Molecular Systems Biology 5; Article number 282; doi:10.1038/msb.2009.40
Citation: ***Molecular Systems Biology*** **5:282**

www.molecularsystemsbiology.com

molecular systems biology

A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability

Amy K Schmid[1], David J Reiss[1], Min Pan[1], Tie Koide[1,3] and Nitin S Baliga[1,2,*]

[1] Institute for Systems Biology, Seattle, WA, USA and [2] Department of Microbiology, University of Washington, Seattle, WA, USA
[3] Present address: E-mail: tiekoide@gmail.com
* Corresponding author. Baliga Lab, Institute for Systems Biology, 1441 N 34th St., Seattle, WA 98103-8904, USA. Tel.: +1 206 732 1266; Fax: +1 206 732 1299; E-mail: nbaliga@systemsbiology.org

Received 20.1.09; accepted 15.5.09

During evolution, enzyme-coding genes are acquired and/or replaced through lateral gene transfer and compiled into metabolic pathways. Gene regulatory networks evolve to fine tune biochemical fluxes through such metabolic pathways, enabling organisms to acclimate to nutrient fluctuations in a competitive environment. Here, we demonstrate that a single TrmB family transcription factor in *Halobacterium salinarum NRC-1* globally coordinates functionally linked enzymes of diverse phylogeny in response to changes in carbon source availability. Specifically, during nutritional limitation, TrmB binds a *cis*-regulatory element to activate or repress 113 promoters of genes encoding enzymes in diverse metabolic pathways. By this mechanism, TrmB coordinates the expression of glycolysis, TCA cycle, and amino-acid biosynthesis pathways with the biosynthesis of their cognate cofactors (e.g. purine and thiamine). Notably, the TrmB-regulated metabolic network includes enzyme-coding genes that are uniquely archaeal as well as those that are conserved across all three domains of life. Simultaneous analysis of metabolic and gene regulatory network architectures suggests an ongoing process of co-evolution in which TrmB integrates the expression of metabolic enzyme-coding genes of diverse origins.
Molecular Systems Biology 5: 282; published online 16 June 2009; doi:10.1038/msb.2009.40
Subject Categories: chromatin & transcription; signal transduction
Keywords: archaea; central metabolism; ChIP-chip; transcription regulation; TrmB

Introduction

Archaeal genomes encode unusual metabolic enzymes with homologs in either eukarya or bacteria (Siebers and Schonheit, 2005). Several homologous gene replacement events are speculated to have an important function in evolution to integrate these enzymes into archaeal metabolic networks that are otherwise comprised of enzymes conserved across two or more domains of life (Galperin and Koonin, 1999; Siebers and Schonheit, 2005). If so, then this raises important questions regarding the evolution and the architecture(s) of gene regulatory networks (GRNs) that integrate and coordinate enzyme-coding genes within archaeal metabolic networks in the face of unique environmental challenges.

GRNs evolve by internalizing environmental factor changes to coordinate the efficient uptake and usage of limited nutritional resources (Tagkopoulos *et al*, 2008). Not surprisingly, the activity of many transcription factors (TFs) in these GRNs reflects cellular adaptations to environmental niches. For example, greater than half of all TFs in bacteria are thought to bind small molecules to monitor changes in environmental and cellular status (Madan Babu and Teichmann, 2003). Likewise, at least 50 eukaryotic TFs coordinate central metabolic pathways in multiple cellular compartments (Herrgard *et al*, 2006; Reece *et al*, 2006). Although limited information exists on archaeal GRNs, it is known that the pre-initiation complex (PIC) is made up of orthologs of the eukaryotic general transcription factors (GTFs): transcription factor II B (TFB), a TATA-binding protein (TBP), and a eukaryotic RNA-Pol II-like polymerase (Geiduschek and Ouhammouch, 2005). In contrast, many of their sequence-specific repressors and activators of transcription share ancestry with bacterial transcription regulators (Bell, 2005). However, only 10 of these regulators have been characterized

to date (Bell, 2005), and even fewer have a known function *in vivo* (Lie *et al*, 2005; Muller and DasSarma, 2005; Lee *et al*, 2008). Those that have a known function *in vivo* include transcription regulators for glycolytic/gluconeogenic, nitrogen, and lysine usage pathways (Brinkman *et al*, 2002; Lie *et al*, 2005; Kanai *et al*, 2007). For example, the TrmB transcription factor (*t*hermococcus *r*egulator of *m*altose *b*inding) acts as a repressor for genes encoding glycolytic enzymes and as activator for genes encoding gluconeogenic enzymes (Kanai *et al*, 2007). In these systems, TrmB also binds to glucose, maltose, trehalose, maltodexterins, and sucrose molecules to differentially regulate the genes encoding corresponding sugar uptake systems in a sequence-specific manner (van de Werken *et al*, 2006; Lee *et al*, 2008). However, the regulation of all other metabolic pathways in archaea is currently unknown.

The limited information on regulation of metabolism in archaea is a significant handicap in comparative analysis for understanding evolutionary similarities and differences in the architecture(s) of GRNs. Here we have characterized the TrmB regulatory network in the halophilic archaeon *Halobacterium salinarum NRC-1* by integrating three disparate sources of evidence (protein–DNA interactions measured globally with ChIP-chip, transcriptional responses of genetically and environmentally perturbed strains using microarray analysis, and genome-wide distribution of a conserved TF-binding motif signature) with a metabolic reconstruction. These results demonstrate that the haloarchaeal TrmB ortholog (VNG1451C) coordinates the transcription of more than 100 central metabolic enzyme-coding genes with genes involved in *de novo* synthesis of their cognate cofactors. We hypothesize that this balanced regulation allows the cell to modulate redox and energy status. More importantly, we show that the TrmB-dependent metabolic network integrates the transcription of enzyme-coding genes that are uniquely archaeal with those that are conserved across all three domains of life. In sum, this study provides insight into how the architecture of a large metabolic network and an associated GRN may have co-evolved using components of diverse origins, and how this assembly may be conserved across the archaeal lineage.

Results

We used a combination of classical genetics, genome-wide experimental, and computational approaches to identify the TrmB ortholog and characterize the architecture of the network it specifies to control central metabolism in the archaeon *H. salinarum*. These approaches included (i) sequence analysis to identify a putative TrmB homolog; (ii) phenotypic characterization of a Δ*trmB* deletion strain; (iii) transcriptomic analysis of the Δ*trmB* strain under defined growth conditions associated with the defective phenotypes; (iv) ChIP-chip (genome-wide *in vivo* localization of TrmB binding); (v) genome-wide distribution of a conserved motif signature discovered *de novo* within experimentally mapped TrmB-binding sites; (vi) promoter: reporter fusion assays to validate TrmB targets identified by the high-throughput methods; (vii) computational integration of the results of these experiments and data from earlier studies to construct transcriptional and metabolic networks governed by TrmB. We conclude from the results of these experiments that TrmB is a bifunctional regulator that governs the transcription of genes in central metabolic pathways of diverse ancestry to manage cellular redox and energy status. Results of these experiments are described in detail below.

Sequence analysis suggests that *VNG1451C* encodes a putative sugar-binding transcription regulator

Given the central nature of sugar metabolism in cellular physiology, we searched for putative TFs that may control this process. At least seven proteins in the *H. salinarum NRC-1* proteome (http://baliga.systemsbiology.net) have significant matches to protein family signatures and sequences of known sugar metabolism regulators. Among these candidate regulators, the VNG1451C amino-acid sequence (Figure 1A) significantly matches (*e*-value=2×10^{-8}) the 50aa TrmB family signature (PF01978, http://pfam.sanger.ac.uk/) with 21% identity to the consensus sequence (Figure 1B). According to ClustalW analysis, VNG1451C possesses at least three active site residues known to be critical for sugar binding in the characterized TrmB orthologs (Krug *et al*, 2006; Kanai *et al*, 2007; Lee *et al*, 2008). Interestingly, although the TrmB signature is conserved across 175 bacterial and archaeal species, no TrmB orthologs have been identified to date in bacteria (Lee *et al*, 2008). TFs of the TrmB family have been implicated in the regulation of maltose and glucose usage in thermophilic archaea (van de Werken *et al*, 2006; Kanai *et al*, 2007; Lee *et al*, 2008). In these archaea, the genetic loci encoding TrmB also harbor genes coding for the maltose and/or trehalose ABC transporters. Notably, these genes are absent in chromosomal vicinity of VNG1451C in the *H. salinarum* genome (Figure 1A). This combined evidence suggests that *VNG1451C* encodes a widely conserved regulator with a putative function related to sugar metabolism.

Phenotypic analysis suggests that TrmB is involved in sugar metabolism and maintenance of redox balance

We investigated the phenotypic consequence of deleting *trmB* in diverse environments. This revealed a severe growth defect in the mutant under nearly every condition tested, including standard growth in rich media, nutrient starvation in defined media, metal depletion and excess, and oxidative stress (Figure 2A; Supplementary Tables 1 and 2; Materials and methods). In addition, the NAD^+/NADH ratio in mid-logarithmic phase Δ*trmB* cultures was, on average, significantly lower than in the parent strain in the absence of glucose (Figure 2B). Wild-type growth rates were recovered by functional complementation *in trans* with a plasmid-borne copy of *trmB* in the Δ*trmB* background, ruling out polar effects of the gene deletion on surrounding genes (Supplementary Table 2). The addition of glucose to the growth media also complemented both the growth defect and NAD^+/NADH ratio imbalance (Figures 2A and B). Partial complementation of the

Figure 1 VNG1451C is a candidate for the regulation of metabolism. (**A**) *VNG1451C* locus view. Gene positions are indicated by the gray arrows, with *VNG1451C* in black. Chromosomal positions in base pairs are indicated by the scale under the genes. TSS, Transcription start site at chromosomal position 1074876 (Koide *et al*, 2009). TTS, transcription termination site at 1076143. Translation starts at 1074878. Blue box indicates in-frame deleted region from 1074899–1075949. Annotation of surrounding genes: *yfkN*, 2′,3′-cyclic-nucleotide 2′-phosphodiesterase; *eif3bd*, translation initiation factor eIF-2B subunit δ; *VNG1448H* and *VNG1453H*, unknown function. (**B**) ClustalW alignment with characterized TrmB homologs is shown. Boxed and labeled regions represent protein-domain architecture. Asterisks below the alignment denote identical residues and dots represent similar residues. Dashes designate gaps in the alignment. Bold italicized residues depict those involved in interaction with non-reducing glucosyl residues of maltose in *T. litoralis* (Krug *et al*, 2006). Residues in larger font (G320 and E326) are those which, when mutated in *T. litoralis*, most drastically reduce binding affinity for substrate (Krug *et al*, 2006). Organism abbreviations are as follows: Tko, *Thermococcus kodakaraensis*; Pfu, *Pyrococcus furiosus*; Hwa, *Halobacterium walsbyi*; Tli, *Thermococcus litoralis*.

mutant growth phenotype was observed in the presence of glycerol. However, the inability of sucrose, galactose, raffinose, maltose, and pyruvate to remedy these defects ($P<0.001$) indicated some nutrient specificity in the function of this regulator (Figure 2A; Supplementary Table 2). Together these observations suggest that the function of TrmB is associated with glucose or glycerol metabolism and potentially linked to the maintenance of cellular redox balance.

