Figure 1). For other genes, this method explained certain aspects of the dynamics, but not all. For yet other
genes, the model was unable to explain the changes in gene expression (Figure 6C and 2.6D). The modeling and prediction for each of these groups is described in turn below.

Figure 6: ODE model fit to NanoString gene expression data in response to glucose is predictive of gene expression in the Δura3 (parent strain) and Δura3ΔtrmB mutant background for some but not all genes. (A) Phosphoglycerate mutase (gpm). (B) Pyruvate kinase (pykA). (C) Cobalamin biosynthesis gene (cbiG). (D) 5-
phosphoribosylglycinamide (GAR) synthetase (purD, purine biosynthesis pathway). Black lines represent mRNA level in the Δura3 strain. Red lines represent mRNA level in the ΔtrmB strain. The gray line and the dotted orange line show model fits for Δura3 and Δura3ΔtrmB, respectively. Error bars depict standard error from the average of two biological replicates of the gene expression data. Model fits to data for the remaining 96 genes are exhibited in Supplementary Figure S1.

In order to characterize both the role and the temporal dynamics of additional regulators working with TrmB over time in response to glucose, we calculated the difference in gene synthesis rate between the model prediction and the observed data. We then clustered the synthesis rate residuals (actual minus predicted) in the Δura3 strain for each gene over the glucose response time course using k-means. This approach classified the accuracy of model fits to the data into five different groups (Figure 7). All five clusters were highly enriched for specific COG functions (Table 2). In two of the five clusters (Cluster 4&5, Figure 7D), containing 62 of 100 genes, no pattern was found in the residuals. This suggests that these genes exhibited either little change over the time course, or that expression predictions from the TrmB-enrichment model were accurate (Figure 7D). As expected, all three of the no-change control genes were members of these two clusters. Further, carbohydrate metabolism genes were enriched in one of these clusters: they are predominantly TrmB regulated and were therefore accurately predicted (Figures 2.5 and 2.6).
Figure 7: The difference between model-predicted and actual gene synthesis rates in $\Delta ura3$ (black) and $\Delta ura3\Delta trmB$ (gray) in clusters over time in response to glucose. (A) Cluster 1, enriched in genes coding for enzymes involved in purine biosynthesis; (B) Cluster 2, enriched in genes coding for enzymes involved in amino acid metabolism; (C) Cluster 3, enriched in genes coding for enzymes involved in cobalamin biosynthesis; (D) Clusters 4 and 5, containing genes whose expression is well-fitted by our model. The error bars represent the standard error from the average of the model fit residuals in each cluster.
Table 2: P-values of enrichment of COG terms in k-means clusters of Δura3 model residuals

<table>
<thead>
<tr>
<th>Cluster number</th>
<th>COG term</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>Nucleotide transport and metabolism</td>
<td>6.30E-03</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Amino acid transport and metabolism</td>
<td>3.19E-02</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Coenzyme transport and metabolism</td>
<td>1.39E-06</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>Carbohydrate transport and metabolism</td>
<td>7.24E-03</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>Energy production and conversion</td>
<td>4.73E-02</td>
</tr>
</tbody>
</table>

The three remaining clusters (Clusters 1, 2, 3; Figure 7 A, B, C) showed significant dynamic patterns in their model residuals between 0 and 180 minutes after perturbation with glucose. This indicates that the model prediction deviated significantly from the observed data over certain time points. Further, this deviation suggested that other regulators besides TrmB may be involved.

2.4.5 Inferring regulatory network topology from dynamic gene expression output.

To identify how such regulators may be involved, we reversed biological circuit design principles (Alon, 2007a) to infer GRN structure from dynamical output. Patterns of regulation that could not be predicted from TrmB-promoter binding data were analyzed using logic approximations. We applied this to clusters 1 and 3, which were enriched for functions in purine and vitamin B-12 biosynthesis, respectively (Figure 7, Table 2).
The $\Delta ura3$ residuals in cluster 3 (enriched for genes encoding cobalamin biosynthesis proteins) exhibited a peak centered at 20-45 minutes (Figure 7C, Table 2). Consistent with the gene expression data, the spike was absent in the $\Delta ura3\Delta trmB$ strain. The global time delay from changes in transcript to changes in protein level following perturbation in $H.\ salinarum$ has been estimated at ~16 minutes based on parallel mRNA and proteomics time course data (Schmid et al., 2007). The similarity between global transcription-translation time lag and the lag between glucose addition and the peak observed in cobalamin biosynthesis gene residuals at 20-45 minutes suggest that TrmB and a second TrmB-dependent regulator may be involved in a feed forward loop (FFL) to regulate cobalamin biosynthesis. Since we observed an immediate decrease in gene expression after TrmB dissociated from the DNA, we reasoned that either TrmB or the FFL regulator could activate cobalamin gene expression. This represents an OR logic gate in the FFL. To test whether our data were consistent with such a feed-forward regulatory topology, we simulated the proposed network using logic approximations (Alon, 2007b, Figure 8A). We then fitted the parameters of the simulation to our cobalamin biosynthesis ($cob / cbi$) gene expression data over the first 60 minutes following glucose addition (Methods). The fitted degradation rate of the FFL effector was 0.034*min$^{-1}$. This indicates a response time of ~20 minutes, which is consistent with the 16 minute lag between transcription and translation (Schmid et al., 2007), and supports the transcription-translation FFL hypothesis (Figure 8B). From this model, the
degradation constant for the \textit{cob} / \textit{cbi} mRNA was estimated at \(~5.5\) minutes, which is similar to the average empirically determined half-life of the \textit{cob} / \textit{cbi} mRNA (8.1 minutes, Hundt \textit{et al.}, 2007). Furthermore, the dynamic profiles of these genes make it clear that TrmB is acting as an activator of \textit{cob} / \textit{cbi} genes, and that the second regulator is acting as an activator as well.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure8}
\caption{(A) An FFL is the most likely topology for the regulation of cobalamin synthesis. Model of proposed FFL involved in cobalamin biosynthesis is shown. (B) Fit of regulatory logic simulation with FFL topology and gene expression time course data. The thick gray line represents the best model fit to the \textit{cob/cbi} average expression (black line). The dotted gray lines represent the scaled expression profiles of individual genes.}
\end{figure}

In contrast to the lagged residuals of the cobalamin biosynthesis cluster, the residuals of Cluster 2, enriched in arginine and serine metabolism genes, exhibited an immediately changing dynamic residual profile in the \textit{\Delta ura3} parent strain (Figure 7B, Figure 8).
Table 2). Model residuals increase immediately upon glucose addition in the $\Delta ura3$ parent strain. This suggests that TrmB is not a primary regulator of these genes. In contrast to the previous cluster, because of the diversity of amino acid metabolic pathways and the likelihood of multiple independent regulators, further understanding of network topology of this cluster will require additional studies.

The residuals of Cluster 1, enriched for functions in purine biosynthesis (Figure 7A, Table 2), increased immediately following glucose addition. Furthermore, the residuals were similar in both $\Delta ura3$ and $\Delta ura3\Delta trmB$. Since the response time through a transcription factor would be on the order of the half-life of this factor (~10 min, Alon, 2007a), these data implicate a secondary regulator already present in the cytoplasm. Although TrmB may ultimately control the concentration of such a secondary regulator (Supplementary Figure 2), initial dynamics are likely TrmB-independent. Since the input signals to this secondary regulator are unknown, this network is not amenable to a logic simulation. Nevertheless, these data support the hypothesis that a secondary regulator may work through an independent pathway to impinge upon the purine biosynthesis genes coordinately with TrmB (Figure 6D).

In summary, predictions from ODE and logic models suggest that TrmB is responsible, in various capacities, for the temporal regulation of metabolism as well as the overall gene expression levels. We conclude that integrating temporal gene expression data in response to environmental and genetic perturbation assists in the
clarification of cause-effect relationships and prediction of the GRN topology that
regulates metabolic pathways.

2.5 Discussion

In this study, we integrated dynamic gene expression data with ChIP-qPCR
measurement of promoter binding site occupancy over time using a two-parameter ODE
model in order to better understand the regulatory causality and topology of the glucose
responsive network in *H. salinarum*. We found that TrmB operates directly, indirectly,
and in conjunction with other factors to regulate gene expression levels and temporal
dynamics in response to glucose.

The cobalamin biosynthesis pathway appears to be regulated transcriptionally
both by TrmB and a TrmB-dependent secondary regulator. Using logic approximations,
we show that FFL topology is consistent with our data (Figure 8B). Previous ChIP-chip
data also suggests that TrmB binds the promoters of four transcription factors:
VNG0156C, VNG0247C, VNG0878G and VNG1179C (Schmid et al., 2009). The TrmB cis-
regulatory motif has been identified in the upstream regions of several more
transcription factors, including VNG1899G. To determine which of these TrmB-
dependent TFs are the most likely candidates for the secondary regulator, we explored
the Environmental and Gene Regulatory Influence Network (EGRIN). EGRIN is a
computationally inferred gene regulatory network for *Halobacterium salinarum* learned
from gene expression data across a range of genetic and environmental conditions (Bonneau et al., 2007). Of the directly TrmB regulated and potentially TrmB regulated TFs, we predict from EGRIN that VNG0156C, VNG1899G, or VNG1179C are more likely candidates for the regulation of cobalamin biosynthesis.

Alternatively, feed-forward topology could be achieved via feedback from a secondary metabolite, such as adenosylcobalamin (vitamin B-12), rather than direct regulation by another transcription factor. For example, in Salmonella typhimurium B-12 biosynthetic enzymes are regulated by riboswitches. These 5’ untranslated region (UTR) mRNA elements bind to B-12 with high affinity to control translation of target mRNAs (Nahvi et al., 2004).

It is also possible that a more complicated regulatory topology underlies the dynamics we observed. For instance, in Salmonella typhimurium, transcriptional control of B-12 biosynthesis involves the interaction of several general and specific transcription factors. Genes coding for enzymes in B-12 synthesis appear to be transcriptionally activated by the relevant global transcription factor, Crp (via cAMP), under aerobic conditions and ArcA/ArcB under anaerobic conditions. Besides affecting the transcription of the cob / cbi genes, the condition-appropriate global regulator also activates PocR. PocR binds DNA and further activates cobalamin synthesis and propanediol metabolism (the pathway for which cobalamin is needed, Ailion et al., 1993).
Despite the potential for complicated control systems, we favor the transcriptional FFL network proposed for cob / cbi regulation. Riboswitch motifs have been identified in numerous bacteria, but few archaea (Rodionov et al., 2003). Computational analysis of the H. salinarum transcriptome reveals little evidence for extensive 5’ UTRs (Koide et al., 2009) or riboswitch motifs. Furthermore, the timing of the feed forward dynamics is consistent with a transcriptional feed forward system rather than with more rapid posttranscriptional regulation. Although it remains formally possible that other more complex regulatory topologies involving TrmB are acting on the B-12 biosynthesis cluster, our data most parsimoniously support the less complex FFL motif model under the conditions tested here.

The FFL motif is widely distributed in GRNs across the bacterial and eukaryotic domains. In E. coli, FFL motifs are statistically overrepresented in network analyses (Shen-Orr et al., 2002). For example, the global TF Crp and the specific TF AraC control the arabinose utilization operon in a FFL motif with AND logic (Shen-Orr et al., 2002), i.e. to transcribe the arabinose utilization genes, the lack of glucose AND the presence of arabinose must be sensed. FFL motifs are similarly overrepresented in yeast (Jothi et al., 2009). Although relatively few putative FFLs have been specifically identified in archaea (Schmid et al., 2011), many of the same evolutionary drivers that have led to feed forward motifs in other genomes may also be relevant (Cordero & Hogeweg, 2006).
Purine biosynthesis, on the other hand, appears to be co-regulated by a non-TrmB-dependent transcription factor in *H. salinarum* (Figure 4). The EGRIN network suggests several other regulators may be in play, including VNG5009H, VNG2614H, and VNG2163H; however, none of these are directly TrmB regulated. TrmB is required for the appropriate equilibrium level of these genes, suggesting that it plays at least a partial role in their regulation (Supplementary Figure 2). In *E. coli* and *Bacillus subtilis*, for example, purine biosynthesis is transcriptionally regulated by the PurR repressor in response to the small molecule hypoxanthine (Cho *et al.*, 2011, Rappu *et al.*, 1999). Hypoxanthine integrates exogenous purine availability, salvage, and de novo biosynthesis to maintain cellular purine levels. Although we could not identify a PurR homologue in *Halobacterium salinarum* on the basis of sequence, an analogous topology seems likely based on our analysis.