Transcriptome analysis reveals that TrmB might regulate functions in diverse metabolic pathways

To characterize the TrmB regulon, genome-wide transcriptional changes were analyzed in the Δ*trmB* deletion background during growth in the presence and absence of glucose (Figure 2C; Supplementary Table 3). On culturing in the absence of glucose, the transcription of 182 genes showed

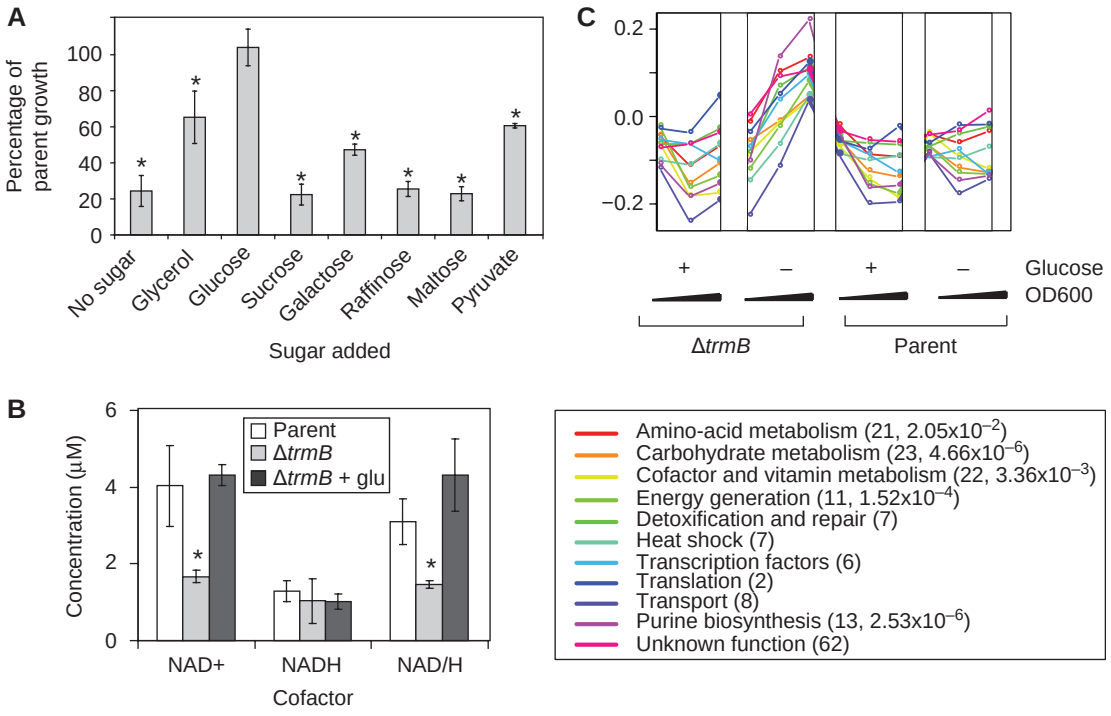


Figure 2 Phenotype of parent versus Δ*trmB* mutant strain during growth. (**A**) Comparison of growth of the parent strain versus mutant in complete defined medium (CDM) supplemented with various sugars (*x*-axis). Ratio of Δ*ura3* parent to Δ*trmB* growth doubling time is represented on the *y*-axis. Error bars represent the s.d. from the mean of at least 3 biological replicate samples, 3 technical replicates each for a total of at least 9 observations and at most 20. Asterisks represent confidence of $P<0.001$ of mutant versus parent in non-parametric paired *t*-tests (i.e. mutant grew significantly slower than parent in all cases except for glucose). (**B**) NAD^+/H assay shows that energy status is perturbed in mutant cells. 'Glu', glucose was added to cultures. Adding glucose to parent strain cultures did not change NAD^+/H levels significantly (not shown). Error bars represent the s.d. from at least three biological replicate experiments. Asterisks represent $P<0.05$ difference between parent and Δ*trmB* in the absence of glucose in non-parametric paired *t*-tests. (**C**) The Δ*trmB* mutant has a defect in expression of genes associated with diverse metabolic processes. Figure represents results of mRNA expression profiling using microarrays. Log10 expression ratios of Δ*trmB* and Δ*ura3* parent versus the common reference RNA (Materials and methods) (*y*-axis) are plotted throughout the growth curve (black triangles beneath graph). The three time points depicted in the expression graph represent RNA collected from cultures in early log, mid-log, and stationary phase relative to parent strain in the presence (+) and absence (−) of glucose (*x*-axis). Colored lines on the graph represent the mean expression profile for genes in each functional category. In the legend, numbers in parentheses indicate the number of genes and the Gene Ontology (GO) *P*-value of enrichment in each category and colors of corresponding categories are indicated. Those categories with no *P*-values shown had no annotations in KEGG or GO and were therefore hand annotated based on protein functional information from databases (Materials and methods).

perturbed expression in the Δ*trmB* background compared with the isogenic parent (16 down and 166 upregulated; Figure 2C; Supplementary Table 3). These genes were grouped into 11 functional categories according to GO and KEGG databases (Ashburner *et al*, 2000; Kanehisa and Goto, 2000). As expected from the aforementioned experiments, genes whose products function in carbohydrate metabolism were significantly over-represented in these categories ($P\sim5\times10^{-6}$; e.g. *ppsA*, PEP synthase; *pykA*, pyruvate kinase; Figure 2C; Supplementary Table 3). Surprisingly, genes encoding additional metabolic pathways were also significantly perturbed in the mutant; including amino acid, cofactor, vitamin, and purine biosynthetic pathways (Figure 2C). Consistent with the phenotypic data, growth phase-specific regulation of these genes was restored in Δ*trmB* on the addition of glucose (Figure 2C). Two possible molecular mechanisms could lead to this result: (i) direct TrmB binding to affected promoters, including those of other TFs; and (ii) an indirect consequence of perturbing sugar metabolism.

TrmB binds target promoters that function in diverse metabolic pathways in the absence of glucose or glycerol

TF-binding location analysis with ChIP-chip

To differentiate between direct and indirect regulatory influences of TrmB, its transcription-factor-binding sites (TFBS) were localized throughout the genome using ChIP-chip. This procedure localizes DNA fragments within transcription factor complexes enriched with chromatin immunoprecipitation (ChIP) using whole genome tiling arrays (chip). We mapped TFBSs in the presence and absence of varying concentrations of glucose or glycerol (Materials and methods). TrmB bound to 113 sites throughout the chromosome in the absence of glucose or glycerol (Figure 3A). Interestingly, no TrmB binding was observed across the genome in the presence of high glucose or glycerol (Figures 3B and C; Supplementary Figures 1 and 2). This finding is consistent with the observation that both glucose and glycerol can complement

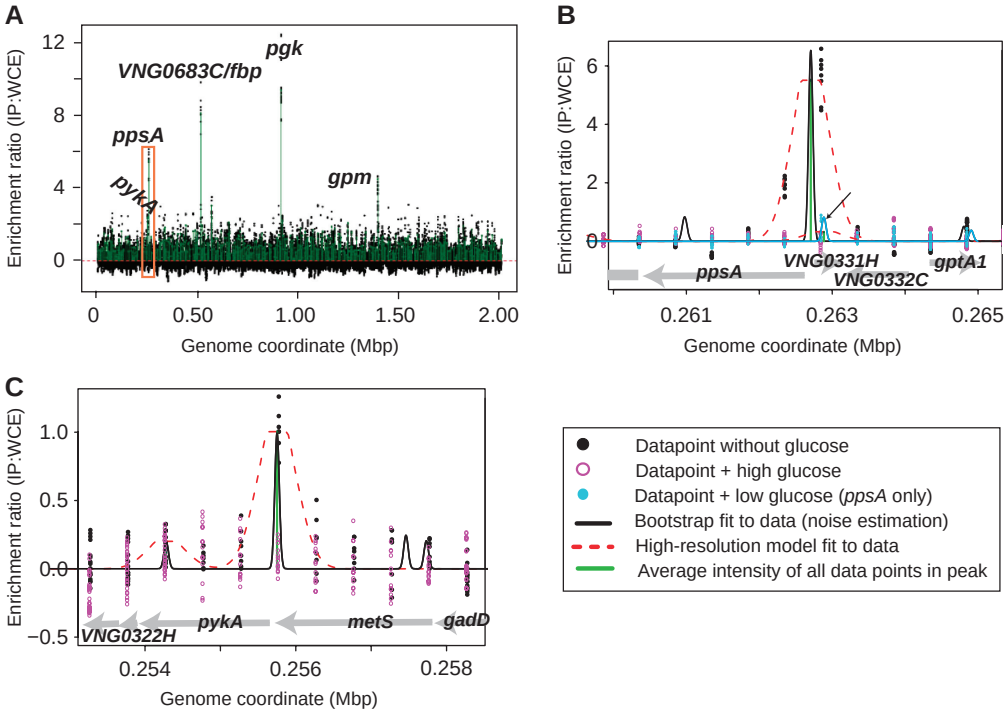


Figure 3 TrmB directly regulates target promoters in a glucose- and glycerol-dependent manner. (**A**) Genome-wide DNA-binding distribution of TrmB in the absence of added glucose. Enrichment ratio of the immunoprecipitated DNA–protein (TrmB) complexes to the mock unprecipitated chromatin is shown on the *y*-axis. Genome position in megabase pairs (Mbp) is displayed on the *x*-axis. Black points are raw intensities from a ChIP-chip experiment representative of the five biological replicates. Green lines represent the MeDiChI-based binding site fit to the data (Reiss *et al*, 2008) (Materials and methods). The brightest peaks are labeled according to the designation of the nearest gene (within 250 bp of the ChIP-chip peak): *pykA*, pyruvate kinase; *ppsA*, PEP synthase (zoomed view of region in orange box shown in (C)); *VNG0683C*, fructose-1,6-bisphosphate aldolase; *fbp*, fructose-1,6-bisphosphatase; *gpm*, phosphoglycerate mutase. (**B**) TrmB may bind with high affinity to the *ppsA* promoter. Graph shows a zoomed view of the genome region bounded by the orange box in (A). Black peak represents the MeDiChI-based bootstrap fit to the black data points (no added glucose). Light blue peak (see arrow) represents the bootstrap fit to the blue data points (low concentration of glucose). Purple data points represent the ChIP-chip data in the presence of high glucose. *x*- and *y*-axes are as in (A). Gray arrows below the data points indicate the gene sizes and direction of transcription. (**C**) TrmB binds with relatively lower affinity to the *pykA* promoter. Graph shows a zoomed view as in (B).

the physiological consequences of deleting TrmB (Figure 2) and with earlier reports on TrmB function in other systems (Lee *et al*, 2008).

The 113 binding sites were ranked according to statistical confidence (Table I). The top seven high-confidence ($P<10^{-6}$) hits reside in intergenic regions upstream of genes encoding functions in glycolysis and gluconeogenesis (functional enrichment P=1.11E−02) (Table I; Figures 3A–C). These genes also exhibited the most significant enrichment in immunoprecipitated TrmB–DNA complexes (*gpm, ppsA, pgk, pykA, VNG0683C*; Figure 3A). In addition, one gene, *ppsA*, remained bound under low glucose concentrations, suggesting a high-affinity interaction with TrmB at this site (Figure 3B).

Other high-confidence targets in the list ($10^{-3}>P>10^{-6}$) also showed a strong transcriptional perturbation in the Δ*trmB* background (Table I; Supplementary Table 4). Consistent with the transcriptome data, genes coding for biosynthesis of purine nucleotides, cobalamin, thiamin, and amino acids were significantly overrepresented across all 113 targets (Table I; Supplementary Table 4). We also observed binding to the intergenic region upstream of five TFs, including VNG0156C, VNG0247C, VNG0878G, putative regulators; VNG1179C, a regulator of copper homeostasis (Kaur *et al*, 2006); and TrmB itself (Supplementary Table 4). This could explain the differential regulation of a large number of genes whose promoters are not directly bound by TrmB. Together these data suggest that TrmB directly controls the expression of genes functioning in diverse metabolic pathways.

Although TF binding in intergenic regions is generally considered as evidence for direct regulation of downstream genes, we found that ∼40% (45 of 113) of TrmB-binding sites were inside coding sequences. However, given that the *H. salinarum* genome is ∼85% coding, our sample of binding sites is actually somewhat enriched in intergenic regions (P=0.18), with 60% of these binding sites falling within 250 bp of an experimentally determined transcription start site (Koide *et al*, 2009). ChIP-chip studies for other bacterial TFs have found up to 70% of targets in intergenic regions (Shimada *et al*, 2008). Combined, these results suggest that these binding events might be functional. However, further investigation is required to elucidate the physiological function of these unusual TrmB targets, at least two of which are near loci that encode newly discovered putative noncoding RNAs (Koide *et al*, 2009).

Identification of a *cis*-regulatory sequence motif

To further define the TrmB-binding site, we searched for a conserved *cis*-regulatory sequence motif within 250 bp of its genomic binding locations identified by ChIP-chip. Locations

Table I High-confidence TrmB target sites from ChIP-chip data

Alias	Annotation	Location[a]	Position[b]	Intensity[c]	*P*-value[d]	TSS[e]	ATG
VNG0683C/fbp[f]	Fructose 1,6-bisphosphate aldolase class I	IG	519847	2.5380	1.85E−16	519909	519914
pgk	Phosphoglycerate kinase	IG	918276	3.6682	2.70E−13	918077	918068
ppsA	Phosphoenolpyruvate synthase	IG	262626	1.7335	3.98E−10	262557	262549
pykA	Pyruvate kinase	IG	255785	1.2462	6.85E−10	255673	255653
gap	Glyceraldehyde-3-phosphate dehydrogenase	IG	714142	1.0375	1.37E−06	714416	714164
gpm	Phosphoglycerate mutase	IG	1397677	2.6769	1.55E−06	1397289	1396895
gapB	Glyceraldehyde 3-phosphate dehydrogenase	IG	81287	2.0801	4.63E−06	ND	81452
cobAI/cbiP	Cobalamin adenosyltransferase	IG	1175752	2.2966	2.24E−05	ND[g]	1176275
aldY2<->VNG0772H/acd2[h]	Aldehyde dehydrogenase (retinol)	IG	582082	0.9210	2.76E−05	581952	581962
aroE/VNG0383H/trpE2/trpG2	Shikimate 5-dehydrogenase	IG	297177	1.2650	3.89E−05	296272	296344
purL2	Phosphoribosylformylglycinamidine synthase I	ORF	1433241	0.6947	5.50E−05	1433519	1433516
lon	Putative protease La homolog type	ORF	238792	1.4968	9.25E−05	237828	237843
korA/korB	Putative 2-ketoglutarate ferredoxin oxidoreductase (α)	IG	858771	0.9335	1.71E−04	858577	858582
ush	5′-nucleotidase	ORF	1046327	0.9919	4.70E−04	1046393*	1048121
VNG0631C/purK/purE	Phosphoribosylaminoimidazole carboxylase ATP-binding subunit	IG	482660	1.1603	1.10E−03	483056	483366
glpK	Glycerol kinase	IG	1452723	1.1001	1.27E−03	1452759	1452694
cobN/cbiCJ	Cobalamin biosynthesis protein	IG	1168174	0.7996	2.16E−03	ND	1167667
cxp/adh2	Probable carboxypeptidase	ORF	1958110	0.4903	3.20E−03	1957532	1957549
gdhA1	Glutamate dehydrogenase	IG	478241	0.8282	5.58E−03	478459	478615
cbiLFGH	Precorrin-2 C20 methyltransferase	ORF	1157269	0.7865	1.33E−02	1156869	1156903

[a]Location of the center of MeDiChI-predicted binding site. IG, intergenic location (i.e. within 250 bp of the transcription start site (TSS)). ORF (open reading frame), binding site located within the coding sequence of the gene.
[b]Position, chromosomal coordinate marking the center of the density peak from all five replicate MeDiChI-predicted binding sites (see Materials and methods).
[c]Intensity, ratio of enrichment of chromosmal location in TrmB immunoprecipitated samples relative to mock precipitated control. Intensities shown represent the mean intensities for five biological replicate experiments.
[d]*P*-values represent the product of *P*-values from five biological replicate samples that were subjected to MeDiChI analysis.
[e]TSS, transcription start site chromosmal coordinate position determined experimentally by high-density tiling array (Koide *et al*, 2009).
[f]Gene names combined with a slash represent genes in operons with the gene closest to the binding site listed first. Annotation for the first gene in the operon is provided for brevity.
[g]ND, transcription start site not detected in high-density tiling array experiments (Koide *et al*, 2009).
[h]aldY2 and acd2 are divergently transcribed with the binding site equidistant from each transcription start site. Data for *aldY2* is provided.

of transcription start sites (Koide *et al*, 2009), translation start sites (http://baliga.systemsbiology.net) (Ng *et al*, 2000), and putative GTF-binding sites (Facciotti *et al*, 2007) were used to constrain the sequence search space (Materials and methods). We identified a conserved *cis*-element [TACT-N (7-8)-GAGTA ($P<2\times10^{-5}$)] (Figure 4A) within 250 bp of 115 ($P=8.7\times10^{-50}$) of all genes nearby TrmB-binding sites identified by ChIP-chip (Figure 4B). Matches to this signature were also detected in the vicinity of other sites, albeit farther from the ChIP-chip location (>250 bp) (Supplementary Table 4). This motif is divergent from other characterized TrmB-binding sites (e.g. Thermococcales glycolytic motif (TGM), TATCAC-N5-GTGATA) (van de Werken *et al*, 2006). However, consensus sequences for different HTH-domain containing TFs can be quite divergent (Rigali *et al*, 2004). A genome-wide pattern searching algorithm identified a total of 317 matches to this motif signature (Supplementary Table 5; Materials and methods) which are nearby 396 genes in operons, suggesting that TrmB may bind to additional loci across the genome. This motif signature is enriched in intergenic regions ($P\sim0.002$). However, given that functional promoter binding is a product of combinatorial interactions of TFs, GTFs, cofactors, and RNA polymerase; further studies will determine which of these additional putative TrmB-binding sites are indeed functional.

In vivo validation of key TrmB-binding sites using promoter: reporter fusion assays

To ensure that the TrmB-binding motif represents a physiologically relevant regulatory region, the *ppsA* promoter was fused to the GFP reporter and assayed in the wild type and Δ*trmB* backgrounds (Figure 4C). As expected from the microarray and ChIP-chip data, FACS assays validated that transcription from the *ppsA* promoter is activated three-fold in the absence of glucose in the parent background, whereas Δ*trmB* cells are impaired for induction (Figure 4C). Strikingly, when the distal TrmB-binding site is removed, the glucose responsiveness is reduced in the parent strain (Figures 4C and D). Activity from this shorter promoter remains at background levels in the Δ*trmB* mutant regardless of condition (Figure 4C). Together these data verify that the *cis*-regulatory TrmB-binding site identified by high-throughput methods is physiologically relevant and requires TrmB for regulation in the absence of glucose. Further, both the promoter proximal and distal binding sites contribute to TrmB-mediated response to glucose, with both binding sites required for full activation, suggesting binding site synergy (Figure 4D).

TrmB governs an integrated transcriptional and metabolic network to balance the expression of evolutionarily diverse cofactor and enzyme-coding genes

TrmB is a bifunctional regulator that activates some targets and represses others

To construct the *H. salinarum* TrmB-dependent transcriptional network, we calculated the significance of the overlap between the integrated ChIP-chip, transcriptome, and motif location data generated here. This was further integrated with genome-wide transcription start site data (Koide *et al*, 2009) and ChIP-chip data for seven GTFs (Facciotti *et al*, 2007). This enabled identification of a significant group of 37 genes (organized among 20 operons) in the overlap between integrated system-wide datasets (Figures 4A and B). Functional annotations (Materials and methods) revealed that these 37 genes encode enzymes of (i) glycolytic and gluconeogenic pathways, (ii) purine biosynthesis, (iii) cobalamin biosynthesis, (iv) TCA cycle, and (v) glutamate dehydrogenase (Figure 4A).

The mechanism by which TrmB activates and/or represses these genes was investigated further by analyzing the locations of its binding sites relative to those of seven GTFs (Facciotti *et al*, 2007) (Figure 5A). TrmB-binding sites were located upstream of GTF-binding sites in promoters of genes it activates (e.g. *ppsA*; Figure 5B; $P<0.03$). In contrast, no GTF-binding locations were detected within promoters of genes that are repressed by TrmB (e.g. *gap*; Figures 5A and C). Alternatively, GTF-binding sites were located upstream of repressed genes (e.g. *VNG0303G*, *VNG0382G*, *VNG1128G*, Figure 5A). Thus, our system-wide ChIP-chip data support a model in which TrmB binds to its motif downstream of the PIC to occlude transcription. In contrast, TrmB-binding upstream of the PIC facilitates transcription at weak promoters (Figure 5D). Although this model was suggested earlier (Kanai *et al*, 2007; Lee *et al*, 2008), this study presents the first *in vivo* experimental evidence of these interactions and places them in a global context. Similar analysis of the remaining TrmB-bound promoters revealed a more complex distribution of GTF and TrmB-binding sites, which may indicate the involvement of additional mechanisms in the regulation of those genes (Facciotti *et al*, 2007). Nevertheless, this integrated systems analysis has provided both a global perspective and valuable mechanistic insight into combinatorial regulation by GTFs and a sequence-specific transcription regulator (Bell *et al*, 1999; Kanai *et al*, 2007).