By integrating dynamic measurements of gene expression with promoter occupancy, this study has provided insight into TrmB-mediated transcriptional regulation of metabolism. Previous studies on TrmB in archaea have focused on steady state gene expression and promoter binding site occupancy to determine network structure and infer regulatory relationships (Schmid *et al.*, 2009, Kanai *et al.*, 2007). Under steady state analysis, it is difficult to discern the role of TFs in activating or repressing genes because the signal from primary TFs is frequently confused by secondary regulation. For example, it is difficult to infer the sign of edges in a FFL (activation or
repression), since the equilibrium level of gene expression output is affected by both input edges. We have demonstrated that adding a temporal dimension to gene expression measurement during metabolic adjustment deconvolves such regulatory relationships. Using a network inference approach, we hypothesize a FFL regulatory topology from the signal at observable nodes. By capturing the temporal separation of the activity of primary and secondary regulators, we lay the groundwork for establishing regulatory causality. This method of dynamic measurement of gene expression in response to genetic and environmental perturbation may be a generally feasible method for reconstructing GRNs in other organisms across all three domains of life.

2.6 Supplementary Data

Supplementary Data are available at NAR online (http://nar.oxfordjournals.org/content/suppl/2013/07/07/gkt659.DC1): Supplementary Tables 1-7, Supplementary Figures 1-3. NanoString data can be accessed under GEO series GSE47201 and platform GPL17188.

2.7 Acknowledgement

We thank Samuel Ramirez for the primer design and initial validation of RT-qPCR. We extend special thanks to Fred Nijhout for helpful discussions, Paul Magwene for
assistance with the logic simulations, and Nick Buchler for comments on the modeling and manuscript.
3. Dynamic metabolite profiling in an archaeon connects transcriptional regulation to metabolic consequences.

3.1 Abstract

In previous work, we showed that the TrmB transcription factor is responsible for regulating the expression of many enzyme-coding genes in the hypersaline-adapted archaeon *Halobacterium salinarum* via a direct interaction with a *cis*-regulatory sequence in their promoter. This interaction is abolished in the presence of glucose. Although much is known about the effects of TrmB at the transcriptional level, it remains unclear whether and to what extent changes in mRNA levels directly affect metabolite levels. In order to address this question, here we performed a high-resolution metabolite profiling time course during a change in nutrients using a combination of targeted and untargeted methods in wildtype and ΔtrmB strain backgrounds. We found that TrmB-mediated transcriptional changes resulted in widespread and significant changes to metabolite levels across the metabolic network. Additionally, the mis-regulation of gluconeogenesis in the ΔtrmB mutant strain in the absence of glucose observed previously results in low PRPP levels. These low PRPP levels are associated with a metabolic block in *de novo* purine synthesis, which is partially responsible for the growth defect of the ΔtrmB mutant strain in the absence of glucose. In conclusion, we show how

---

2 This chapter is in preparation for submission to a peer reviewed journal. Please look for an updated version in PubMed.
transcriptional regulation of metabolism affects metabolite levels and ultimately, phenotypes.

3.2 Introduction

Recent studies have revealed a startling diversity of microbes in many environments. In order to understand their role in these environments, one must understand their metabolic abilities. Although many central metabolic processes are well conserved and have been thoroughly characterized, variations in these canonical pathways and peripheral metabolic processes have not been extensively studied. The availability of high-throughput untargeted metabolomics techniques such as LC/MS-MS has led to a better understanding of the metabolic abilities and preferences of different microbes. However, a complete understanding of metabolism and its regulation is still missing in many understudied organisms such as the archaee. Although archaee represent much of the microbial diversity in many environments, relatively little is known about their metabolism and its regulation. Work in several archaeeal species have suggested that transcriptional regulation plays a central role in controlling metabolism (Siebers et al., 2005). However, a global analysis of the effects of transcription on metabolism has not yet been performed. Such an analysis can provide a deeper understanding of how cells respond to nutrient perturbations and how these changes affect the physiology of the organism.
Previous work in the hypersaline-adapted model archaeal species, *Halobacterium salinarum*, identified a winged helix-turn-helix (wHTH) transcription factor, TrmB, as a central regulator of carbon metabolism. TrmB was found to bind to the promoters of 113 genes in the absence of glucose. Many of these genes encode enzymes involved in diverse metabolic pathways such as gluconeogenesis, amino acid metabolic, cobalamin (B12) synthesis and purine synthesis. Despite sharing a common regulator, dynamic gene expression profiling suggests that these pathways are differentially regulated by TrmB and other regulators, which raises the possibility that TrmB is a global metabolic regulator. When bound to the genome, TrmB activates the expression of some genes and represses the expression of others (Schmid et al., 2009). The addition of glucose abolishes TrmB-DNA binding, likely via a direct interaction with TrmB which causes a conformational change (Schmid et al., 2009, Krug et al., 2013). The abolition of TrmB binding results in de-repression of repressed genes and de-activation of previously activated genes. A ΔtrmB mutant strain has been generated, which exhibits altered gene expression, impaired growth, and other abnormal physiology in the absence of glucose but has wildtype growth rate and physiology when glucose is added to the media (Schmid et al., 2009). Interestingly, *H. salinarum* appears to be unable to metabolize glucose for energy, suggesting that TrmB responds primarily to endogenously produced glucose. A primary use of glucose in *H. salinarum* appears to be glycosylation of the S-layer glycoprotein (Todor et al., 2014). Since the amount of S-layer produced is
proportional to the instantaneous growth rate, TrmB activity may also link cell growth rate and metabolism (Todor et al., 2014).

Although the gene expression changes caused by TrmB and their evolutionary rationale have been explored (Schmid et al., 2009, Todor et al., 2013, Todor et al., 2014), it remains unclear whether TrmB affects metabolite levels and, if so, how these changes are connected to the growth defect observed in the ΔtrmB mutant strain under gluconeogenic conditions. Additionally, many TrmB-regulated pathways, including purine metabolism and aromatic amino acid synthesis, and how these pathways connect to central metabolism, remain only partially understood in H. salinarum and many other halophiles. For example, although many of the enzymes involved in de novo purine synthesis have been identified by homology to characterized enzymes (Brown et al., 2011), the pentose phosphate pathway that provides the precursor to purine synthesis remains unknown (Soderberg, 2004). For these reasons, here we conducted metabolomics profiling experiments in the ΔtrmB mutant strain and the Δura3 isogenic parent strain over a time course of 10 hours, which included a glucose stimulus. A combination of targeted and untargeted metabolomic approaches revealed that TrmB-mediated changes in gene expression are associated with significant shifts in metabolite levels across the metabolic network, and that de novo purine synthesis is inhibited due to a lack of phosphoribosylpyrophosphate (PRPP).
3.3 Methods

3.3.1 Strains

*Halobacterium salinarum* NRC-1 (ATCC strain 700922) was used as the wild-type strain background. Experiments were performed in two previously constructed strains, one containing an in-frame deletion of $\text{VNG1451C} \ (\Delta\text{ura3}\Delta\text{trmB}; \text{Schmid et al., 2009})$ and its isogenic parent strain, $\Delta\text{ura3} \ (\text{Peck et al., 2000})$.

3.3.2 Growth conditions and growth rate calculation

Cells were grown routinely in Complete Defined Medium (CDM) containing 19 amino acids (Todor et al., 2013) as indicated in the figures. The medium was supplemented with 50 µg ml$^{-1}$ uracil to complement the $\text{ura3}$ deletion. Cultures were routinely grown at 42°C while shaking at 225 r.p.m. under low ambient light.

In order to assess instantaneous growth rate during glucose addition, 200 µl cultures of the $\Delta\text{trmB}$ mutant strain and its isogenic parent strain $\Delta\text{ura3}$ were grown at 42°C under continuous shaking (~225 r.p.m.) in a Bioscreen C automated growth curve analysis system (Growth Curves USA, Piscataway, NJ) to mid-logarithmic phase. Glucose, sucrose, or nothing was added to the wells once mid-logarithmic phase was reached. Optical density at 600nm was measured every 10 minutes for each sample during the duration of the experiment. Instantaneous growth rate was calculated from
the log transformed and LOWESS-smoothed data using the bsd Analysis Function previously described (Todor et al., 2014). Instantaneous growth rate in the cultures to which glucose was added was calculated by adding the difference of the glucose and sucrose instantaneous growth rates to the control (+nothing) growth rate in order to account for osmotic effects.

To measure the growth of the ΔtrmB mutant strain and its isogenic parent strain Δura3 in CDM supplemented with 100µM concentration of various purines, 200 µl cultures were grown as described above in a Bioscreen C. Optical density at 600nm was measured every 30 minutes for up to 200 culture samples simultaneously. Area under the log-transformed growth curve (growth potential) was used as the growth metric to convey information about both growth rate and carrying capacity of the culture. These calculations are packaged into the bsd Analysis Function previously described (Todor et al., 2014).

### 3.3.3 Collection and quantification of metabolomics time course samples

Cells were grown to mid-logarithmic phase (OD ~ 0.3) in CDM as described above. For metabolomics time courses, 10 ml aliquots were removed from the continuously shaking cultures at time points as indicated in Table 3. Cells were immediately filtered using a Millipore SteriFlip filter, and washed with 10 ml basal salts media. Since H. salinarum lyses readily in methanol and other solvents because of its
requirement for high salinity, cells were lysed on the membrane using 1 ml of either ice-cold 80% methanol with extraction standards, ice-cold 100% methanol with extraction standards, or 1ml 50% acetonitrile with 0.3% Formic acid as indicated (Table 3). Samples were immediately snap frozen at -80°C and analyzed at Metabolon Corp. (Durham, NC) or the Stedman Center for Nutrition and Metabolism (Durham, NC) using previously validated methods (Table 3). For the Metabolon samples, the first minute of the LC column was diverted to avoid salt buildup in the instrument.

Table 3: Summary of metabolomics experiments

<table>
<thead>
<tr>
<th>Method</th>
<th>Strains Analyzed</th>
<th>Time Points Analyzed</th>
<th>Extraction Solvent</th>
<th>Analysis Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Targeted OA and AA Panel</td>
<td>Δura3, ΔtrmB</td>
<td>-60,0,5,10,20,45,90,180,360</td>
<td>50% Acetonitrile + 0.3% Formic Acid</td>
<td>Ferrara et al., 2008</td>
</tr>
<tr>
<td>Untargeted</td>
<td>Δura3, ΔtrmB</td>
<td>-240, -60,0,5,10,20,45,90,180,360</td>
<td>80% Methanol</td>
<td>Albrecht et al., 2014</td>
</tr>
<tr>
<td>Purine Panel</td>
<td>+Glucose, -Glucose</td>
<td></td>
<td>100% Methanol</td>
<td>Sinha et al., 2014</td>
</tr>
</tbody>
</table>

3.3.4 Normalization, outlier removal, and clustering of Metabolon data

Each sample from the Metabolon (untargeted GC/LC-MS) data was normalized by the coefficients of a linear regression between its log-transformed values and the mean of all log-transformed samples. Outlier removal was performed after normalization using Dixon’s test with a p-value cutoff of 0.05 for each metabolite in each
strain at each time point. Remaining data were averaged for each time point, strain and metabolite. For a given time point, when a metabolite was not detected in any of the 5 replicates, the minimum value of the metabolite from the remaining data was used as the value for that time point. The log-transformed scaled and centered metabolite measurements were clustered by hierarchical clustering with 6 clusters using the complete linkage method.

3.3.5 Analysis of purine metabolites and purine genes

The log-transformed, mean-centered and normalized time course gene expression data (Todor et al., 2013) for genes encoding enzymes involved in purine synthesis (prsA, purB, purC, purD, purE, purF, purH, purK, purL2, purM, purU, VNG2371C) were averaged at each time point during the glucose addition time course in the ΔtrmB mutant strain and Δura3 parent strain (n=20). These measurements were correlated using a linear model to the averaged log-transformed mean-centered and normalized data for guanosine and adenosine measured at the same time points.
3.4 Results and Discussion

3.4.1 TrmB-driven transcriptional regulation directly and specifically affects the levels of certain metabolites.

Many metabolic processes are regulated both transcriptionally and post-transcriptionally in *H. salinarum*. Therefore, in order to determine whether and how TrmB-mediated transcriptional regulation of enzyme-coding genes affects metabolite levels, we performed untargeted GC/LC-MS metabolomics analysis over a high-resolution time course in a strain harboring an in-frame deletion of *trmB* (Δ*trmB*) and its isogenic parent strain (Δ*ura3*). The Δ*trmB* mutant strain and the Δ*ura3* isogenic parent strain were grown on amino acids as the source of carbon and energy to mid-logarithmic phase (OD600 0.2-0.5) in complete defined medium (CDM), then sampled three times before and seven times after the addition of glucose (Methods). In order to confirm that these data are reproducible and sensitive, we compared the untargeted metabolomics data to data from targeted GC-MS and LC-MS assays for organic acids and amino acids performed on cells grown under identical conditions (Methods, Ferrara *et al.*, 2008). We found a strong correlation between metabolite levels obtained with the two methods (Figure 9a, $R^2 > 0.5$, $p < 10^{-16}$). To determine the detection limit of our untargeted assay, we converted the relative concentrations returned by the untargeted assay to absolute concentrations based on the regression with the targeted organic acid and amino acid data. We were able to consistently detect the presence of compounds with intracellular
concentrations in the single milli-molar range (Figure 9b). Taken together, these data suggest that metabolomics profiling in *H. salinarum* is a robust and sensitive method for assessing intracellular metabolite concentrations (Figure 9).