The combined evidence presented thus far strongly suggests that TrmB acts as both a transcriptional activator and a repressor in response to carbon source availability. Further, these results suggest that TrmB directly and coordinately controls genes significantly overrepresented for functions in central metabolism and its associated pathways in the metabolic network.

Metabolic network reconstruction analysis suggests that TrmB coordinates the expression of evolutionarily diverse enzymes

To gain a systems-scale perspective on the role of TrmB in metabolism, we reconstructed the metabolic network of *H. salinarum* in the context of the TrmB transcription regulatory network and known archaeal reactions reported in the literature (Figure 6; Supplementary Table 6). Many TrmB-regulated enzymes catalyze reactions at critical regulatory branch points but not the end stages of several metabolic pathways, including amino acid, purine, thiamine, cobalamin biosynthesis. (Figure 6; Falb *et al*, 2008; Gonzalez *et al*, 2008). This is illustrated by TrmB-mediated transcriptional control of genes encoding branch points between purine and thiamine metabolism (purK, purE, and purM; reactions 46 and 47, Figure 6), carbon and nitrogen metabolism (gdhB, korAB, and gdhA1; reactions 31, 32, and 39, respectively, Figure 6)

A

Gene/operon	ORF	Motif	Annotation
gapB	*VNG0095G*	cgcgca**GACTC**atactcg-**TAGGA**tcgAcc-107-ATG	Glyceraldehyde 3-phosphate dehydrogenase
lon	*VNG0303G*	aaggggg**TACGG**ctacgag-**GTGTA**catggac	Putative La protease homolog
pykA	*VNG0324G*	gaatgg**AACTC**agtatc--**GAGTA**aagagc-35-ATG-N24-ct**c**gg	Pyruvate kinase
ppsA P1	*VNG0330G*	gccggg**TAAAA**cccatc--**GAGTA**gaaacg-84-**a**cATGgctgtaCgct (p1)	Phosphoenolpyruvate synthase
ppsA P2	*VNG0330G*	aacgac**AACTC**ggttcC--**GAGTA**ccatat-60-**a**cATGgctgtaCgct (p2)	
aroE/VNG0383H /trpE2/trpG2	*VNG0382G*	ga**t**cga**AACCA**ggcagc--**GAGTA**aaaaac-46-ATG	Shikimate 5-dehydrogenase
gdhA1	*VNG0628G*	ccgtca**TGCTT**attcgtt-**g**AGTTtgaaag-143ATG	Glutamate dehydrogenase
VNG0631C/purK /purE	*VNG0631C*	gggcaa**TTCTT**atgcag-**g**GTGTAggtggg-68-ATG	Phosphoribosylaminoimidazole carboxylase ATP-binding subunit
VNG0683C	*VNG0683C*	aacgtt**CACTC**agaaca-**GAGTT**aaaacG-58-**c**aatc-14-ATG	Fructose 1,6-bisphosphate aldolase class I
aldY2<->VNG0772H/acd2	*VNG0771G*	tcggcg**TGCTC**gtgttg--**GATTA**cctgc-522-ATG-9-c**c**a	Aldehyde dehydrogenase (retinol)
gap	*VNG0937C*	gaatct**TTCTT**cggtatt-**GAGTA**aactct-27-ATG	Glyceraldehyde 3-phosphate dehydrogenase, NAD dependent
korAB	*VNG1128G*	acgatt**TAGTG**gcgacc**a**c**GCGTA**cg--------ATGcc	Putative 2-alphaketogluterate ferredoxin reductase
pgk	*VNG1216G*	accctt**TACTC**ggtccc--**GAGTA**-64-**c**cgatacATG	Phosphoglycerate kinase
ush	*VNG1408G*	cgaccg**TGCTG**aacgag--**GAGTA**cacgct	5′-nucleotidase
sucDC	*VNG1542G*	gccgtc**GACTT**caccaac-**GAGTT**gtacgtc	Succinyl-CoA synthetase
cbiLFGH	*VNG1551G*	gcgacc**TCCTT**cgactcc-**GAGAA**ccccgt-45-ATG	Precorrin-2 C20 methyltransferase
cobN/cbiCJ	*VNG1566G*	ttgtaa**TAC**t**c**caatg---**GAGGT**ttaacag-4-GTG	Protoporphyrin IX magnesium chelatase
cobAI/cbiP	*VNG1574G*	ggggtc**TACCT**gcccggg-**GGGTA**ccccga-441ATG	Cobalamin adenosyltransferase
purL2	*VNG1945G*	ggcggg**TTCTC**gtacggc-**GACTA**cctctg	Phosphoribosylformylglycinamidine synthase I
glpK	*VNG1967G*	actgcg**TACCG**gcccccct-**GAGTG**-305-**t**-64-ATG	Glycerol kinase
cxp/adh2	*VNG2616G*	ctcatc**GACCT**caaacgg-**GAGTA**cgctaa	Probable carboxypeptidase/aldehyde dehydrogenase

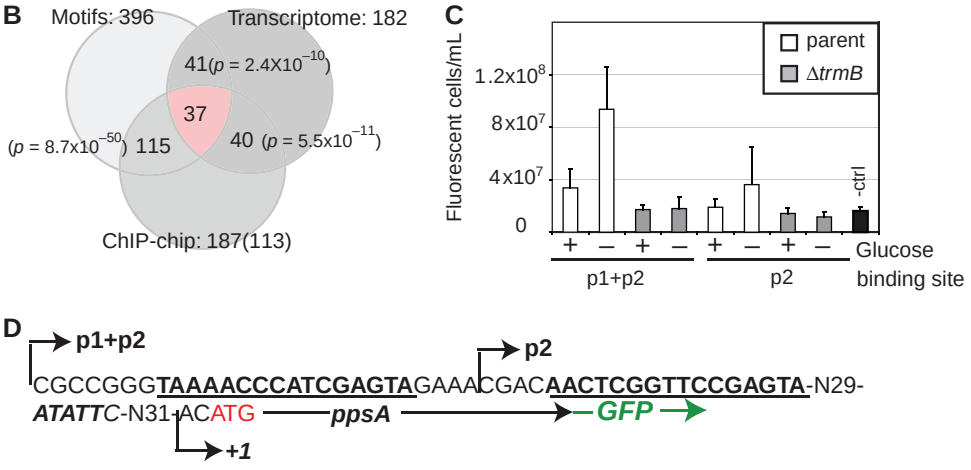


Figure 4 TrmB binds to an inverted palindromic sequence in the promoters or ORFs of target genes. (**A**) Table shows the location of the motif in the vicinity of high-confidence TrmB-binding sites. Genes in operons are separated by a slash. Divergently transcribed genes are separated by a double-headed arrow. The TrmB-binding motif is indicated in uppercase bold text, experimentally determined transcription start site (TSS) in lowercase bold italic (Koide *et al*, 2009), and translation start site (ATG) is underlined. The location of the center of the ChIP-chip hit nearby the motif is designated in uppercase unbolded font. Annotations of translation start site and gene function are derived from the *H. salinarum* complete genome sequence database (baliga.systemsbiology.net). Numbers within each sequence indicate the number of nucleotides between the motif and downstream elements (e.g. TSS and ATG). Many transcripts in *H. salinarum* are thought to be devoid of 5′ untranslated regions (UTR) (Brenneis *et al*, 2007), hence the close spacing of the TSS and ATG for some genes. Dashes are inserted for the purposes of motif alignment because of variable gap spacing between motif half sites. In those cases in which no ATG is listed, the ChIP-chip site and motif were located in middle of the coding sequence. (**B**) Integration of high-throughput experimental data reveals a set of 20 high-confidence targets of TrmB, which corresponds to 37 genes in operons. These 37 genes are represented in the table (A). The intersection (red) of genes identified by motif searches, transcriptome data, and ChIP-chip datasets is shown in the Venn diagram.187 genes, including operon members, were designated as TrmB targets because they fell within 250 bp of the 113 TrmB-binding sites identified through ChIP-chip. *P*-values represent the likelihood that the pairwise overlap between datasets is due to chance. (**C**) Validation of TrmB-binding site using GFP promoter: reporter fusions. GFP expression levels driven by the two different *ppsA* promoter constructs is shown (p1 + p2, contains both TrmB-binding sites; p2 contains only the promoter proximal motif). White bars indicate expression levels in the parent strain, whereas gray bars depict expression in the Δ*trmB* mutant. Black bar labeled '−ctrl' (negative control) represents GFP expression background in the vector construct devoid of a promoter region (Supplementary information). Glucose growth conditions are shown on the *x*-axis, and the number of fluorescent cells counted (normalized to spiked-in fluorescent beads) is shown on the *y*-axis. Error bars represent the s.d. from the average of three independent biological replicate experiments. (**D**) Sequence of the *ppsA* promoter constructs fused to GFP. Bent arrows depict the beginning of the p1 + p2 long promoter construct, p2 truncated construct, and transcription start site, respectively. Translation start site is depicted in red text.