Table 4: Number of Compounds detected in different metabolite classes.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unidentified</td>
<td>34</td>
</tr>
<tr>
<td>Amino acid</td>
<td>34</td>
</tr>
<tr>
<td>Lipids</td>
<td>16</td>
</tr>
<tr>
<td>Carbohydrate</td>
<td>13</td>
</tr>
<tr>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>7</td>
</tr>
<tr>
<td>Xenobiotics</td>
<td>7</td>
</tr>
<tr>
<td>Nucleotide</td>
<td>6</td>
</tr>
<tr>
<td>Peptide</td>
<td>6</td>
</tr>
<tr>
<td>Secondary metabolism</td>
<td>2</td>
</tr>
</tbody>
</table>
Figure 9: (A) Metabolite levels measured by Metabolon are well-correlated with targeted measurements of amino acids and organic acids. Log-log linear regression of metabolites and time points in common is shown in red. (B) Metabolon’s untargeted assay can measure metabolites with levels in the millimolar intracellular range in both the Δura3 parent strain (dark bars) and the ΔtrmB mutant strain (red bars).
A total of 125 metabolites were detected in the untargeted samples. Of these, 91 were positively identified and classified into 8 pathways (Table 4). The temporal profiles of all 125 metabolites were grouped into 6 clusters using hierarchical clustering in order to determine their response to glucose perturbation and the dependence of this response on TrmB (Figure 10). We found that metabolites in specific pathways were significantly enriched in certain clusters (Table 5, Supplementary Table 1). Additionally, while some clusters showed similar patterns in the ΔtrmB knockout mutant and the Δura3 isogenic parent strain, many showed distinct patterns between strains (Figure 10, Supplementary Figure 1). Specifically, cluster 3, which is significantly enriched for nucleotides such as adenosine and guanosine \((p < 0.005)\) and cluster 4, which is significantly enriched in peptides \((p < 0.01)\) are very different in the ΔtrmB knockout mutant relative to the Δura3 isogenic parent strain before \((p < 1 \times 10^{-10})\), but not after \((p > 0.05)\) the addition of glucose. Together, these data suggest that TrmB, a transcription factor regulating metabolic enzyme-coding genes (Schmid et al., 2009, Todor et al., 2013), is required to maintain wild type metabolite levels in specific pathways, including purines and peptides, during nutrient perturbations.
Figure 10: Figure depicts hierarchical clustering using six clusters on the mean-scaled and normalized combined Δura3 (black lines) and ΔtrmB (red lines) data. Each cluster graph depicts metabolite data for individual metabolites (thinner lines) and the mean expression profile for the cluster (thicker lines).
### Table 5: p-values of enrichment of functional categories in clusters.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Pathway</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster 1</td>
<td>N/A</td>
<td>6.65E-06</td>
</tr>
<tr>
<td>Cluster 2</td>
<td>Carbohydrate</td>
<td>6.08E-09</td>
</tr>
<tr>
<td>Cluster 3</td>
<td>Xenobiotics</td>
<td>0.000901</td>
</tr>
<tr>
<td></td>
<td>Nucleotide</td>
<td>0.003562</td>
</tr>
<tr>
<td></td>
<td>Secondary metabolism</td>
<td>0</td>
</tr>
<tr>
<td>Cluster 4</td>
<td>Peptide</td>
<td>0.008786</td>
</tr>
<tr>
<td>Cluster 5</td>
<td>Lipids</td>
<td>0.00306</td>
</tr>
<tr>
<td></td>
<td>Peptide</td>
<td>0.004867</td>
</tr>
<tr>
<td>Cluster 6</td>
<td>Amino acid</td>
<td>1.90E-06</td>
</tr>
</tbody>
</table>

### 3.4.2 Temporal dynamics of growth recovery

In order to understand how changes in metabolite levels affect the physiology of *H. salinarum* before and after the addition of glucose, we sampled the optical density of the cultures of the ΔtrmB mutant strain and the Δura3 isogenic parent strain every 10 minutes before, during, and after the addition of glucose during mid-log growth in CDM (Methods). We calculated the instantaneous growth rate during each of these 10 minute periods in order to determine how growth rate changes after the addition of glucose. We found that the ΔtrmB mutant strain recovers a wildtype growth rate approximately ~200 minutes after glucose addition (Figure 11b). The addition of glucose to the Δura3 isogenic parent did not appear to cause a major change in growth rate (Figure 11a). In order to determine which metabolites may be linked to growth rate, we correlated the level of each metabolite at each time point with the instantaneous growth in both the ΔtrmB knockout mutant and the Δura3 isogenic parent strain. We found that
12 out of the 125 metabolites detected were significantly correlated with instantaneous growth rate after correcting for multiple hypothesis testing using the Bonferroni method (Table 6). Adenosine, guanosine, NAD\(^+\), and N-carbamoylaspartate were positively correlated with growth while N-acetylthreonine, tryptophanol, gamma-glutamyl-valine, gamma-glutamyl-alanine, and gamma-glutamyl-2-aminobutyrate were negatively correlated with instantaneous growth rate. The strong positive correlation between instantaneous growth rate and purine compounds such as adenosine, guanosine, and NAD\(^+\) led us to ask whether part of the growth defect of the $\Delta trmB$ knockout mutant strain may be due to a lack of purine synthesis, which may be due to TrmB transcriptional regulation of this pathway.

Table 6: Metabolites significantly correlated with growth rate in the $\Delta trmB$ mutant strain and the $\Delta ura3$ isogenic parent strain.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>Slope</th>
<th>$R^2$</th>
<th>$p$-Value</th>
<th>Corrected $p$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>X - 16071</td>
<td>-4.64E-07</td>
<td>0.5128</td>
<td>0.0004</td>
<td>0.048</td>
</tr>
<tr>
<td>X - 20525</td>
<td>7.87E-09</td>
<td>0.5379</td>
<td>0.0002</td>
<td>0.0292</td>
</tr>
<tr>
<td>X - 16682</td>
<td>-4.22E-08</td>
<td>0.5763</td>
<td>0.0001</td>
<td>0.013</td>
</tr>
<tr>
<td>NAD(^+)</td>
<td>4.55E-07</td>
<td>0.5915</td>
<td>7.00E-05</td>
<td>0.0092</td>
</tr>
<tr>
<td>N-acetylthreonine</td>
<td>-5.54E-07</td>
<td>0.5927</td>
<td>7.00E-05</td>
<td>0.009</td>
</tr>
<tr>
<td>tryptophanol</td>
<td>-2.02E-07</td>
<td>0.6294</td>
<td>3.00E-05</td>
<td>0.0038</td>
</tr>
<tr>
<td>gamma-glutamylvaline</td>
<td>-2.53E-07</td>
<td>0.6441</td>
<td>2.00E-05</td>
<td>0.0026</td>
</tr>
<tr>
<td>adenosine</td>
<td>7.93E-08</td>
<td>0.6444</td>
<td>2.00E-05</td>
<td>0.0026</td>
</tr>
<tr>
<td>guanosine</td>
<td>5.53E-06</td>
<td>0.6667</td>
<td>1.00E-05</td>
<td>0.0014</td>
</tr>
<tr>
<td>gamma-glutamylalanine</td>
<td>-5.55E-08</td>
<td>0.668</td>
<td>1.00E-05</td>
<td>0.0014</td>
</tr>
<tr>
<td>N-carbamoylaspartate</td>
<td>4.40E-07</td>
<td>0.6687</td>
<td>1.00E-05</td>
<td>0.0013</td>
</tr>
<tr>
<td>gamma-glutamyl-2-aminobutyrate</td>
<td>-2.66E-08</td>
<td>0.7107</td>
<td>3.00E-06</td>
<td>0.0004</td>
</tr>
</tbody>
</table>
Figure 11: Instantaneous growth rate of the Δura3 parent strain (A) and ΔtrmB mutant strain (B) during the glucose addition (red lines) and control (black lines) time course. Error bars represent the standard error of 3 biological replicates.
3.4.3 Purine synthesis is negatively regulated by purine levels

Since TrmB is a transcriptional regulator that has been shown to target the promoters of genes encoding enzymes involved in de novo purine synthesis (Schmid et al., 2009), we wondered whether the defect in purine levels we observed in the ΔtrmB mutant in the absence of glucose (Figure 10c) was due to mis-regulation of genes encoding purine synthesis enzymes due to the lack of TrmB. In order to address this question, we compared the expression level of genes encoding enzymes involved in purine biosynthesis (Todor et al., 2013) to the measured purine levels (Figure 12, Methods) in both the ΔtrmB mutant strain and its isogenic Δura3 parent strain at each time point. We found a strong negative correlation ($R^2 > 0.55, p < 10^{-3}$) between gene expression and purine level which is consistent across both strains and nutrient conditions. This suggests that the defect in purine levels in the ΔtrmB mutant strain in the absence of glucose is not due to transcriptional misregulation of de novo purine synthesis. Rather, the data is consistent with the hypothesis that the transcription of purine synthesis genes is negatively regulated by the availability of purines. Taken together, the elevated expression of genes coding for enzymes involved in purine biosynthesis and the lack of purines in the ΔtrmB strain in the absence of glucose suggest a metabolic block in de novo purine biosynthesis.
3.4.4 Defect in purine synthesis may be due to lack of PRPP

Since the ΔtrmB mutant strain has a severe growth defect in the absence of glucose which is rescued by the addition of glucose, and since purine levels are well correlated with instantaneous growth rate in the ΔtrmB knockout mutant and the Δura3 isogenic parent strain, we wondered whether the lack of sufficient de novo purine synthesis may be responsible for part of the growth defect of the ΔtrmB mutant strain. In
order to assess whether the growth defect of the ΔtrmB mutant strain in the absence of glucose is due to inhibited purine synthesis, we assayed the growth of the ΔtrmB mutant and its isogenic parent strain Δura3 in the presence and absence of various purines (Methods). Previous work has shown that purines are readily taken up by *H. salinarum* and integrated into biomass and nucleic acids (Stuer-Lauridsen & Nygaard, 1998). We found that adenosine significantly, but incompletely, complemented the growth defect of the ΔtrmB strain in the absence of glucose (*p* < 0.01, Figure 13, Table 7). Once taken up by the cell, adenosine is converted to inosine monophosphate (IMP). IMP can be converted to either GMP or AMP. These nucleotides can then be assimilated into RNA, phosphorylated to tri-nucleotides and otherwise function in the cell (Stuer-Lauridsen & Nygaard, 1998). Although adenosine partially complemented growth defect of the ΔtrmB strain in the absence of glucose, adenine did not (Figure 13, Table 7). Since the enzymatic activity of adenine phosphoribosyltransferase, which uses PRPP to phosphoribosylate adenine into AMP, has been detected in *H. salinarum* (Stuer-Lauridsen & Nygaard, 1998), these data suggest that there is insufficient PRPP present to phosphoribosylate the adenine into AMP and complement the growth defect. Guanosine was also unable to complement the growth defect of the ΔtrmB strain in the absence of glucose (Figure 13, Table 7) likely because *H. salinarum* lacks GMP reductase which is required to convert GMP from guanosine to AMP. Interestingly, supplementation with guanine causes no change in the growth of the Δura3 isogenic parent strain, but appears
to adversely affect the growth of the \( \Delta trmB \) knockout \((p < 4 \times 10^{-5}, \text{Figure 13, Table 7})\). This may be due to guanine phosphoribosyltransferase further depressing PRPP levels.

Taken together, these data suggest that PRPP levels are depressed in the \( \Delta trmB \) mutant in the absence of glucose. Since PRPP is also required in the absence of purine supplementation for the initial step of de novo purine biosynthesis, PRPP may be the cause of the defect in purine levels in the \( \Delta trmB \) mutant in the absence of glucose.

**Table 7: Bonferroni corrected p-values of the T-test of difference from CDM.**

<table>
<thead>
<tr>
<th>Condition</th>
<th>( \Delta ura3 )</th>
<th>( \Delta trmB )</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDM+Gluc</td>
<td>1</td>
<td>4.95E-10</td>
</tr>
<tr>
<td>CDM+Ade</td>
<td>0.170055</td>
<td>2.300766</td>
</tr>
<tr>
<td>CDM+AdeN</td>
<td>0.430323</td>
<td>0.007913</td>
</tr>
<tr>
<td>CDM+Gua</td>
<td>0.119496</td>
<td>3.96E-05</td>
</tr>
<tr>
<td>CDM+GuaN</td>
<td>0.006139</td>
<td>0.526128</td>
</tr>
</tbody>
</table>
Figure 13: The average growth potential (Methods) of the \( \Delta ura3 \) parent strain (dark bars) and \( \Delta trmB \) mutant strain (light bars) in CDM, CDM supplemented with 5% glucose (CDM+ Gluc), and CDM supplemented 100 \( \mu \)M of adenine (CDM+Ade), adenosine (CDM+AdeN), guanine (CDM+Gua), and guanosine (CDM+GuaN). Error bars represent the standard error of 7 biological replicates.

3.4.5 PRPP is found at a low level in \( \Delta trmB \) mutant strain in the absence of glucose

In order to determine if PRPP limitation is the cause of the growth defect of the \( \Delta trmB \) mutant growing in the absence of glucose, we assayed PRPP levels using a targeted GC-MS method (Sinha et al., 2014) in the \( \Delta trmB \) mutant strain and its \( \Delta ura3 \) isogenic parent strain in the presence and absence of glucose. We found that PRPP levels were significantly lower in the \( \Delta trmB \) mutant strain in the absence of glucose than in the
ΔtrmB mutant strain with glucose and Δura3 isogenic parent strain with or without glucose (Figure 14, Supplementary Table 2). Taken together with our other metabolomics data and with the growth complementation assay, this observation strongly suggests that the growth defect of the ΔtrmB mutant strain in the absence of glucose is partially caused by the lack of PRPP and therefore the lack of de novo purine synthesis.

Figure 14: PRPP levels in the Δura3 parent strain (dark bars) and the ΔtrmB mutant strain (light bars). Glucose complements the ΔtrmB mutant’s PRPP deficiency. Error bars represent the standard error of 8 biological replicates. Detection limit calculated from standard curve is shown as a dashed grey line.
3.4.6 The oxidative pentose phosphate pathway proceeds through gluconate

Despite the fact that previous research (Gochnauer & Kushner, 1969, L.O. Severina, 1990, Sonawat et al., 1990) suggests that *H. salinarum* is unable to metabolize glucose for energy, the growth defect of the Δ*trmB* mutant strain in the absence of glucose suggests that glucose is nonetheless required for cell growth. Previous research (Todor et al., 2014) suggests that one use of glucose is to decorate the S-layer glycoprotein in order to maintain cell shape. Since purine levels are low in the Δ*trmB* mutant strain in the absence of glucose due to the lack of PRPP, another use of glucose may be the production of PRPP. Despite the lack of a non-oxidative pentose phosphate pathway and the ability of *H. salinarum* to grow without added nucleotides, previous studies (Soderberg, 2004) have only identified two enzymes of the oxidative pentose phosphate pathway (OPPP): a putative 6-phosphogluconate dehydrogenase (*VNG2553G*) and a ribose phosphate pyrophosphokinase (*VNG2203G*). It has been hypothesized that *H. salinarum* uses a modified version of the OPPP, proceeding from glucose through gluconate to 6-phosphogluconate (and on to ribulose-5-phosphate and PRPP, Falb, *et al.*, 2008). However, experimental evidence for this pathway is lacking. In order to determine if the OPPP in *H. salinarum* proceeds through these compounds, we examined the temporal pattern of gluconate and adenosine in the PRPP-starved Δ*trmB* mutant strain after the addition of glucose (Figure 15). We found that gluconate began to
accumulate approximately 20 minutes after glucose addition. Adenosine levels, on the other hand, did not increase until after gluconate accumulation (after 45 minutes), consistent with the hypothesis that the OPPP proceeds through gluconate. To the best of our knowledge, this is the first experimental evidence for the operation of such a pathway.

![Log-transformed glucose (black line), gluconate (blue line), and adenosine (red line) levels in the ΔtrmB mutant strain are temporally separated](image)

Figure 15: Log-transformed glucose (black line), gluconate (blue line), and adenosine (red line) levels in the ΔtrmB mutant strain are temporally separated suggesting these metabolites are sequential steps of the OPPP and purine synthesis pathways.
3.6 Conclusions

Understanding the regulation of metabolism in archaea remains a central challenge to successfully tapping the vast reservoir of pathways and compounds in these organisms. Previous studies have identified TrmB as a central transcriptional regulator of metabolism in *H. salinarum*. In order to better understand how transcriptional regulation of enzyme coding genes affects metabolism, we performed targeted and untargeted measurements of the metabolome of *H. salinarum* during a change in nutrient conditions. We found that the transcriptional regulation causes substantial changes in intracellular metabolite levels, which cause changes in physiology. Specifically, we identified a metabolic block in de novo purine synthesis in the ΔtrmB mutant strain in the absence of glucose which is partially responsible for the observed growth rate defect. Combining dynamic metabolite and gene expression data, we showed that this block was due to a defect in PRPP caused by mis-regulation of genes encoding enzymes involved in gluconeogenesis and therefore a lack of OPPP substrates. Taken together, this data suggests that transcriptional regulation of enzyme coding genes is an important mechanism for regulating metabolism.

3.7 Supplementary Data

Supplementary Data is in Appendix A
4. A transcription factor links growth rate and metabolism in the hypersaline adapted archaeon *Halobacterium salinarum*.

4.1 Abstract

Coordinating metabolism and growth is a key challenge for all organisms. Despite fluctuating environments, cells must produce the same metabolic outputs to thrive. The mechanisms underlying this “growth homeostasis” are known in bacteria and eukaryotes, but remain unexplored in archaea. In the model archaeon *Halobacterium salinarum*, the transcription factor TrmB regulates enzyme-coding genes in diverse metabolic pathways in response to glucose. However, *H. salinarum* is thought not to catabolize glucose. To resolve this discrepancy, we demonstrate that TrmB regulates the gluconeogenic production of sugars incorporated into the cell surface S-layer glycoprotein. Additionally, we show that TrmB-DNA binding correlates with instantaneous growth rate, likely because S-layer glycosylation is proportional to growth. This suggests that TrmB transduces a growth rate signal to co-regulated metabolic pathways including amino acid, purine, and cobalamin biosynthesis.

---

Remarkably, the topology and function of this growth homeostastic network appear conserved across domains despite extensive alterations in its protein components.

4.2 Introduction

The coordination of growth with catabolic and anabolic metabolism is one of the most important and complicated tasks an organism faces (Brauer et al., 2008). Cells must regulate diverse processes in response to environmental conditions and nutrient availability so that the appropriate metabolic outputs are produced in the correct proportions for the cell to continue to grow, replicate, and maintain energetic balance. This balance, here referred to as growth homeostasis, must also be coordinated with cell cycle, shape, and division (Johnston et al., 1977). Because of its importance to all organisms, various aspects of growth homeostasis have been elucidated for several organisms in the bacterial and eukaryotic domains (Brauer et al., 2008, Weart et al., 2007). Despite the importance of growth homeostasis, the connection between cell growth and metabolism remains understudied in the archaea. Elucidation of the mechanisms for coordinating growth rate and metabolism in this third domain of life is necessary to provide a complete perspective on the evolution of growth homeostatic mechanisms in different environmental niches.

In previous work, we identified and characterized a transcription factor, TrmB, in the hypersaline archaeon Halobacterium salinarum. TrmB regulates genes coding for
enzymes involved in diverse metabolic pathways in response to glucose (Schmid et al., 2009). TrmB is a helix-turn-helix transcription factor whose amino acid sequence and cis-regulatory DNA binding motif are strongly conserved throughout the archaeal domain (Schmid et al., 2009, Lee et al., 2003, Lee et al., 2007, Perez-Rueda & Janga, 2010). This transcription factor has been shown to be involved in the regulation of maltose, trehalose, and glucose uptake and metabolism throughout the archaeal domain (Schmid et al., 2009, Lee et al., 2008, Kanai et al., 2007). In H. salinarum, TrmB binds to the promoters of over 100 genes that encode enzymes functioning throughout central and peripheral metabolism during growth on amino acids as the sole source of carbon and energy (i.e. in the absence of any added sugars; (Schmid et al., 2009). While bound to the DNA, TrmB represses some promoters (e.g. glycolytic enzyme-coding genes) and activates others (e.g. gluconeogenic genes). TrmB-DNA binding is inhibited by a direct interaction with glucose in hyperthermophilic archaea (Krug et al., 2006) and by glucose addition to the medium in H. salinarum. This leads to de-repression and de-activation of bound promoters. Through this mechanism, TrmB dynamically regulates the expression of enzyme coding genes and transcription factors in response to nutrient changes (Schmid et al., 2009, Todor et al., 2013).

Despite the importance of glucose in regulating central metabolism via TrmB in H. salinarum, previous studies suggest that H. salinarum may not be able to catabolize glucose. Both the encoding gene and the enzyme activity for phosphofructokinase are
undetectable in this organism (Sonawat et al., 1990, Ng et al., 2000). H. salinarum also cannot use glucose as the sole source of carbon and energy (Gochnauer & Kushner, 1969), and does not actively transport glucose from the media (L.O. Severina, 1990); although glucose can passively enter the cell. Previous label tracing NMR experiments have also failed to detect the conversion of 13C-labeled glucose to pyruvate (Sonawat et al., 1990). This evidence raises the question of why H. salinarum regulates much of its metabolism in response to glucose. Here we address this question by testing the hypothesis that one possible function of glucose is to serve as a precursor molecule to the N and O-linked glycans that decorate the cell surface glycoprotein.

In most archaea, the cell envelope consists of a paracrystalline S-layer protein (Mescher et al., 1974). The S-layer is solely responsible for maintaining the characteristic rod shape of H. salinarum (Mescher & Strominger, 1976). It consists of a single protein that comprises approximately 50% of the protein and most of the non-lipid carbohydrate of the cell envelope (Mescher et al., 1974). The S-layer protein of H. salinarum contains two different N-linked glycans, a repeating unit pentasaccharide (one site) and small sulfated oligosaccharides (10-15 sites), as well as a single O-linked disaccharide unit (10-15 sites at C-terminus; Wieland, 1988). These carbohydrate modifications are required for the stability and shape maintaining function of the S-layer in halophilic archaea related to H. salinarum. For example, strains deleted for genes encoding enzymes that
attach the glycoprotein to the S-layer (ΔalgB) exhibit rounded morphology and shed S-layer proteins into the medium (Abu-Qarn et al., 2007).

Here we present evidence that *H. salinarum* produces monosaccharides through gluconeogenesis that serve as precursors for the glycans that decorate the S-layer cell surface protein, enabling cells to maintain their characteristic rod shape during active growth. Additionally, we present metabolomics and gene expression data suggesting that TrmB is required for appropriate levels of these monosaccharides. TrmB activity is correlated with instantaneous growth rate across conditions and time scales, likely because growth requires additional S-layer and therefore additional glycans. We hypothesize that by sensing the concentration of a gluconeogenic metabolite, TrmB regulates genes encoding enzymes involved in diverse metabolic processes in response to cell growth rate.

### 4.3 Methods

#### 4.3.1 Strains and plasmids

*Halobacterium salinarum* NRC-1 (ATCC strain 700922) was used as the wild-type strain background. Experiments were performed in previously constructed strains, one containing an in-frame deletion of VNG1451C (Δura3ΔtrmB, Schmid et al., 2009) and its isogenic parent strain, Δura3 (Peck et al., 2000).
4.3.2 Growth Conditions

Cells were grown in either Complete Defined Medium (CDM) containing 19 amino acids (Todor et al., 2013) or rich complete medium (CM; 250 NaCl, 20 g/l MgSO4 7H2O, 3 g/l sodium citrate, 2 g/l KCl, 10 g/l bacteriological peptone) as indicated in the figures. Each medium formulation was supplemented with 50 μg ml-1 uracil to complement the *ura3* deletion. Cultures were routinely grown at 42°C while with shaking at 225 r.p.m. under low ambient light. In order to assess growth under different conditions, 200 μl cultures were grown at 42°C under continuous shaking (~225 r.p.m.) in a Bioscreen C automated growth curve analysis system (Growth Curves USA, Piscataway, NJ). Optical density at 600nm was measured every 30 minutes for up to 200 culture samples simultaneously. Maximum instantaneous growth rate was calculated from the log transformed and LOWESS smoothed data. Area under the log transformed growth curve (growth potential) was used as an additional growth metric to convey information about both growth rate and carrying capacity of the culture. These calculations are packaged into the bsd Analysis Function freely available at http://gaggle.systemsbiology.net/svn/gaggle/BioscreenUtilsR/trunk/bsdAnalysisFunc.R.
4.3.3 Cell lysis, S-layer enrichment, and staining protocol

S-layer glycosylation levels were quantified in whole cell extracts and in S-layer enrichments from each of the Δura3 and ΔtrmB strains. For preparation of whole cell extracts, H. salinarum cultures were grown to stationary phase (OD 600 ~ 1) in CM medium with or without glucose. Equivalent cell numbers from each strain were pelleted by centrifugation at 21,130 r.c.f. for 30 seconds and lysed in 500 µl of water followed by vigorous pipetting to lyse and homogenize the samples. Several aliquots of different volumes (50, 100, 200 µl) were precipitated using trichloroacetic acid and separated by SDS-PAGE as described below. S-layer enrichment was performed using a modified version of the protocol described in (Sumper et al., 1990). Briefly, H. salinarum cultures were grown to mid-log phase (OD 600 ~ 0.6) in CM medium with or without glucose. Approximately 10 ml of culture was pelleted at 4,500 r.c.f. for 10 minutes, resuspended in basal salts buffer (CM lacking peptone) containing EDTA equimolar to the Mg\(^{2+}\) concentration (80mM), and incubated for 30 minutes at 225 r.p.m at 42°C to remove the S-layer. Cells were removed by centrifugation for 5 minutes at 12,000 r.c.f. The supernatant, which contains primarily S-layer protein, was collected. Proteins were precipitated using 10% TCA, solubilized in Laemmli buffer containing SDS, and separated on two identically loaded BioRad Mini Protean TGX gels for 1 hr at 125V (BioRad Laboratories, Hercules, CA). One gel was stained to visualize protein using Colloidal Coomassie Blue (CCB, Invitrogen, Carlsbad, CA), and the other was stained to
visualize carbohydrate using Periodic Acid Schiff (PAS, Sigma, St. Louis, MO). After staining, gels were scanned using an Epson Perfection V700 Photo scanner and the band corresponding to the S-layer glycoprotein, known to run at ~200 kD (Lechner & Wieland, 1989), was quantified on each stained gel using ImageJ software. The PAS/CCB ratio for each of Δura3 and ΔtrmB with and without glucose was normalized to the average ratio of Δura3 without glucose for every gel, and the results from at least 2 bands were averaged to calculate the final values.