A

ORF	GENE	Δ*trmB* mRNA	GTF-TrmB	A/R
VNG0095G	*gapB*	–0.0315	6	A
VNG0303G	*lon*	0.178	–115	R
VNG0324G	*pykA*	0.278	—	R
VNG0330G	*ppsA*	–0.2115	65	A
VNG0382G	*aroE/VNG0383H /trpE2/trpG2*	0.195	–14	R
VNG0628G	*gdhA1*	–0.0465	—	
VNG0632G	*VNG0631C/purK /purE*	0.3245	—	R
VNG0683C	*VNG0683C*	–0.099	125	A
VNG0771G	*aldY2<-> VNG0772H/ acd2*	0.091	—	R
VNG0937G	*gap*	0.6585	—	R
VNG1128G	*korAB*	0.1215	–15	R
VNG1216G	*pgk*	–0.085	5	A
VNG1408G	*ush*	0.1455	—	R
VNG1542G	*sucDC*	0.3935	–284	R
VNG1551G	*cbiLFGH*	–0.05	—	
VNG1565G	*cobN/cbiCJ*	0.098	—	R
VNG1574G	*cobAI/cbiP*	0.2135	—	R
VNG1945G	*purL2*	0.3195	—	R
VNG1967G	*glpK*	0.0095	55	
VNG2616G	*cxp/adh2*	0.143	—	R

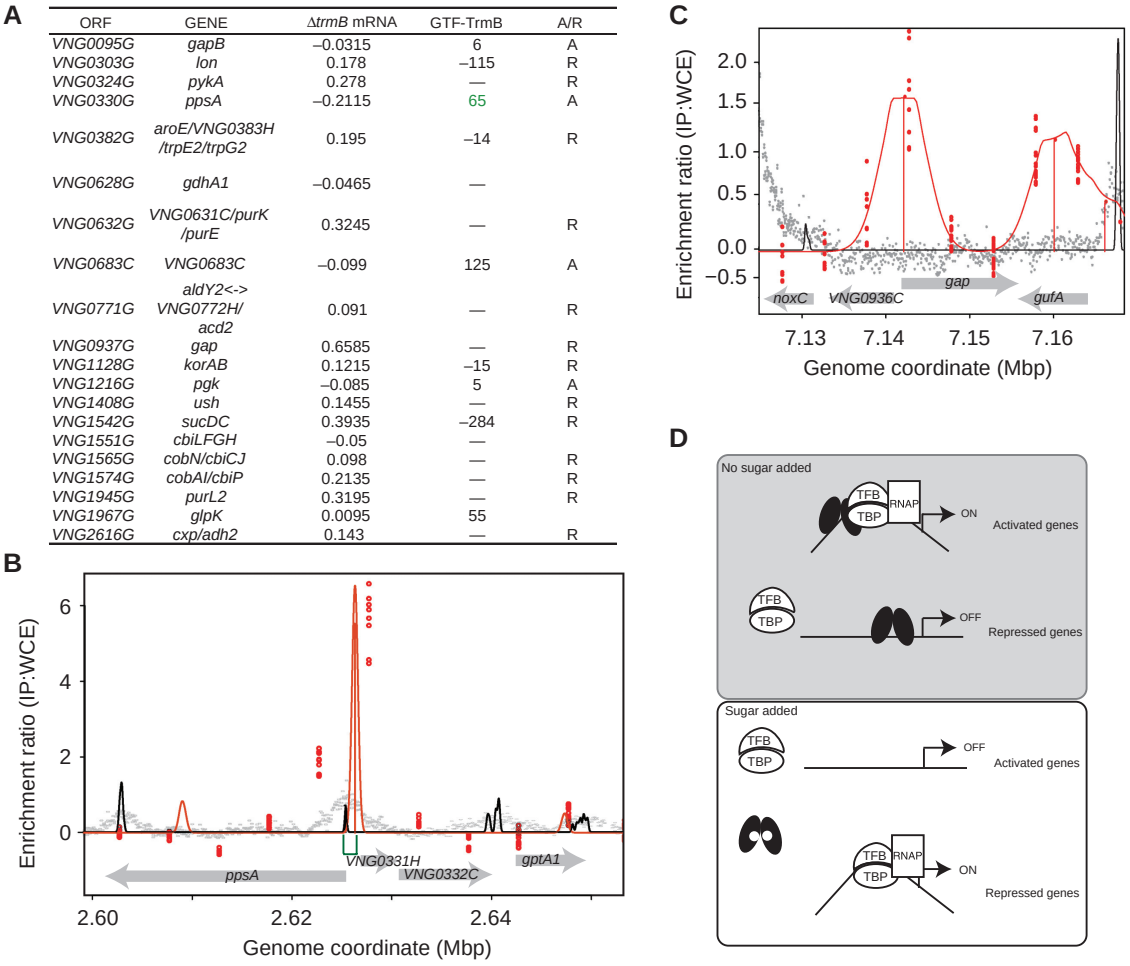
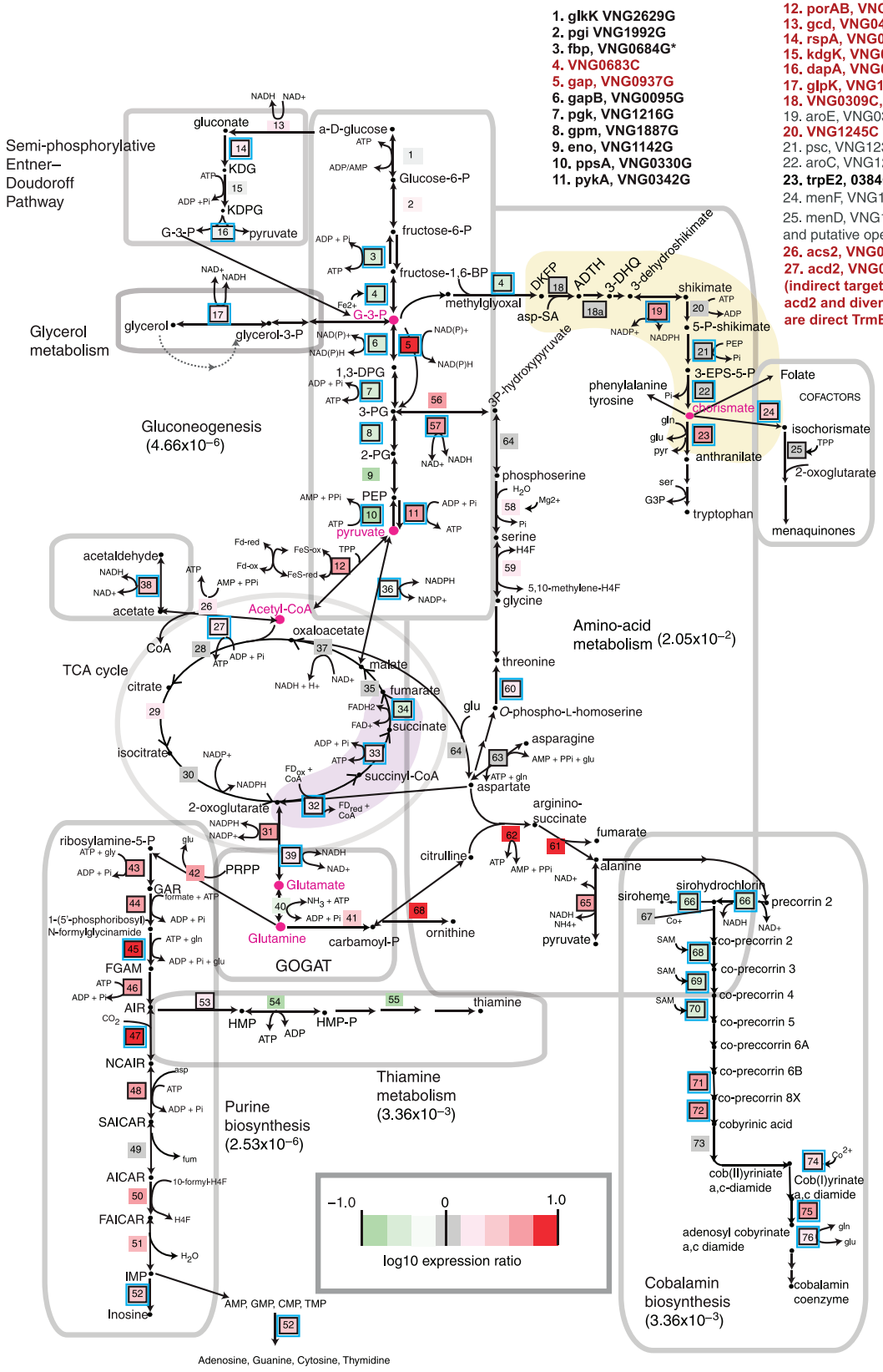


Figure 5 Co-occurrence of TrmB and general transcription factors (GTFs)-binding sites provides *in vivo* evidence for the putative TrmB mechanism. (**A**) Correlation of mRNA expression and binding site locations of TrmB target genes. The Δ*trmB* mRNA column lists the log10 expression ratio for each gene in the Δ*trmB* mutant strain in the absence of glucose in stationary growth phase. GTF–TrmB column lists the number of base pairs separating the two motifs (i.e. the chromosomal coordinate of the TrmB-binding site subtracted from that of the GTF site). —, no GTF-binding site was identified within 250 bp of the TrmB site. A/R, activated or repressed. These A/R calls were made based on the significant correlation between the distance separating binding sites and transcription ratio across the 37 target genes considered ($P < 0.03$; see Materials and methods). Gene names separated by slashes indicate operon memberships and double-headed arrow represents divergently transcribed genes. (**B**) Example ChIP-chip data for a TrmB-activated gene, *ppsA*. Data points representing enrichment intensities for TrmB are shown in red and that of TfbD is shown in gray. Red lines represent the MeDiChI-based fit to the data for TrmB and black depicts the fit to the TfbD data. Gray arrows beneath the chart represent the locations and transcriptional directions of open reading frames (ORFs). (**C**) Example ChIP-chip data for a TrmB-repressed gene, *gap*. Colors and labels are as in (B). (**D**) Model for the TrmB mechanism of interaction with the general transcription factor machinery (TFB/TBP) at target promoters. Black ellipses represent TrmB. Small white circles inside TrmB represent the binding effector, which enables TrmB to dissociate with the DNA in the presence of glucose. TFB, transcription factor II B; TBP, TATA-binding protein; RNAP, RNA polymerase.

We also observed several instances of direct TrmB-mediated transcriptional control of metabolic enzymes with the biosynthesis of their cognate cofactors. Three examples support this assertion. (i) TrmB directly controls over 10 enzymes that require adenosine phosphates (AXP; Figure 6, e.g. reactions 3, 7, 10, 11, 27, 33, 63) and six genes that encode the biosynthesis of these cofactors (Figure 6, reactions 43, 44, 45, 46, 47, 52). (ii) TrmB target genes encoding functions such as 5-phosphoribosyl-*N*-formylglycinamidine (FGAM) synthase (*purL2*; Figure 6, reaction 45) and asparagine synthase (*asnA*, reaction 63) use glutamate or glutamine as co-reactants and ATP as a cofactor. Concomitantly, TrmB regulates GOGAT pathway genes (reactions 31, 39). (iii) TrmB coordinately controls thiamin (TPP) synthesis (*thiC*) with thiamine-requiring enzymes (*porAB* and *menD*, reactions 12 and 25, respectively) (Rodionov *et al*, 2003).