4.3.4 Microscopy and image quantitation

Cells were mounted on agarose pads and imaged using a Zeiss Axio Scope.A1 (Carl Zeiss, Oberkochen, Germany) and a PixelLINK CCD (PixelLINK, Ottawa, Canada) camera. Images were captured using a 100x oil immersion objective under phase contrast illumination. Image analysis was performed in ImageJ. Briefly, each image was automatically thresholded using the IsoData iterative thresholding scheme (Velasco, 1980). Ellipses were fit to all shapes with circularity (minor/major axis) between 0.1-1 and area between 300 and 3000 pixels (equivalent to 0.78 µM2 – 7.8 µM2). The major and minor axes were used as the cell length and width, respectively. Significance of differences between the ratio of length/width of ΔtrmB vs Δura3 was calculated using the 2-tailed Welch’s t-test.
4.3.5 Collection and analysis of GC-MS samples

Cells were grown to mid-log phase (OD 600 ~ 0.3-0.5) in CDM with or without the addition of 5% glucose as described above. Each flask was split into 2x25 ml technical replicates, for a total of 3 biological replicates and 2 technical replicates. Each sample was pelleted by centrifugation for 10 minutes at 4,500 r.c.f., washed twice with BS, lysed in 200 µl of H2O, and centrifuged at 18,000 r.c.f. for 5 minutes. 1ml of methanol was then added to 100µl of lysate in order to precipitate macromolecules. The samples were then vortexed and and centrifuged at 18,000 r.c.f. for 5 minutes. The methanolic supernatant was then transferred to a clean tube, and 5 µl of 2.5 mg/ml myristic acid-D27 was added as an internal standard. Samples were then dried under forced nitrogen at 45°C. Spin-vacuum and azeotropic treatment under ethyl acetate were applied to allow complete drying of high-salt samples. Dried samples were derivatized by stepwise addition of (i) 100 µl of 20 mg ml\(^{-1}\) methoxyamine hydrochloride in pyridine (Sigma-Aldrich), and (ii) 300 µl MSTFA with 1% TMCS (Thermo Fisher Scientific, Waltham, WA) with 50°C incubations (stagnant) for 30 min following addition of each chemical. After derivatization, samples were cooled, mixed, pulse centrifuged, and the liquid phase was transferred to GC/MS vials. Non-targeted metabolomics via GC/MS and peak area calculation were performed as described in Scholtens et al. 2014. Raw peak areas (unlogged values) for all sugars were summed independently for each strain, condition and biological replicate. Each of these sums were divided by the sum of all identified
metabolites from the same sample and normalized to the mean ratio for the parent strain grown without glucose. The log transformed mean across biological replicates of these normalized values are reported in Fig. 4.4.

4.3.6 Analysis of microarray and growth data

To infer TrmB-DNA binding activity across a range of conditions, we adapted the concept of a gene expression proxy from work in *S. cerevisiae* (Airoldi et al. 2009). The proxy was defined as the average of gene expression of VNG0095G, VNG0330G, VNG0683C, VNG0684G, VNG1216G, VNG1887G (TrmB-activated genes in the hypothesized gluconeogenic pathway) and the reciprocal of VNG032G4 and VNG0937G (TrmB-repressed genes coding for enzymes that catalyze the reverse reaction of VNG0330G and VNG0095G, respectively). Gene expression values were obtained from published microarray data for *H. salinarum* throughout growth in batch culture (Facciotti *et al.*, 2010); GEO accession numbers: GSE14832, GSE14835 and GSE14836), from *H. salinarum* undergoing the transition from anaerobic to aerobic conditions (Schmid *et al.*, 2007); GEO accession number: GSE7559), from *H. salinarum* being exposed to H2O2 (Kaur *et al.*, 2010); GEO accession number: GSE17515), and from *H. salinarum* growing in media of different salinity (Beer *et al.*, 2014); GEO accession number: GSE53544). Growth rate at each time point was calculated as the weighted average of the growth rate before and after each time point. The first 4 points of the oxygen time course were excluded.
due to optical density changes caused by the foaming of the medium during sparging of oxygen (Schmid et al., 2007). The batch culture growth data was binned according to optical density range. Figure 4.5a shows the bin averages and standard error of 8 independent time course experiments comprising 56 separate microarrays.

4.4 Results

4.4.1 The ΔtrmB mutant is round in the absence of glucose.

The ΔtrmB deletion mutant has been shown to be defective in gene regulation and exhibits significantly impaired growth rate in batch culture using amino acids as a source of carbon, nitrogen, and energy. These phenotypes are complemented upon the addition of glucose to the medium (Schmid et al., 2009, Todor et al., 2013). In order to determine the physiological effects of misregulated gene expression caused by the ΔtrmB mutation at a single cell level, we quantified the size and shape of H. salinarum cells in the ΔtrmB mutant and its isogenic Δura3 parent strain during mid-log phase growth (OD600 ~ 0.5) in complete medium (CM) via phase contrast microscopy. We observed that the ΔtrmB mutant strain was significantly rounder than the normally rod-shaped isogenic Δura3 parent strain in the absence of glucose (p < 10^{-22}, Figure 16, see Supplementary Table 1 for p-values of all comparisons). This effect is unlikely to be caused by slow growth, since H. salinarum normally becomes more rod shaped when
growth is stopped (Herrmann & Soppa, 2002). As expected from batch culture growth rate and gene expression phenotypes, this morphological defect is significantly complemented by the addition of 5% glucose to the medium (Figure 16). In order to test whether this shape defect may be caused by deficient S-layer glycosylation, we added the glycosylation inhibitor 2-deoxy-D-glucose (2DG, Esko JD, 2009) to cultures of the ΔtrmB mutant strain and the isogenic Δura3 parent strain. The Δura3 strain lost its rod shape and became significantly rounder in the presence of 2DG (p < 10^{-16}, Figure 16, see Supplementary Table 1 for p-values of all comparisons). Because both strains take on similar morphology in the presence of a glycosylation inhibitor, these results are consistent with the hypothesis that the lack of proper glycosylation can cause a shape defect such as that observed in the ΔtrmB mutant strain.
Figure 16: The ΔtrmB mutant strain exhibits altered cell morphology in the absence of glucose. Cells were visualized using phase-contrast microscopy and quantified by fitting an ellipse using ImageJ. (A) Bar graph shows the ratio of length to width in the Δura3 parent strain (dark bars) and the ΔtrmB mutant strain (light bars). Glucose complements the ΔtrmB mutant’s morphological phenotype, while the glycosylation inhibitor 2DG causes Δura3 cells to assume the same morphology as the ΔtrmB mutant strain. Error bars represent standard error of at least 39 measurements with an average of 114 cells. (B) Representative phase-contrast images of the Δura3
parent strain and the ΔtrmB mutant strain growing in CM, CM + 5% glucose, and CM + 2% 2DG

4.4.2 Carbohydrate staining suggests that ΔtrmB morphological defects are associated with reduced S-layer glycosylation.

In order to assess directly whether morphological defect of the ΔtrmB mutant strain is caused by a glycosylation defect, we assayed the carbohydrate to protein ratio of the S-layer glycoprotein from the ΔtrmB mutant strain and its isogenic Δura3 parent strain using Colloidal Coomassie Blue (CCB) and Periodic Acid Schiff (PAS) stains on SDS-PAGE. The carbohydrate to protein staining ratio is known to correlate with the amount of protein glycosylation (Gralnick et al., 1982). Because the S-layer represents a large fraction of total protein (~5%, Van et al., 2008) and migrates substantially slower than any other highly expressed protein (~200 kD), it is easily distinguishable from other proteins by SDS-PAGE. Whole cell lysates (Experimental Procedures) from the ΔtrmB mutant and its isogenic Δura3 parent strain growing in either CM, CM supplemented with 5% glucose, or CM supplemented with 2% of the glycosylation inhibitor 2DG were compared. We observed that the ΔtrmB mutant exhibited a significantly reduced carbohydrate to protein ratio in the S-layer protein compared to the Δura3 parent strain in the absence of glucose (Figure 17, p < 10⁻⁵, see Supplementary Table 2 for p-values of all comparisons). The ΔtrmB mutant strain also showed decreased carbohydrate staining of other proteins (Figure 17b). Glucose increased the carbohydrate to protein ratio of S-
layer glycoprotein from ΔtrmB mutant cells to the level of the Δura3 parent strain (see Supplementary Table 2 for p-values). In contrast, glucose did not substantially change the ratio in the Δura3 parent strain. Consistent with its role as a glycosylation inhibitor and with the morphological data (Figure 16), 2DG reduced the carbohydrate to protein ratio in the Δura3 strain to the same level as the ΔtrmB mutant strain (Figure 17), and only slightly further reduced this ratio in the ΔtrmB mutant strain. These trends are recapitulated in S-layer enriched preparations (Supplementary Figure 1). Taken together, these data suggest that the ΔtrmB mutant strain has reduced levels of protein glycosylation, which is especially apparent in the S-layer glycoprotein.
Figure 17: S-layer protein glycosylation is deficient in the ΔtrmB mutant strain in the absence of glucose. (A) Bar graph shows the ratio of PAS to CCB intensity in whole cell extracts. In the Δura3 parent strain (dark bars) S-layer protein glycosylation is constant with and without glucose and is inhibited by 2DG. In the ΔtrmB mutant strain (light bars), glycosylation is impaired in the absence of glucose, and inhibited by 2DG. Error bars represent standard error of an average of 10 replicates.(B) Representative images of PAS stained (left, pink bands) and CCB stained (right, blue bands) SDS-PAGE gels of *H. salinarum* whole cell extracts, showing decreased S-layer and general glycosylation in the ΔtrmB mutant strain in the absence of glucose and in both strains in the presence of 2DG. The black arrows to the right identify the S-layer band with glycosylation (top arrow) and without (bottom arrow).
4.4.3 The $\Delta trmB$ mutant glycosylation defect is associated with an inadequate supply of carbohydrate precursors.

TrmB has been established as a direct positive regulator of gluconeogenic enzyme-coding genes in *H. salinarum* (Schmid *et al.*, 2009, Todor *et al.*, 2013). This transcription factor directly regulates the expression of genes encoding enzymes leading from the TCA cycle, through pyruvate, phosphoenolpyruvate, and up to fructose 6-phosphate. Since genes encoding gluconeogenic enzymes are expressed at a lower level in the $\Delta trmB$ mutant than in the $\Delta ura3$ parent strain in the absence of glucose, we hypothesized that the glycosylation defect observed in the $\Delta trmB$ mutant strain may be due to reduced supply of glycosylation precursors. However, TrmB regulated gene products are also involved in glycerolipid anabolic pathways. We reasoned that defects in the dolichol phosphate lipid carriers on which the S-layer glycans are built and delivered (Lechner *et al.*, 1985) might also lead to the glycosylation defect of the $\Delta trmB$ mutant strain. To test this alternative hypothesis, we compared the growth of the mutant and parent strains in rich medium (CM) supplemented with varying amounts of 2DG and the antibiotic bacitracin (Figure 18a, see Supplementary Table 3a for p-values). Bacitracin acts to inhibit glycosylation by preventing dephosphorylation of the dolichol lipid pyrophosphate sugar carrier (Kato *et al.*, 1980), whereas 2DG acts by competing with glucose for incorporation into oligosaccharide chains and terminating these chains upon incorporation in *S. cerevisiae* (Johnson, 1968). To confirm that 2DG is a competitive inhibitor of glycosylation in *H. salinarum*, we assayed growth in rich medium...
supplemented with varying amounts of glucose and 2DG. We found that glucose rescued 2DG toxicity, suggesting that 2DG is a competitive inhibitor (Supplementary Figure 2). The ΔtrmB mutant was also strikingly more susceptible to 2DG than the parent strain. While the parent strain was able to grow in the presence of up to 3% 2DG, the ΔtrmB mutant ceased growth in the presence of 1% 2DG (Figure 18a). Interestingly, we found no difference in the growth inhibitory effect of bacitracin on the ΔtrmB mutant strain and the isogenic Δura3 parent strain (Figure 18b, see Supplementary Table 3a for p-values). This observation is consistent with the hypothesis that the ΔtrmB mutant is inhibited for glycosylation at the level of substrate availability rather than at the level of the lipid carriers involved in delivering sugar moieties to the S-layer. Combined with our previous knowledge that TrmB does not directly regulate the level of S-layer protein or the archaeal glycosylation genes (agl), and that genes coding for enzymes involved in gluconeogenesis are down regulated in the ΔtrmB mutant, these growth data strongly support the hypothesis that the glycosylation defect of the ΔtrmB mutant strain is caused by a defect in precursor availability.
Figure 18: The ΔtrmB mutant (light grey line) exhibits impaired growth at lower 2DG (A) concentrations than the Δura3 parent strain (dark grey line). In contrast, the growth inhibition due to bacitracin (B) is constant between the ΔtrmB mutant (light line) and the Δura3 parent strain (dark line). Growth was measured as the area under the log transformed curve and normalized to the highest growth rate for each strain. Error bars represent standard error of at least two biological replicates, each with at least two technical replicates.
4.4.4 Direct detection of intracellular metabolite concentrations reveals inadequate supply of carbohydrates in the ΔtrmB mutant strain.

In order to test whether the ΔtrmB mutant strain is defective in precursor availability, we measured intracellular metabolites in cell extracts from both strains growing in the presence and absence of glucose. Cells were grown in complete defined medium (CDM) in order to positively ascribe compounds to the organism. We used an untargeted mass spectrometric metabolomics approach (Methods) to identify candidate intermediates, as the specific precursor sugars and their assembly process are not fully characterized for this organism. Several monosaccharides were detected, including aldohexoses and ketohexoses. Sugar levels (relative to all identified metabolites) were significantly lower in the ΔtrmB mutant compared to those of the Δura3 parent in the absence of glucose (Figure 19, $P < 6 \times 10^{-3}$), but not in the presence of glucose ($P > 0.8$). The sugar levels in each strain under each condition followed the pattern observed for cell surface carbohydrate staining (Figure 17) and cell shape (Figure 16). This suggests that transcriptional regulation of gluconeogenic enzyme-coding genes by TrmB (Schmid et al., 2009) directly affects intracellular metabolite concentrations, which correlate strongly with other ΔtrmB mutant phenotypes. This is consistent with the hypothesis that an inadequate supply of monosaccharide precursors causes the glycosylation defect in ΔtrmB.
Figure 19: The ΔtrmB mutant strain (light bars) is deficient in sugars in the absence of glucose. Glucose complements this phenotype, returning sugar levels to those of the Δura3 isogenic parent strain (dark bars). Each bar represents the log-transformed ratio of summed glucose peak areas versus all identified metabolites normalized to the parent strain grown without glucose. Error bars represent standard error from the mean of three biological replicates.

4.4.5 TrmB activity is correlated with instantaneous growth rate.

Our data suggest that TrmB is required for regulating the supply of S-layer glycosylation precursors by controlling gluconeogenesis. Since cell growth and division require additional S-layer production to accommodate the changing cell surface area, and therefore additional glycosylation precursors, we reasoned that TrmB activity could be correlated with instantaneous growth rate. To test this, we reanalyzed three time course expression experiments from the literature (Facciotti et al., 2010, Schmid et al., 2007, Beer et al., 2014, Kaur et al., 2010). Instantaneous growth rate was calculated from the optical density of the culture over time. Since TrmB-DNA binding was not measured
in these experiments, we derived a gene expression proxy for TrmB binding. To generate this proxy, we calculated a composite average for the expression of genes whose activation (genes encoding enzymes in gluconeogenesis) or repression (genes encoding enzymes in glycolysis) depends exclusively upon TrmB binding (Todor et al., 2013; specific genes listed in Experimental Procedures). We applied this metric to gene expression data from H. salinarum cells (1) throughout the growth curve in batch culture (Facciotti et al., 2010, Figure 20a); (2) undergoing the transition from anoxic to oxic physiology (Schmid et al., 2007, Figure 20b); and (3) being challenged with a sub-lethal dose of H2O2 (Kaur et al., 2010, Figure 20c). We found that instantaneous growth rate and TrmB activity were significantly and positively correlated across all of these conditions and time scales (Pearson correlation = 0.579, p-value < 10^-7). These data suggest that TrmB responds to instantaneous growth rate. In order to confirm that the correlation between growth rate and TrmB-DNA binding was not due to nutrient changes over time, we reanalyzed data from a recent paper (Beer et al., 2014) in which H. salinarum was grown in ten media of varying ion composition but invariant nutrients. All cultures were harvested during balanced growth in mid-log phase (OD600 ~0.4-0.5). We found that under these conditions, TrmB-DNA binding still correlated strongly with growth rate (Pearson correlation = 0.448, p-value < 0.05, Figure 20d). Together with other data presented here (Figures 17-19), this proxy is consistent with TrmB linking
metabolism with growth rate, in order to ensure an appropriate supply of glycosylation precursors under varied environmental conditions.

Figure 20: Inferred TrmB–DNA binding (black line) correlates with growth rate (grey line) across growth in batch culture (A), during the shift from anoxic to oxic conditions in a turbidostat (B), during a sublethal challenge with H$_2$O$_2$ (C), and during mid-log phase growth in media of varying salinity (D). The dotted black and grey lines in (C) represent the TrmB–DNA binding and growth rate (respectively) prior to H$_2$O$_2$ addition as inferred from the data shown in (A). The dashed grey line in (D) depicts the linear regression between the proxy and growth rate in varying media formulations.
4.5 Discussion

In previous work, we showed that the transcription factor TrmB is important for regulating genes encoding enzymes across the metabolic network of *H. salinarum* (Schmid *et al.*, 2009). Specifically, genome-wide expression and TrmB–DNA binding location measurements demonstrated that TrmB regulates the promoters of 113 genes coding for enzymes involved in central carbon metabolism, purine metabolism, cobalamin biosynthesis, and other metabolic processes in response to glucose (Schmid *et al.*, 2009). Using dynamic gene expression profiling, we were able to identify the topology of TrmB regulation of central carbon metabolism, purine metabolism, and cobalamin biosynthesis (Todor *et al.*, 2013). However, the strong effect of glucose on the many TrmB regulated promoters and the gross morphological (Figure 16) and growth rate (Schmid *et al.*, 2009) defects of the ΔtrmB mutant strain in the absence of glucose are at odds with observations suggesting that *H. salinarum* cannot metabolize glucose for energy (Gochnauer and Kushner, 1969; Severina *et al.*, 1990; Sonawat *et al.*, 1990). In order to resolve this discrepancy, we performed a series of focused phenotypic profiling experiments. Combining these data with our previous knowledge of TrmB function, we hypothesize that TrmB is required for proper glycosylation of the S-layer protein by regulating the level of gluconeogenic enzymes that supply monosaccharide precursors from amino acids (Figure 17-19, Schmid *et al.*, 2009, Todor *et al.*, 2013).
Metabolomics analysis suggests that the concentration of monosaccharides is decreased in the ΔtrmB mutant in the absence of glucose. Although the assembly pathway for S-layer glycans is not currently known for *H. salinarum*, the low levels of sugars such as hexoses in the ΔtrmB mutant in the absence of glucose suggest a link between TrmB regulation of gluconeogenesis and S-layer glycoprotein assembly. In *H. salinarum*, TrmB directly regulates the expression of genes coding for gluconeogenic enzymes leading to fructose-6-phosphate, but derivatives of this intermediate other than glucose have not yet been investigated to determine if they complement the ΔtrmB mutant phenotype. Taken together with the ΔtrmB mutant’s sensitivity to 2DG and its lack of increased sensitivity to bacitracin, these data suggest that TrmB is required for maintaining the level of S-layer glycan precursors, likely by regulating gluconeogenic enzyme-coding genes. Our metabolomics work therefore paves the way for future characterization of glycan synthesis pathways in this organism.

The amount of S-layer protein present in a culture of *H. salinarum* is proportional to the surface area of the cells. Since each S-layer monomer is decorated with specific glycans at specific residues (Wieland, 1988), it follows that the amount of glycosylation is also proportional to the surface area of the cell. Because we did not observe a substantial change in the surface area to volume ratio of cells at different phases of growth (Supplementary Fig. S3), we reasoned that the rate of change in surface area of a culture is proportional to its instantaneous growth rate. Therefore, we propose that the
synthesis rate of glycosylation precursors will be proportional to the growth rate of \( H.\) \textit{salinarum}. Using a gene expression proxy for TrmB–DNA binding, we found that growth rate is correlated with TrmB–DNA binding activity (Figure 20). This proxy is a valid metric for TrmB–DNA binding activity because (i) it is well correlated with empirical TrmB–DNA binding measured under other conditions (Schmid \textit{et al.}, 2009, Supplementary Fig. S4), (ii) the \textit{PpssA::GFP} promoter fusion shows an expression pattern across growth consistent with inferred TrmB–DNA binding activity (Supplementary Fig. S5), and (iii) as expected, the proxy reports a constant low level in the \( \Delta \text{trmB} \) mutant in the absence and presence of glucose and in the \( \Delta \text{ura3} \) parent strain in the presence of glucose (Supplementary Fig. S6). Because TrmB dissociates from the DNA in the presence of glucose, and because the level of TrmB itself appears to be constant throughout growth (Bonneau \textit{et al.}, 2007; Schmid \textit{et al.}, 2009), changes in TrmB–DNA binding are likely due to changes in intracellular monosaccharide concentration. Because allosteric regulation of gluconeogenic enzymes is weak in many archaea (Fukuda \textit{et al.}, 2004; Say and Fuchs, 2010), we hypothesize that higher growth rates deplete the monosaccharide pool as it is used for S-layer glycosylation, thereby increasing TrmB–DNA binding. In contrast, lower growth rates may lead to accumulation of monosaccharides and decrease TrmB–DNA binding. In this way, TrmB could co-ordinate gluconeogenic flux with respect to growth.
The current study provides additional clarity into the evolutionary rationale of TrmB co-regulation of peripheral metabolic pathways such as amino acid metabolism, purine synthesis, and cobalamin synthesis (Schmid et al., 2009, Todor et al., 2013). Purine and cobalamin demand is influenced both by instantaneous growth rate and by other factors, such as availability in the environment. Because TrmB activity is correlated with growth via the incorporation rate of sugars into the S-layer, TrmB may function as a growth rate-specific regulator for these peripheral pathways as well. Therefore, appropriate expression of these genes encoding enzymes in peripheral pathways requires co-ordinate regulation by TrmB and yet to be determined regulators.

TrmB is widely conserved throughout the archaeal domain, suggesting that it could function as a growth rate responsive regulator in other archaea as well. In the hyperthermophilic archaeon Pyrococcus furiosus, a TrmB-like transcription factor, TrmBL1, acts as a global regulator of gluconeogenic and glycolytic genes. Although its physiological inducer is yet to be determined (Lee et al., 2008), TrmBL1 appears to function similarly to the Tgr transcription factor of Thermococcus kodakarensis. Tgr functions to repress glycolytic genes and to activate gluconeogenic genes in response to glucose (Kanai et al., 2007). The Δtgr mutant grows slowly in the absence of glucose, suggesting the potential for a global regulatory role such as the one that had been postulated for H. salinarum. Perhaps the most diverged TrmB homologue identified is the MreA protein of the crenarchaeon Methanosarcina acetivorans (Reichlen et al., 2012). In
this organism, MreA controls acetotrophic and methylotrophic pathways in response to an unknown inducer. The ΔmreA strain grows slower than the wild-type on acetate due to constitutive downregulation of acetotrophic genes. Although TrmB homologues have been identified and studied in a variety of archaea, TrmB targets and dynamics have only been investigated in *H. salinarum*. Given the wide range of habitats and metabolic activities in the archaeal domain, further study of TrmB homologues and their respective regulons would be fruitful, especially in closely related haloarchaea with different nutrient requirements.