In addition, among direct TrmB targets were genes encoding enzymes of uniquely archaeal lineage (Figure 6; Supplementary Table 6). Of the enzyme-coding genes under TrmB control included in the metabolic pathways shown in Figure 6, 13 are unique to archaea (Supplementary Table 6) and 35 are conserved across species from all three domains of life. Integrated analysis of the metabolic and gene regulatory network architecture reveals two opposing scenarios. (i) In the



1. glkK VNG2629G
2. pgi VNG1992G
3. fbp, VNG0684G*
4. VNG0683C
5. gap, VNG0937G
6. gapB, VNG0095G
7. pgk, VNG1216G
8. gpm, VNG1887G
9. eno, VNG1142G
10. ppsA, VNG0330G
11. pykA, VNG0342G
12. porAB, VNG0473/4G
13. gcd, VNG0446G
14. rspA, VNG0442G
15. kdgK, VNG0158G
16. dapA, VNG0444G*
17. glpK, VNG1967G
18. VNG0309C, 18a. VNG0310C*
19. aroE, VNG0382G
20. VNG1245C
21. psc, VNG1232G
22. aroC, VNG1230C*
23. trpE2, 0384G*; trpG2, VNG0386G*
24. menF, VNG1083G
25. menD, VNG1081G*
and putative operon menBACE*
26. acs2, VNG0997G
27. acd2, VNG0775G,
(indirect target acs3, VNG0940G).
acd2 and divergently transcribed aldY2
are direct TrmB targets.
28. citZ, VNG2102G
29. can, VNG2574G
30. icd, VNG1873G
31. gdhB, VNG0161G
32. korAB, VNG1125/8G
33. sucCD, VNG1541G
and VNG1542G
34. sdhABCD,
VNG1306-1310G
35. fumC, VNG1356G
36. mdh, VNG1624G
37. mdhA, VNG2367G
38. aldY2, VNG0771G
39. gdhA1, VNG0628G
40. glnA, VNG2093G
41. carAB, VNG1814/5G
42. purF, VNG1493G
43. purD, VNG1305G
44. purU, VNG1946G
45. purL2, VNG1945G;
purL, VNG0864G;
46. purM, VNG0876G
47. purK/E, VNG0632/3G
48. purC, VNG1939G
49. purB, VNG0415G
50. purH, VNG0414G
51. purO, VNG2371C
52. ush, VNG1408G
53. thiC, VNG0715G
54. thiD, VNG2606G
55. thi1, VNG2604G
56. serA1, VNG2424G
57. serA3, VNG0104G
58. serB, VNG2423G
59. glyA, VNG1414G
60. thrC2, VNG0541G
61. argH,VNG2436G
62. argG,VNG2437G
63. asnA, VNG0867G
64. aspC1, VNG2418G/
aspC2, VNG1121G
65. ocd2, VNG1364G
66. VNG1775C
67. cbiX, VNG1561C
68. cbiL, VNG1551G*
69. cbiH1/cbiH2,
VNG1555/7G*
70. cbiF, VNG1553G*
71. cbiT/cbiJ,
VNG1550G/1568G*
72. cbiC, VNG1567G*
73. cbiA, VNG1573G
74. cobN, VNG1566G
75. cobA, VNG1574G*
76. cbiP, VNG1576G*

shikimate biosynthesis pathway, only one gene encoding shikimate kinase (*VNG1245C*, reaction 20) is of archaeal origin, whereas all other genes are conserved throughout evolution (orange shaded area, Figure 6; Supplementary Table 6). Strikingly, all genes of this pathway *except* shikimate kinase are direct TrmB targets (Figure 6), suggesting that shikimate kinase may have been acquired by homologous gene replacement (Galperin and Koonin, 1999). This finding is especially surprising given that enzymes in bacterial metabolic pathways without branch points tend to be co-regulated (Seshasayee *et al*, 2008). (ii) The latter half of the TCA cycle from 2-oxoglutarate to fumarate is directly TrmB-dependent (purple shaded area, Figure 6), *including* the genes encoding 2-oxoglutarate oxidoreductase, the only uniquely archaeal enzyme in the pathway (Supplementary Table 6). These findings suggest that TrmB governs the transcription of a metabolic network with hybrid evolutionary origin.

Discussion

TrmB coordinately regulates metabolic enzyme-coding genes with cofactor genes

From the evidence presented in this study, we conclude that TrmB governs a sugar-responsive global metabolic regulatory network to coordinate the expression of genes with diverse evolutionary ancestry. Remarkably, TrmB coordinates the transcription of enzyme-coding genes involved in the synthesis of cofactors required for the function of these metabolic enzymes. We hypothesize that this TrmB-directed coordination may enable redox and energy balance. Specifically, our data suggest that TrmB represses the semi-phosphorylative Entner–Doudoroff (E–D) glycolytic pathway (Kanai *et al*, 2007; Danson *et al*, 2007; van der Oost and Siebers, 2007; Pfeiffer *et al*, 2008) (e.g. *gap, pykA, VNG0442G*) (Figures 5 and 6), and induces gluconeogenesis (e.g. *ppsA*; Figure 2). Thus, a deletion in *trmB* would lead to an inability to generate energy through gluconeogenesis in the absence of glucose. Secondly, Δ*trmB* cultures grown in the absence of glucose overexpress genes coding for enzymes that convert $NAD(P)^+$ to NAD(P)H, this may force the cell toward an oxidized state (e.g. Figure 6, reactions 19, 31). If so, then this would lead to a shortage of reducing equivalents. The hypersensitivity to oxidative stress and reduced NAD/H ratio observed in Δ*trmB* mutant cells are consistent with this hypothesis (Supplementary Table 2; Figure 2B). We conclude that TrmB acts to maintain redox and energy balance in response to nutrient availability in *H. salinarum*.

Interestingly, our evidence points to additional regulatory mechanisms that may cooperate with the TrmB transcription network. First, TrmB does not seem to regulate the end stages of several pathways (e.g. amino acid, purine, thiamine, cobalamin biosynthesis). Second, TrmB is not only autoregulated, but also seems to directly control four other TFs (Supplementary Table 4). Third, the transcription of some genes with a direct TrmB–promoter interaction remains unchanged in the Δ*trmB* strain (e.g. amino-acid biosynthesis gene *asnA*). Finally, in several instances TrmB regulates an entire pathway except for one gene (e.g. enolase, shikimate kinase). Together, this evidence suggests the cooperation of other regulatory mechanisms with the TrmB transcription network. This interpretation is in line with earlier metabolic regulatory network analyses in *Escherichia coli*, which found that the majority of central metabolic pathways are controlled by multiple TFs (Seshasayee *et al*, 2008).

Evolutionary context for the TrmB transcription-metabolism network

Among known global regulators of central metabolic genes in prokaryotes, no single transcription factor has been shown to directly control both metabolic enzyme and cognate cofactor biosynthesis genes (Grainger *et al*, 2005; Supplementary Table 4). For example, CRP, a global regulator of carbon and nitrogen metabolic pathways in enteric bacteria, is required for the condition-specific transcriptional induction of cobalamin biosynthesis genes (*cob*). However, a direct CRP–*cob* promoter interaction has not been established (Ailion *et al*, 1993; Grainger *et al*, 2005). Instead, the *cob*-specific transcription factor PocR may be a more likely candidate for direct regulation (Ailion *et al*, 1993). Similarly, in *Bacillus subtilis*, CcpA controls global targets in carbon metabolism (Sonenshein, 2007), whereas the PurR repressor is specific for purine biosynthetic genes (Saxild *et al*, 2001).

In contrast, in other archaeal species, it is possible that TrmB or TrmB-type global transcriptional control of metabolism is operative. For example, in species in which TrmB is known to control glycolysis and gluconeogenic pathways (e.g. *Pyrococcus furiosus*), an unknown transcription factor upregulates genes in other metabolic pathways such as the TCA cycle and chorismate synthesis in response to maltose (Schut *et al*, 2003). In addition, divergent TrmB-binding sites can be

Figure 6 TrmB is a global bifunctional regulator, which coordinates the expression of evolutionarily diverse metabolic enzyme genes. Pathway diagram represents the reconstructed TrmB-specified metabolic network for *H. salinarum* (see Supplementary Figure 3 and Supplementary information for reconstruction method). Numbered, colored boxes represent mRNA expression level for each enzyme-encoding gene in the Δ*trmB* mutant in the absence of glucose. Numbers correspond to gene names listed to the right. Gene names shown in bold black type represent those whose encoded enzymes have been biochemically characterized in archaea (see also Supplementary Table 6 for details on enzyme functions). Gene names in bold red indicate those that are indicated by the literature to be unique to the archaeal domain (Supplementary Table 6). Asterisks next to gene names denote a gene in the second or higher position in an operon directly controlled by TrmB. Red boxes represent those that are induced in the Δ*trmB* knockout background in the absence of glucose; green, repressed; the intensity of the color corresponds to the extent of change (see *legend*). Boxes bounded by black lines are direct targets of TrmB (according to ChIP-chip data), and those bounded by blue lines represent targets for which a TrmB-binding motif was identified. Magenta dots indicate central metabolic intermediates. Dotted lines between glycerol and glyceraldehyde-3-phosphate represent putative alternative pathways for entry of glycerol into the trunk portion of glycolysis (Gonzalez *et al*, 2008). Not all 113 TrmB targets are listed in the figure for the sake of clarity. Explanations for abbreviations of metabolic intermediates are listed in Supplementary Table 6. Purple and orange shaded areas represent examples of pathways with mixed evolutionary ancestry (see Results).

detected in the vicinity of the transcription start site for some of these genes in this and other thermophilic archaeal species, although they have not been experimentally validated (van de Werken *et al*, 2006). It will be interesting to confirm these preliminary findings, because it suggests that the network motif of integrated control of cofactor and enzyme genes could be widespread in archaea. Therefore, the metabolic network model presented here will be useful as a structural framework for other archaeal systems and a starting point for evolutionary comparisons with other understudied representatives in other domains of life.

In light of this evolutionary context, it was striking to observe that genes encoding enzymes of uniquely archaeal lineage were included among the direct TrmB targets in *H. salinarum* (Figure 6; Supplementary Table 6). Combined with the observation that the TrmB regulon genes encode enzymes of diverse ancestry, these results are consistent with the hypothesis that several lateral gene transfer or homologous gene replacement events occurred in the evolutionary compilation of the TrmB network (Galperin and Koonin, 1999). However, additional information is required to determine whether the TrmB regulatory network architecture (e.g. promoter elements, nutrient responsiveness) was in place before or after the acquisition of these new genes, because our data support both scenarios (e.g. compare the opposing shikimate pathway and TCA cycle examples. See shaded areas in Figure 6). Nevertheless, the TrmB regulatory network may represent a unique window into an active evolutionary process.

Conclusion

In summary, this study reveals that TrmB regulatory control is restricted primarily to central metabolism and branch points, suggesting combinatorial control between TrmB and pathway-specific regulatory mechanisms. In addition, TrmB seems to balance the expression of genes coding for metabolic enzymes with those of their cognate cofactors in a sequence-specific manner. Finally, the TrmB metabolic regulatory network is an evolutionary mosaic, controlling genes coding for uniquely archaeal enzymes with those that are more widely distributed, even within the same metabolic pathway.

Materials and methods

Strains, media, plasmids, and growth curve assays

H. salinarum NRC-1 (ATCC700922) was grown routinely in complex medium (CM; 250 NaCl, 20 g/l $MgSO_4$ 7H2O, 3 g/l sodium citrate, 2 g/l KCl, 10 g/l peptone) or a complete defined medium (CDM) containing 19 amino acids (Supplementary Table 1). For growth in CDM, starter cultures were grown in CM to mid-logarithmic phase and washed three times in basal salts buffer (CM lacking peptone) and resuspended in CDM at OD600 ~0.1. Subsequent growth was conducted for all media conditions at 37°C with 225 r.p.m. shaking in the presence of low light intensity (24.6 μmol photons/m^2 s from fluorescent lamps).

For routine culturing and growth assays of the Δ*ura3* parent and Δ*VNG1451C* mutant strains, CM or CDM was supplemented with 0.05 mg/ml uracil. For growth curve assays, NAD^+/H assays and growth complementation experiments in the Δ*ura3* parent and Δ*VNG1451C* mutant strains, CM or CDM was supplemented with various sugars (Figure 2) at 7% w/v except for glycerol at 0.08% v/v. Samples were grown in 200 μl cultures for 6 days under continuous ~225 r.p.m. shaking in a Bioscreen C (Growth Curves USA, Piscataway, NJ), set to measure optical density (OD) at 600 nm automatically every 30 min for 200 culture samples simultaneously. The average and standard deviation of the doubling time for Δ*ura3* and Δ*ura3*Δ*VNG1451C* during the logarithmic phase is shown in Figure 2, which represents at least 9 and at most 20 biological replicate samples for each strain under each condition. Two sets of non-parametric paired *t*-tests were performed comparing (i) wild type versus mutant growth (asterisks in figure represent this set of tests); and (ii) mutant growth without carbon source versus mutant growth in each of the carbon sources shown in Figure 2. By the latter measure, mutant growth rates in the absence of carbon source were significantly different from growth in glycerol and glucose but not for any other source tested. For TFBS location array experiments with this strain, glucose was either omitted or added at 0.01% or 7% w/v or glycerol at 0.08% v/v where indicated in the text. Strain constructions and NAD^+/H assays are described in detail in Supplementary information.

Gene expression arrays

10 ml of *H. salinarum NRC-1* Δ*ura3*Δ*VNG1451C* and Δ*ura3* parent strain sample cultures grown in CM or CDM in the presence or absence of glucose were collected at three time points throughout the growth curve (OD_{600} ~0.2, 0.6, and 1.2). Cells were immediately pelleted by room temperature centrifugation at 8820 *g* for 8 min at 4°C and snap-frozen on a dry-ice ethanol bath. Sample pellets were stored overnight at −80°C, followed by RNA preparation using the Absolutely-RNA kit (Stratagene, La Jolla, CA) according to the manufacturer's instructions. RNA quality was checked using the Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA) and freedom from DNA contamination was ensured by PCR amplification of 200 ng of RNA sample. 5 μg of each quality-checked RNA sample was hybridized against the *H. salinarum NRC-1* reference RNA prepared under standard conditions (mid-logarithmic phase batch cultures grown at 37°C in CM). This common reference RNA has been used across all ~950 microarray experiments in the *H. salinarum NRC-1* microarray data repository (Bonneau *et al*, 2007). Samples were hybridized to a 70-mer oligonucleotide array containing the 2400 nonredundant open reading frames (ORFs) of the *H. salinarum NRC-1* genome as described in Baliga *et al* (2004). Each ORF was spotted on each array in quadruplicate and dye flipping was conducted (to rule out bias in dye incorporation) for all samples, yielding eight technical replicates per gene per time point. At least two independent biological replicates exist for all experimental conditions for a total of 16 replicates per gene per condition. Direct RNA or DNA (TFBS location arrays, see below) labeling, slide hybridization, and washing protocols were performed as described earlier (Facciotti *et al*, 2007; Schmid *et al*, 2007). Raw intensity signals from each slide were processed by the SBEAMS-microarray pipeline (Marzolf *et al*, 2006) (www.SBEAMS.org/microarray), in which resultant data were median normalized and subjected to significant analysis of microarrays (SAM) and variability and error estimates (VERA) analysis. Each data point was assigned a significance statistic, λ, using maximum likelihood (Ideker *et al*, 2000).

Microarray data were analyzed using the TM4 MultiExperiment Viewer (MeV) application (http://www.tm4.org/) within the Gaggle data analysis environment (Shannon *et al*, 2006). Specifically, all 2400 genes across mutant and wild-type microarray experiments, described above, were subjected to three independent analyses: significance analysis of microarrays (SAM), KMEANS clustering, and hierarchical clustering. Resultant clusters of genes in the union of all three analyses that displayed significant differential transcription between the parent and mutant strains in the presence and/or absence of glucose were considered to be *VNG1451C* dependent. Biological replicates were considered independently to ensure statistical rigor.

TFBS location array analysis

ChIP of *VNG1451C*-cmyc-tagged constructs was performed as described (Ren *et al*, 2000; Facciotti *et al*, 2007) in cultures grown in the presence or absence of glucose or glycerol. Resultant TFBS location data were analyzed for statistically significant enrichment of features in the ChIP-chip sample versus the unenriched sample using MeDiChI,

a regression-based deconvolution algorithm (Reiss *et al*, 2008). Enrichment lists from each of the five independent MeDiChI runs were combined into a density algorithm (Koide *et al*, 2009) to find TFBS locations overrepresented in the data. To be considered as part of the final TFBS enrichment list (Supplementary Table 4), we required that each enrichment peak from the density output be composed of at least two biological replicate peaks with a combined MeDiChI *P*-value <0.001 (product of replicate *P*-values). A peak was considered to be 'intergenic' if it fell within 250 bp of a transcription start site or termination site (a conservative estimate of the resolution of the data from MeDiChI-derived binding sites; (Reiss *et al*, 2008)). Subsequently, the resultant binding sites from the combined dataset were compared with orthogonal datasets (i.e. genome-wide mRNA expression and binding motif searches, details below).

To analyze the TrmB TFBS location data in the context of the GTF TFBS data (Facciotti *et al*, 2007), the distance of the genomic position for each high-resolution GTF-binding site (Reiss *et al*, 2008) to that of TrmB (GTF coordinate—TrmB coordinate=relative position) at each target promoter was calculated. The Pearson correlation between this distance and the mRNA expression data in Δ*trmB* for the gene of interest was then calculated. The *P*-value for these correlations reported in the text were calculated based on 100 000-fold resamplings of the data.

DNA-binding motif searching

To find the consensus binding motif for *VNG1451C*, the sequence search space was limited for each putative promoter region enriched in the TFBS location data through several constraints: (i) sequence ± 250 bp from the center of each MeDiChI-based peak; (ii) sequence ± 20 bp from the putative transcription start site (Koide *et al*, 2009); (iii) annotated translation start site location (Ng *et al*, 2000). Resultant sequences were used as input for three independent motif-finding algorithms: (i) Bioprospector (http://ai.stanford.edu/~xsliu/BioProspector/), which finds gapped motifs in query sequences (Liu *et al*, 2001); (ii) MEME/MAST (http://meme.sdsc.edu/meme/) (Bailey and Gribskov, 1998; Bailey *et al*, 2006); and (iii) RSAT pattern finding (http://rsat.ulb.ac.be/rsat/). Only motifs represented in the intersection of all the three algorithm outputs were considered in further analysis. To generate the *P*-value for each motif, the three algorithms were re-run on randomized query sequences. Results were compared with algorithm outputs from original sequences using the Wilcoxon test (Frith *et al*, 2008). One motif had a statistically significant *P*-value ($P=2.0\times10^{-5}$), and the resultant consensus motif described in the text was generated using weblogo.berkeley.edu/logo.cgi. To scan the remainder of the genome for the resultant motif, we used the pattern-finding program at rsat.ulb.ac.be/rsat/ with the parameters of (i) no more than 1 bp away from the consensus motif, (ii) unbiased for genomic position (i.e. coding *and* noncoding sequences were searched); (iii) containing a 7 or 8 bp gap within the motif; (iv) located on the chromosome (as no TrmB hits were found on either of pNRC100 or pNRC200).

GFP promoter: reporter fusion validation experiments

The *ppsA* p1 + p2 construct contains both putative TrmB-binding sites in a 115-bp fragment upstream of the translation start site of the *ppsA (VNG0330G)* gene (Figure 4D). The *ppsA* p2 construct is an 88-bp truncated version of the p1 + p2 construct and lacks the promoter-distal TrmB-binding site (Figure 4D). These promoter fragments were fused to a red-shifted GFP variant optimized to function in haloarchaea, which was adapted from (Reuter and Maupin-Furlow, 2004). See Supplementary information for details on strain construction. Constructs were transformed into the Δ*ura3* parent strain and the Δ*trmB* deletion mutant. Resultant constructs were grown in CM media in the presence or absence of 7% glucose to mid-logarithmic phase ($OD_{600}\sim0.4$–0.8). Cultures were diluted to $OD_{600}=0.2$ and fixed in 0.25% formaldehyde dissolved in basal salts (CM lacking peptone) for 10 min at 4°C and subsequently washed in basal salt to remove fixative. Fluorescence of fixed cells was measured by flow cytometry on a FACS-Calibur instrument (Becton Dickinson, San Jose, CA) in the presence of 1 μM fluorescent beads (Polysciences, Warrington, PA) spiked in at a concentration of 4×10^8 beads/ml. A negative control strain carrying the empty GFP vector with no promoter insert was treated identically to gauge background fluorescence levels (black bar in Figure 4C). Resultant data were analyzed using Flow-Jo software (Tree Star, Inc., Ashland, OR). The average absolute fluorescent cell counts normalized to bead counts from three biological replicate experiments ± s.d. are shown in the graph (Figure 4C).

Data integration analysis

To assess the extent of agreement between the three system-wide datasets presented in this study (gene expression, ChIP-chip, and motif search data), the hypergeometric distribution *P*-values were calculated, which reflect the likelihood that the intersection of any two of these three datasets are due to chance. Specifically, we calculated the significance of (i) the number of genes within 250 bp of both the ChIP-chip hits and binding motif sequences (one extra bp in motif degeneracy was allowed for a few of the genes in Figure 4A, which were nearby ChIP-chip hits, which showed a highly significant change in the microarray data;Figure 4A); (ii) the number of genes whose expression changed in the *trmB* mutant, which were also within 250 bp of a ChIP-chip hit; (iii) the number of genes changing in the transcriptome data, which contained a motif within 250 bp of their transcription start site (Figure 4B).

Detailed annotation analysis of the 37 genes in the intersection of the three high-throughput datasets (Figure 4B) was conducted using protein functional data from online databases (http://baliga.systemsbiology.net/halobacterium; KEGG, GO, STRING, HaloLex) (Bonneau *et al*, 2004; Bare *et al*, 2007; Pfeiffer *et al*, 2008; Jensen *et al*, 2009). To build the metabolic network governed by TrmB, a four-step bioinformatic process was conducted according to the flowchart shown in Supplementary Figure 3.