In the most heavily studied representatives of the bacterial and eukaryotic domains, the yeast *S. cerevisiae* and the enterobacteria *E. coli*, the co-ordination of growth rate and metabolism has two components: one dedicated to appropriate carbon flux and one dedicated to appropriate nitrogen flux. In *S. cerevisiae* the sensor and effector system for connecting carbon flux to growth rate is the Ras/cAMP/PKA system, while the primary sensor system for nitrogen flux is the TOR pathway (Ramachandran and Herman, 2011). In *E. coli*, the carbon sensor is Crp, while the nitrogen sensor is Lrp and its downstream effectors (Cho *et al*., 2008, You *et al*., 2013). In *E. coli*, these two systems operate to determine the appropriate proteome partitioning at varied rates of growth and in different environmental conditions (You *et al*., 2013), and the Crp/cAMP system has been shown to respond to nutrients in a growth rate-dependent manner. Given the expansive regulon of TrmB (Schmid *et al*., 2009) as well as its ability to sense nutrient
conditions and growth rate, we hypothesize that TrmB plays a role analogous to the Crp and the Ras/cAMP/PKA in *H. salinarum*. Like Crp and the Ras/cAMP/PKA system, TrmB is the primary regulator of some pathways, and a contributing regulator to many others (Schmid *et al.*, 2009; Busti *et al.*, 2010; Shimada *et al.*, 2011; Todor *et al.*, 2013).

Additionally, mutations in TrmB affect the ability of the cell to make use of certain nutrients and cause a growth defect, similar to mutations in *crp*. Both Crp and the Ras/cAMP/PKA control cell size. Ras/cAMP/PKA has been shown to control mass accumulation (Busti *et al.*, 2010), while Crp indirectly controls the accumulation of UDP-Glucose, a metabolite that interacts with FtsZ to determine cell size (Hill *et al.*, 2013).

Unlike Crp and the Ras/cAMP/PKA system, the direct effector of TrmB in *H. salinarum* remains unknown. While glucose is able to relieve TrmB repression when added to the media, it is not clear whether it is directly interacting with TrmB or whether a second messenger is involved. Regardless, the broad similarities and similar mutant phenotypes of TrmB, Crp, and Ras/cAMP/PKA, suggest that TrmB is acting in an analogous role in *H. salinarum*. Surprisingly, this suggests that despite the lack of protein homology, a conserved transcriptional strategy links growth rate and nutrition across all three domains.

In closing, our findings suggest a transcriptional mechanism for linking cell growth and metabolism in an archaeon. An evolutionarily conserved transcription factor can regulate metabolism by monitoring the anabolic production of a small molecule that
is integrated into the cell surface. Interestingly, both TrmB and its cis-regulatory motif are well conserved throughout the archaeal domain, suggesting that TrmB could be a conserved mechanism for connecting cell growth and metabolism. The variety of lifestyles and metabolic niches occupied by members of the archaea make TrmB an excellent case study for the co-evolution of transcriptional and metabolic networks in diverse environments. Although much work remains to be done to elucidate the mechanisms of co-ordinated control of cell shape, cell cycle, and growth rate in archaea, our findings represent a first step in understanding how archaea sense and respond to changes in cell growth in their extreme and varied environments. Combined with previous research on the TrmB responsive gene regulatory network, this work paves the way for future studies on the conservation of the topology underlying this connection across all three domains.

4.6 Supplementary Data

Supplementary Data are available online at:


4.7 Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant numbers MCB-1052290 and MCB-1417750 to AKS. We would like to
acknowledge Lee Pang for conceptualizing and creating the bsdAnalysis function and
for refining it to work seamlessly with Bioscreen files. We thank Karlyn Beer for
supplying growth data from media of different salinities. The authors have no conflicts
of interest to report.
5. Conclusion and Future Directions

5.1 Conclusion

In this dissertation, we expand the understanding of the transcriptional regulation of metabolism in the halophilic archaeon *H. salinarum*. Previous work had identified TrmB as a central regulator of metabolism. In *H. salinarum*, TrmB was shown to regulate the promoters of 113 genes. Many of these genes encode enzymes involved in various metabolic pathways, such as gluconeogenesis, purine synthesis, and cobalamin synthesis, suggesting that TrmB is a central metabolic regulator. Like TrmB homologues in the hyperthermophilic archaea, TrmB in *H. salinarum* binds to a conserved *cis*-regulatory sequence in the absence of its biological effector sugar. Upon addition of sugar (in this case glucose) TrmB unbinds from the genome, changing gene expression from previously bound promoters. Despite this established mechanism, prior to the current work, how TrmB was able to accurately regulate such diverse pathways remained unclear. In order to address this question, we performed dynamic measurements of gene expression using NanoString and dynamic measurements of TrmB promoter occupancy using ChIP-qPCR. We modeled these time course measurements in a simple ODE model and determined to what extent TrmB was responsible for the observed gene expression patterns for each gene in our dataset. Specifically, we found that TrmB appears to be the sole regulator of gluconeogenesis,
and that it functions with other regulators to regulate peripheral pathways. Purine synthesis was found to be predominantly regulated by another regulator, with only minor input from TrmB, while cobalamin synthesis was found to be regulated by TrmB and another regulator acting together in a feed-forward loop. These regulators remain to be identified and characterized.

In addition to understanding how TrmB transcriptional regulation of metabolic enzyme coding genes occurs, another primary goal of this work was to determine how this transcriptional regulation affects the level of metabolic intermediates, and ultimately how this regulation affects phenotype. In order to address this question, we assayed metabolite levels using a combination of targeted and untargeted approaches over the same experimental time course used for gene expression. We found that TrmB regulation was required for the maintenance of proper metabolite levels before and during a transition in sugar levels. Additionally, we found evidence that low purine levels in the $\Delta trmB$ mutant may be partially responsible for its growth defect. We followed up this observation using targeted quantitative phenotyping techniques and found that a metabolic block in PRPP synthesis caused by a defect in gluconeogenesis in the $\Delta trmB$ mutant in the absence of glucose was responsible for the defect in purine synthesis. Additionally, we found that the defect in purine synthesis was partially responsible for the growth defect of the $\Delta trmB$ mutant strain in the absence of glucose.
Interestingly, despite the apparent importance of glucose for both the regulation and functioning of metabolism in *H. salinarum*, evidence suggests that *H. salinarum* cannot metabolize glucose for energy. This led us to ask why *H. salinarum* regulates so much of its metabolism in response to a compound it cannot use for energy. We resolved this discrepancy by showing that a central use of gluconeogenically produced sugars was glycosylation of the S-layer protein. Therefore, TrmB regulates gluconeogenesis in order to ensure sufficient sugars for glycosylation. Because the rate of S-layer production is proportional to growth rate, we hypothesized that TrmB was able to sense growth rate by this mechanism. Using a gene expression proxy for TrmB binding, we showed that TrmB-DNA binding and growth rate are well correlated across various conditions. Taken together, our observations suggest that TrmB functions as a global regulator of metabolism and that it plays an important role in coordinating metabolism with growth rate.

### 5.2 Future directions

In the current work, we showed that TrmB is a central regulator of metabolism. Although much is known about the function of TrmB in the archaea, several outstanding questions remain.
5.2.1 Biochemical characterization of TrmB binding mechanism

In many of the hyperthermophilic archaea, the effector for TrmB unbinding has been determined using purified TrmB protein and biochemical assays such as EMSA. In addition, the biophysical constants of TrmB sugar and DNA binding have been determined, allowing precise modeling of the interaction. In contrast, in *H. salinarum* the physiological effector has not been conclusively determined. It has been shown that the addition of glucose to the media results in TrmB unbinding within 2 minutes (Todor et al., 2013). However, the possibility of an enzymatic conversion to a different effector sugar remains open. Purification of TrmB from *H. salinarum* and exploration of its biochemical binding would improve our understanding of its function and of its physiological role.

5.2.2 Evolutionary perspectives

As previously discussed, TrmB is well conserved among the archaea. This conservation is especially strong in the haloarchaea, with expect values at the level of amino acid sequence between TrmB homologues of less than $10^{-100}$. Intriguingly, haloarchaea have many diverse metabolic pathways, needs, and abilities. For example, *Halarhabdus utahensis* is unable to use amino acids for energy, but can metabolize glucose, fructose, and xylose (Waino et al., 2000). *Halosimplex carlsbadense* can only use
simple compounds such as pyruvate for energy (Vreeland et al., 2002). Still other halophiles such as *Haloferax volcanii* are able to use both amino acids and sugars for energy (Anderson et al., 2010). This metabolic diversity raises questions about the role of TrmB in each species. Specifically, does TrmB play a global role regardless of metabolic ability, or is it only a global regulator in species that cannot metabolize sugar? To address this question, both computational and experimental methods can be applied to related halophile species in order to determine the TrmB regulon and infer its function. For example, TrmB-DNA binding sites in other halophiles can be determined using ChIP-chip or ChIP-Seq. Alternatively, a computational method may be used to search for the conserved TrmB cis-regulatory motif in the genomes of other halophiles to determine which genes may be regulated by TrmB. By looking at the genes whose promoters TrmB binds, its regulatory scope can be determined, and questions about its potential role as a global regulator of metabolism can be addressed.

### 5.3 Funding

This work was supported by the National Science Foundation [NSF-MCB 1052290 to A.K.S.], the National Institutes of Health [P50 to Duke CSB], and Duke Startup [to A.K.S.].
Appendix A

Supplementary Table 1: List of metabolites in each of the 6 clusters.

### Cluster 1

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>38002</td>
<td>Lipids</td>
<td>Glycerolipids</td>
<td>1,2-propanediol</td>
</tr>
<tr>
<td>1554</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>2-ethylhexanoic acid</td>
</tr>
<tr>
<td>27548</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>2-phenoxyethanol</td>
</tr>
<tr>
<td>12091</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>methyl palmitate (16:0)</td>
</tr>
<tr>
<td>43399</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>tetradeanol</td>
</tr>
<tr>
<td>32634</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 11317</td>
</tr>
<tr>
<td>33026</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 11684</td>
</tr>
<tr>
<td>35327</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 13549</td>
</tr>
<tr>
<td>37807</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15558</td>
</tr>
<tr>
<td>37808</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15559</td>
</tr>
<tr>
<td>37811</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15562</td>
</tr>
<tr>
<td>37812</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15563</td>
</tr>
<tr>
<td>37826</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15577</td>
</tr>
<tr>
<td>37906</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15657</td>
</tr>
<tr>
<td>38840</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 16271</td>
</tr>
<tr>
<td>45121</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20520</td>
</tr>
<tr>
<td>45125</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20524</td>
</tr>
</tbody>
</table>

### Cluster 2

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>15996</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>aspartate</td>
</tr>
<tr>
<td>1125</td>
<td>Amino acid</td>
<td>Branched Chain Amino Acids (OAA derived)</td>
<td>isoleucine</td>
</tr>
<tr>
<td>60</td>
<td>Amino acid</td>
<td>Branched Chain Amino Acids (pyruvate derived)</td>
<td>leucine</td>
</tr>
<tr>
<td>1302</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>methionine</td>
</tr>
<tr>
<td>64</td>
<td>Amino acid</td>
<td>Aromatic amino</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>Compound ID</td>
<td>Super Pathway</td>
<td>Sub Pathway</td>
<td>BIOCHEMICAL</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
<td>------------------------------</td>
</tr>
<tr>
<td>54</td>
<td>Amino acid</td>
<td>acid metabolism (PEP derived)</td>
<td>tryptophan</td>
</tr>
<tr>
<td>1299</td>
<td>Amino acid</td>
<td>Aromatic amino acid metabolism (PEP derived)</td>
<td>tyrosine</td>
</tr>
<tr>
<td>35963</td>
<td>Carbohydrate</td>
<td>Glycolysis</td>
<td>dihydroxyacetone</td>
</tr>
<tr>
<td>37516</td>
<td>Carbohydrate</td>
<td>Amino sugar and nucleotide sugar</td>
<td>arabonate</td>
</tr>
<tr>
<td>1564</td>
<td>Carbohydrate</td>
<td>TCA cycle</td>
<td>citrate</td>
</tr>
<tr>
<td>27753</td>
<td>Carbohydrate</td>
<td>Sucrose, glucose, fructose metabolism</td>
<td>erythrose</td>
</tr>
<tr>
<td>37427</td>
<td>Carbohydrate</td>
<td>Sucrose, glucose, fructose metabolism</td>
<td>erythulose</td>
</tr>
<tr>
<td>577</td>
<td>Carbohydrate</td>
<td>Sucrose, glucose, fructose metabolism</td>
<td>fructose</td>
</tr>
<tr>
<td>587</td>
<td>Carbohydrate</td>
<td>Amino sugar and nucleotide sugar</td>
<td>gluconate</td>
</tr>
<tr>
<td>38507</td>
<td>Carbohydrate</td>
<td>Glycolysis</td>
<td>glucose</td>
</tr>
<tr>
<td>1572</td>
<td>Carbohydrate</td>
<td>Glycolysis</td>
<td>glycerate</td>
</tr>
<tr>
<td>563</td>
<td>Carbohydrate</td>
<td>Sucrose, glucose, fructose metabolism</td>
<td>sorbose</td>
</tr>
<tr>
<td>568</td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Biotin metabolism</td>
<td>biotin</td>
</tr>
<tr>
<td>34732</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>isovalerate (C5)</td>
</tr>
<tr>
<td>45117</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20516</td>
</tr>
</tbody>
</table>