Accession numbers

All ChIP-chip and gene expression array data presented in this study are available at the National Center for Biotechnology Information Gene Expression Omnibus (NCBI GEO) under the accessions GSE13531, GSE13529, and GSE13498.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We are indebted to Ludmila Chistoserdova and Monica Orellana for their critical reading of the paper, Kenia Whitehead for useful discussions, Christopher Bare for software support, and Lee Pang and Noel Blake for assistance with the FACS analysis. This work was supported by grants from NIH (P50GM076547 and 1R01GM077398-01A2), DoE (MAGGIE: DE-FG02-07ER64327), NSF (DBI-0640950) to NSB, and from NIH (5F32GM078980-02) to AKS.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Ailion M, Bobik TA, Roth JR (1993) Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*. *J Bacteriol* **175:** 7200–7208

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25:** 25–29

Bailey TL, Gribskov M (1998) Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14:** 48–54

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* **34:** W369–W373

Baliga NS, Bjork SJ, Bonneau R, Pan M, Iloanusi C, Kottemann MC, Hood L, DiRuggiero J (2004) Systems level insights into the stress response to UV radiation in the halophilic archaeon *Halobacterium* NRC-1. *Genome Res* **14:** 1025–1035

Bare JC, Shannon PT, Schmid AK, Baliga NS (2007) The Firegoose: two-way integration of diverse data from different bioinformatics web resources with desktop applications. *BMC Bioinformatics* **8:** 456

Bell SD (2005) Archaeal transcriptional regulation—variation on a bacterial theme? *Trends Microbiol* **13:** 262–265

Bell SD, Cairns SS, Robson RL, Jackson SP (1999) Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. *Mol Cell* **4:** 971–982

Bonneau R, Baliga NS, Deutsch EW, Shannon P, Hood L (2004) Comprehensive *de novo* structure prediction in a systems-biology context for the archaea *Halobacterium* sp. NRC-1. *Genome Biol* **5:** R52

Bonneau R, Facciotti MT, Reiss DJ, Schmid AK, Pan M, Kaur A, Thorsson V, Shannon P, Johnson MH, Bare JC, Longabaugh W, Vuthoori M, Whitehead K, Madar A, Suzuki L, Mori T, Chang DE, Diruggiero J, Johnson CH, Hood L *et al.* (2007) A predictive model for transcriptional control of physiology in a free living cell. *Cell* **131:** 1354–1365

Brenneis M, Hering O, Lange C, Soppa J (2007) Experimental characterization of *cis*-acting elements important for translation and transcription in halophilic archaea. *PLoS Genet* **3:** e229

Brinkman AB, Bell SD, Lebbink RJ, de Vos WM, van der Oost J (2002) The *Sulfolobus solfataricus* Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. *J Biol Chem* **277:** 29537–29549

Danson MJ, Lamble HJ, Hough DW (2007) Central metabolism. In *Archaea: Molecular and Cellular Biology*, Cavicchioli R (ed), pp 260–287. Washington, DC: American Association for Microbiology Press

Facciotti MT, Reiss DJ, Pan M, Kaur A, Vuthoori M, Bonneau R, Shannon P, Srivastava A, Donohoe SM, Hood LE, Baliga NS (2007) General transcription factor specified global gene regulation in archaea. *Proc Natl Acad Sci USA* **104:** 4630–4635

Falb M, Muller K, Konigsmaier L, Oberwinkler T, Horn P, von Gronau S, Gonzalez O, Pfeiffer F, Bornberg-Bauer E, Oesterhelt D (2008) Metabolism of halophilic archaea. *Extremophiles* **12:** 177–196

Frith MC, Saunders NF, Kobe B, Bailey TL (2008) Discovering sequence motifs with arbitrary insertions and deletions. *PLoS Comput Biol* **4:** e1000071

Galperin MY, Koonin EV (1999) Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* **106:** 159–170

Geiduschek EP, Ouhammouch M (2005) Archaeal transcription and its regulators. *Mol Microbiol* **56:** 1397–1407

Gonzalez O, Gronau S, Falb M, Pfeiffer F, Mendoza E, Zimmer R, Oesterhelt D (2008) Reconstruction, modeling & analysis of *Halobacterium salinarum* R-1 metabolism. *Mol Biosyst* **4:** 148–159

Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ (2005) Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E coli* chromosome. *Proc Natl Acad Sci USA* **102:** 17693–17698

Herrgard MJ, Lee BS, Portnoy V, Palsson BO (2006) Integrated analysis of regulatory and metabolic networks reveals novel regulatory mechanisms in *Saccharomyces cerevisiae*. *Genome Res* **16:** 627–635

Ideker T, Thorsson V, Siegel AF, Hood LE (2000) Testing for differentially expressed genes by maximum-likelihood analysis of microarray data. *J Comput Biol* **7:** 805–817

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C (2009) STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res* **37:** D412–D416

Kanai T, Akerboom J, Takedomi S, van de Werken HJ, Blombach F, van der Oost J, Murakami T, Atomi H, Imanaka T (2007) A global transcriptional regulator in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes. *J Biol Chem* **282:** 33659–33670

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

Kaur A, Pan M, Meislin M, Facciotti MT, El-Gewely R, Baliga NS (2006) A systems view of haloarchaeal strategies to withstand stress from transition metals. *Genome Res* **16:** 841–854

Koide T, Reiss DJ, Bare CJ, Pang WL, Facciotti MT, Schmid AK, Pan M, Marzolf B, Van PT, Lo F-Y, Pratap A, Deutsch EW, Peterson A, Martin D, Baliga NS (2009) Prevalence of transcription promoters within archaeal operons and coding sequences. *Mol Sys Biol* **5:** doi:10.1038/msb.2009.42

Krug M, Lee SJ, Diederichs K, Boos W, Welte W (2006) Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB. *J Biol Chem* **281:** 10976–10982

Lee SJ, Surma M, Hausner W, Thomm M, Boos W (2008) The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus*. *Arch Microbiol* **190:** 247–256

Lie TJ, Wood GE, Leigh JA (2005) Regulation of nif expression in *Methanococcus maripaludis*: roles of the euryarchaeal repressor NrpR, 2-oxoglutarate, and two operators. *J Biol Chem* **280:** 5236–5241

Liu X, Brutlag DL, Liu JS (2001) BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes. *Pac Symp Biocomput* **6:** 127–138

Madan Babu M, Teichmann SA (2003) Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res* **31:** 1234–1244

Marzolf B, Deutsch EW, Moss P, Campbell D, Johnson MH, Galitski T (2006) SBEAMS-Microarray: database software supporting genomic expression analyses for systems biology. *BMC Bioinformatics* **7:** 286

Muller JA, DasSarma S (2005) Genomic analysis of anaerobic respiration in the archaeon *Halobacterium* sp. strain NRC-1: dimethyl sulfoxide and trimethylamine *N*-oxide as terminal electron acceptors. *J Bacteriol* **187:** 1659–1667

Ng WV, Kennedy SP, Mahairas GG, Berquist B, Pan M, Shukla HD, Lasky SR, Baliga NS, Thorsson V, Sbrogna J, Swartzell S, Weir D, Hall J, Dahl TA, Welti R, Goo YA, Leithauser B, Keller K, Cruz R, Danson MJ *et al.* (2000) Genome sequence of *Halobacterium* species NRC-1. *Proc Natl Acad Sci USA* **97:** 12176–12181

Pfeiffer F, Broicher A, Gillich T, Klee K, Mejia J, Rampp M, Oesterhelt D (2008) Genome information management and integrated data analysis with HaloLex. *Arch Microbiol* **190:** 281–299

Reece RJ, Beynon L, Holden S, Hughes AD, Rebora K, Sellick CA (2006) Nutrient-regulated gene expression in eukaryotes. *Biochem Soc Symp* **73:** 85–96

Reiss DJ, Facciotti MT, Baliga NS (2008) Model-based deconvolution of genome-wide DNA binding. *Bioinformatics* **24:** 396–403

Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, Simon I, Zeitlinger J, Schreiber J, Hannett N, Kanin E, Volkert TL, Wilson CJ, Bell SP,

Young RA (2000) Genome-wide location and function of DNA binding proteins. *Science* **290:** 2306–2309

Reuter CJ, Maupin-Furlow JA (2004) Analysis of proteasome-dependent proteolysis in *Haloferax volcanii* cells, using short-lived green fluorescent proteins. *Appl Environ Microbiol* **70:** 7530–7538

Rigali S, Schlicht M, Hoskisson P, Nothaft H, Merzbacher M, Joris B, Titgemeyer F (2004) Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new cis/trans relationships. *Nucleic Acids Res* **32:** 3418–3426

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS (2003) Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res* **31:** 6748–6757

Saxild HH, Brunstedt K, Nielsen KI, Jarmer H, Nygaard P (2001) Definition of the *Bacillus subtilis* PurR operator using genetic and bioinformatic tools and expansion of the PurR regulon with glyA, guaC, pbuG, xpt-pbuX, yqhZ-folD, and pbuO. *J Bacteriol* **183:** 6175–6183

Schmid AK, Reiss DJ, Kaur A, Pan M, King N, Van PT, Hohmann L, Martin DB, Baliga NS (2007) The anatomy of microbial cell state transitions in response to oxygen. *Genome Res* **17:** 1399–1413

Schut GJ, Brehm SD, Datta S, Adams MW (2003) Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides. *J Bacteriol* **185:** 3935–3947

Seshasayee AS, Fraser GM, Babu MM, Luscombe NM (2008) Principles of transcriptional regulation and evolution of the metabolic system in *E coli*. *Genome Res*

Shannon PT, Reiss DJ, Bonneau R, Baliga NS (2006) The Gaggle: an open-source software system for integrating bioinformatics software and data sources. *BMC Bioinformatics* **7:** 176

Shimada T, Ishihama A, Busby SJ, Grainger DC (2008) The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* **36:** 3950–3955

Siebers B, Schonheit P (2005) Unusual pathways and enzymes of central carbohydrate metabolism in Archaea. *Curr Opin Microbiol* **8:** 695–705

Sonenshein AL (2007) Control of key metabolic intersections in *Bacillus subtilis*. *Nat Rev Microbiol* **5:** 917–927

Tagkopoulos I, Liu Y-C, Tavazoie S (2008) Predictive behavior within microbial genetic networks. *Science* **320:** 1313–1317, 1154456

van de Werken HJ, Verhees CH, Akerboom J, de Vos WM, van der Oost J (2006) Identification of a glycolytic regulon in the archaea *Pyrococcus* and *Thermococcus*. *FEMS Microbiol Lett* **260:** 69–76

van der Oost J, Siebers B (2007) *The Glycolytic Pathways of Archaea: evolution by Tinkering*. Oxford, UK: Blackwell Publishing, Inc