**Cluster 3**

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1419</td>
<td>Amino acid</td>
<td>Amines and polyamines</td>
<td>5-methylthioadenosine (MTA)</td>
</tr>
<tr>
<td>35100</td>
<td>Amino acid</td>
<td>Branched Chain</td>
<td>beta-methyllevulinate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Amino Acids (pyruvate derived)</td>
<td>Aspartate family (OAA derived)</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1594</td>
<td>Amino acid</td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Nicotinate and nicotinamide metabolism</td>
</tr>
<tr>
<td>5278</td>
<td></td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Nicotinate and nicotinamide metabolism</td>
</tr>
<tr>
<td>32380</td>
<td></td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Riboflavin and FAD metabolism</td>
</tr>
<tr>
<td>1827</td>
<td></td>
<td>Lipids</td>
<td>Glycerolipids</td>
</tr>
<tr>
<td>21127</td>
<td>Lipids</td>
<td>Lipids</td>
<td>Glycerolipids</td>
</tr>
<tr>
<td>21188</td>
<td></td>
<td></td>
<td>Free fatty acid</td>
</tr>
<tr>
<td>42103</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>39533</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>33185</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>38360</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>41783</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>42193</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>42376</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>45184</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>45119</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>45120</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>45126</td>
<td>N/A</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>555</td>
<td>Nucleotide</td>
<td>Purine metabolism</td>
<td>adenosine</td>
</tr>
<tr>
<td>32342</td>
<td>Nucleotide</td>
<td>Purine metabolism</td>
<td>AMP</td>
</tr>
<tr>
<td>1573</td>
<td>Nucleotide</td>
<td>Purine metabolism</td>
<td>guanosine</td>
</tr>
<tr>
<td>2849</td>
<td>Nucleotide</td>
<td>Purine</td>
<td>5'- GMP</td>
</tr>
</tbody>
</table>
### Secondary metabolism

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>15778</td>
<td>Secondary metabolism</td>
<td>Benzenoids</td>
<td>benzoate</td>
</tr>
<tr>
<td>37499</td>
<td>Secondary metabolism</td>
<td>Benzenoids</td>
<td>p-toluic acid</td>
</tr>
<tr>
<td>38154</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>heptaethylene glycol</td>
</tr>
<tr>
<td>38133</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>hexaethylene glycol</td>
</tr>
<tr>
<td>42361</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>octaethylene glycol</td>
</tr>
<tr>
<td>38158</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>pentaethylene glycol</td>
</tr>
<tr>
<td>27743</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>triethyleneglycol</td>
</tr>
</tbody>
</table>

### Cluster 4

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>6146</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>2-amino adipate</td>
</tr>
<tr>
<td>37253</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>2-hydroxyglutarate</td>
</tr>
<tr>
<td>1494</td>
<td>Amino acid</td>
<td>Glutathione metabolism</td>
<td>5-oxoproline</td>
</tr>
<tr>
<td>22132</td>
<td>Amino acid</td>
<td>Branched Chain Amino Acids (pyruvate derived)</td>
<td>alpha-hydroxyisocaproate</td>
</tr>
<tr>
<td>37092</td>
<td>Amino acid</td>
<td>Glutathione metabolism</td>
<td>gamma-glutamyl-2-aminobutyrate</td>
</tr>
<tr>
<td>57</td>
<td>Amino acid</td>
<td>Glutamate family (alpha-ketoglutarate derived)</td>
<td>glutamate</td>
</tr>
<tr>
<td>23642</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>homoserine (homoserine lactone)</td>
</tr>
<tr>
<td>37076</td>
<td>Amino acid</td>
<td>Serine family (phosphoglycerate derived)</td>
<td>N-acetylserine</td>
</tr>
<tr>
<td>33939</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>N-acetylthreonine</td>
</tr>
<tr>
<td>40699</td>
<td>Amino acid</td>
<td>Aromatic amino acid metabolism (PEP derived)</td>
<td>phenylacetylphenylalanine</td>
</tr>
<tr>
<td>43386</td>
<td>Amino acid</td>
<td>Aromatic amino acid metabolism (PEP derived)</td>
<td>tryptophanol</td>
</tr>
<tr>
<td>Compound ID</td>
<td>Super Pathway</td>
<td>Sub Pathway</td>
<td>BIOCHEMICAL</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>1303</td>
<td>Carbohydrate</td>
<td>TCA cycle</td>
<td>malate</td>
</tr>
<tr>
<td>32484</td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>CoA metabolism</td>
<td>acetyl-CoA</td>
</tr>
<tr>
<td>34829</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 13137</td>
</tr>
<tr>
<td>35247</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 13483</td>
</tr>
<tr>
<td>38514</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 16071</td>
</tr>
<tr>
<td>39259</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 16682</td>
</tr>
<tr>
<td>41331</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 18227</td>
</tr>
<tr>
<td>45189</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20588</td>
</tr>
<tr>
<td>37063</td>
<td>Peptide</td>
<td>gamma-glutamyl gamma-glutamylalanine</td>
<td></td>
</tr>
<tr>
<td>1778</td>
<td>Peptide</td>
<td>gamma-glutamyl gamma-glutamylcysteine</td>
<td></td>
</tr>
<tr>
<td>32393</td>
<td>Peptide</td>
<td>gamma-glutamyl gamma-glutamylvaline</td>
<td></td>
</tr>
</tbody>
</table>

Cluster 5

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>15676</td>
<td>Amino acid</td>
<td>Branched Chain Amino Acids (pyruvate derived)</td>
<td>3-methyl-2-oxovalerate</td>
</tr>
<tr>
<td>22116</td>
<td>Amino acid</td>
<td>Branched Chain Amino Acids (pyruvate derived)</td>
<td>4-methyl-2-oxopentanoate</td>
</tr>
<tr>
<td>33453</td>
<td>Carbohydrate</td>
<td>TCA cycle</td>
<td>alpha-ketoglutarate</td>
</tr>
<tr>
<td>599</td>
<td>Carbohydrate</td>
<td>Glycolysis</td>
<td>pyruvate</td>
</tr>
<tr>
<td>2134</td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Riboflavin and FAD metabolism</td>
<td>FAD</td>
</tr>
<tr>
<td>33963</td>
<td>Lipids</td>
<td>Cutin and cuticular</td>
<td>acetoacetate</td>
</tr>
<tr>
<td>32489</td>
<td>Lipids</td>
<td>Free fatty acid caproate (6:0)</td>
<td></td>
</tr>
<tr>
<td>32492</td>
<td>Lipids</td>
<td>Free fatty acid caprylate (8:0)</td>
<td></td>
</tr>
<tr>
<td>1644</td>
<td>Lipids</td>
<td>Free fatty acid heptanoate (7:0)</td>
<td></td>
</tr>
<tr>
<td>12035</td>
<td>Lipids</td>
<td>Free fatty acid pelargonate (9:0)</td>
<td></td>
</tr>
<tr>
<td>33443</td>
<td>Lipids</td>
<td>Free fatty acid valerate (5:0)</td>
<td></td>
</tr>
<tr>
<td>Compound ID</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 15503</td>
</tr>
<tr>
<td>-------------</td>
<td>-----</td>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>45123</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20522</td>
</tr>
<tr>
<td>45124</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 20523</td>
</tr>
<tr>
<td>1505</td>
<td>Nucleotide</td>
<td>Pyrimidine metabolism</td>
<td>orotate</td>
</tr>
<tr>
<td>34456</td>
<td>Peptide</td>
<td>gamma-glutamyl</td>
<td>gamma-glutamylisoleucine*</td>
</tr>
<tr>
<td>18369</td>
<td>Peptide</td>
<td>gamma-glutamyl</td>
<td>gamma-glutamylleucine</td>
</tr>
<tr>
<td>40045</td>
<td>Peptide</td>
<td>Dipeptide</td>
<td>leucylglycine</td>
</tr>
<tr>
<td>38101</td>
<td>Xenobiotics</td>
<td>Chemicals</td>
<td>tetraethylene glycol</td>
</tr>
</tbody>
</table>

### Cluster 6

<table>
<thead>
<tr>
<th>Compound ID</th>
<th>Super Pathway</th>
<th>Sub Pathway</th>
<th>BIOCHEMICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>605</td>
<td>Nucleotide</td>
<td>Pyrimidine metabolism</td>
<td>uracil</td>
</tr>
<tr>
<td>42887</td>
<td>N/A</td>
<td>N/A</td>
<td>X - 19411</td>
</tr>
<tr>
<td>43546</td>
<td>Lipids</td>
<td>Free fatty acid</td>
<td>5-methylnorleucine</td>
</tr>
<tr>
<td>39541</td>
<td>Cofactors, Prosthetic Groups, Electron Carriers</td>
<td>Thiamine metabolism</td>
<td>thiamin (Vitamin B1)</td>
</tr>
<tr>
<td>1126</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>alanine</td>
</tr>
<tr>
<td>1638</td>
<td>Amino acid</td>
<td>Glutamate family (alpha-ketoglutarate derived)</td>
<td>arginine</td>
</tr>
<tr>
<td>11398</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>asparagine</td>
</tr>
<tr>
<td>1647</td>
<td>Amino acid</td>
<td>Glutamate family (alpha-ketoglutarate derived)</td>
<td>glutamine</td>
</tr>
<tr>
<td>11777</td>
<td>Amino acid</td>
<td>Serine family (phosphoglycerate derived)</td>
<td>glycine</td>
</tr>
<tr>
<td>1301</td>
<td>Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>lysine</td>
</tr>
<tr>
<td>1493</td>
<td>Amino acid</td>
<td>Glutamate family (alpha-</td>
<td>ornithine</td>
</tr>
<tr>
<td></td>
<td>Amino acid</td>
<td>Metabolism Pathway</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>------------</td>
<td>--------------------------------------------------------------------------------</td>
<td>---</td>
</tr>
<tr>
<td>15958 Amino acid</td>
<td>ketoglutarate derived)</td>
<td>aromatic amino acid metabolism (PEP derived)</td>
<td>phenylacetate</td>
</tr>
<tr>
<td>1898 Amino acid</td>
<td>Glutamate family (alpha-ketoglutarate derived)</td>
<td>proline</td>
<td></td>
</tr>
<tr>
<td>1648 Amino acid</td>
<td>Serine family (phosphoglycerate derived)</td>
<td>serine</td>
<td></td>
</tr>
<tr>
<td>1284 Amino acid</td>
<td>Aspartate family (OAA derived)</td>
<td>threonine</td>
<td></td>
</tr>
<tr>
<td>1649 Amino acid</td>
<td>Branched Chain Amino Acids (pyruvate derived)</td>
<td>valine</td>
<td></td>
</tr>
</tbody>
</table>

**Supplementary Table 2: Bonferroni corrected p-values of the T-test.**

<table>
<thead>
<tr>
<th></th>
<th>CDM</th>
<th>CDM + Gluc</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔtrmB</td>
<td>Δura3</td>
<td>Δura3</td>
</tr>
<tr>
<td>CDM</td>
<td>Δura3</td>
<td>2.58E-04</td>
</tr>
<tr>
<td>CDM</td>
<td>ΔtrmB</td>
<td></td>
</tr>
<tr>
<td>CDM + Gluc</td>
<td>Δura3</td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure 1: Figure shows the $-\log_{10}$ p-value between $\Delta ura3$ and $\Delta trmB$ at each timepoint for each of the 6 clusters.
References


Farlow WG (1878) “On the nature of the peculiar reddening of salted codfish during the summer season.” *U.S. Fish Comm., Report of the Commissioner for 1878*, pp 969–973


in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes.” *J Biol Chem* 282: 33659-33670.


Sonawat, H.M., S. Srivastava, S. Swaminathan & G. Govil, (1990) “Glycolysis and
Entner-Doudoroff pathways in Halobacterium halobium: some new observations


characterization of salvage pathways and isolation of mutants resistant to purine

Sumper, M., E. Berg, R. Mengele & I. Strobel, (1990) “Primary structure and
glycosylation of the S-layer protein of Haloferax volcanii.” J Bacteriol 172: 7111-
7118.

families.” Science, 278, 631-637.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V.,
COG database: an updated version includes eukaryotes.” BMC Bioinformatics, 4, 41.

synthetase and pyruvate, phosphate dikinase of Thermoproteus tenax: key pieces

dynamics predict transcriptional response to nutrients in archaea.” Nucleic Acids
Res 41: 8546-8558.

hypersaline adapted archaeon Halobacterium salinarum.” Molecular Microbiology,


Biography

Biography of Horia Todor

I was born in Bucharest, Romania on May 2nd, 1987. In 2009 I received a Bachelor of Science in Biology as well as a Bachelor of Arts in Economics from the University of Virginia. After an internship at the National Institutes of Health, I began a PhD in Biology at Duke University in Durham, North Carolina. My biology dissertation was completed in the spring of 2015.

Fellowships, Grants, and Awards

1. Duke Graduate School Travel Fellowship (2014)
2. Duke University Biology Graduate Fellowship (2010)

Publications

