Analysis of Gene Regulatory Networks in *Pyrococcus furiosus* by ChIP-seq

DISSERTATION ZUR ERLANGUNG EINES DOKTORGRADES DER NATURWISSENSCHAFTEN (DR. RER. NAT.) DER FAKULTÄT FÜR BIOLOGIE UND VORKLINISCHE MEDIZIN DER UNIVERSITÄT REGENSBURG

vorgelegt von

Robert Martin Reichelt

aus Weiden i.d.OPf.

Regensburg im Jahr 2014

Das Promotionsgesuch wurde eingereicht am: 26.05.2014

Die Arbeit wurde angeleitet von: Prof. Dr. Michael Thomm

Unterschrift:

Abstract

Transcription in Archaea relies on a eukaryotic-like transcription machinery composed of general transcription factors, the RNAP and corresponding promoter elements. In contrast, the regulation of archaeal gene expression depends on bacterial-like transcription factors. In the present work the gene regulatory networks of several archaeal transcription factors from the hyperthermophilic euryarchaeon *P. furiosus* were examined by ChIP. Moreover, combining ChIP with high-throughput sequencing (ChIP-seq) enabled the first time genome-wide TF binding site mapping in *P. furiosus*.

ChIP-seq revealed that the global regulator of sugar metabolism in *P. furiosus*, TrmBL1, not only represses and activates expression of sugar-uptake-, glycolysis- and gluconeogenesis-specific-enzymes encoding genes as proposed, but it also controls transcription of genes encoding proteins catalyzing proteolysis, AA metabolism and nucleotide degradation. In addition, TrmBL1 indirectly affects expression of genes involved in functionally-linked metabolic pathways by transcriptional repression of the hitherto not well-characterized secondary transcription factor PF1476. In contrast, the anti-TrmB antibodies were not specific enough to allow investigation of the proposed co-regulation of TrmB and TrmBL1 in sugar metabolism. Furthermore, the genes encoding TrmB and the TM-system were deleted in the *P. furiosus* type strain.

After successfull establishment of the ChIP-seq approach, genome-wide TF binding site mapping was extended to analyze a third paralogous TrmB protein designated as TrmBL2. These studies uncovered numerous potential functions of the TF as both architectural abundant DNA binding protein and global transcriptional regulator in *P. furiosus*. Furthermore, TrmBL2 ChIP-seq analysis using *P. furiosus* cells incubated at 107° for 30 min revealed a novel unexpected role of the protein in the heat shock response.

TFB-RF1 functions as transcriptional activator by recruitment of TFB to an imperfect BRE. Under optimal growth conditions expression of TFB-RF1 is not detectable and the transcription factor had to be over expressed from the *fbpase* promoter. First combination of ChIP-seq with the recently developed *in vivo* over-expression system of *P. furiosus* confirmed binding of the TF to its *in vitro* designated binding site *in vivo*. However, no additional binding sites were identified, which would elucidate a physiological function as regulator.

The presented studies of this work contribute to a better understanding of the mechanisms of transcriptional control mediated by gene regulatory networks to achieve optimal growth of the cells in response to differing environmental as well as intracellular conditions.

Zusammenfassung

Die Transkription in Archaeen beruht auf einem dem eukaryotischem System ähnlichen Transkriptionsapparat bestehend aus allgemeinen Transkriptionsfaktoren, der RNA Polymerase und den entsprechenden Promotorelementen. Im Gegensatz dazu wird die archaeelle Genexpression vor allem von Transkriptionsfaktoren reguliert, die dem bakteriellen System ähnlich sind. In dieser Studie wurden die genregulatorischen Netzwerke von mehreren archaeelen Transkriptionsfaktoren aus dem hyperthermophilen Modelorganismus *P. furiosus* durch Etablierung von ChIP Experimenten untersucht. Die Kombination von ChIP Experimenten mit Hochdurchsatz-Sequenzierung (ChIP-Seq) ermöglichte zum ersten Mal die Identifizierung von Bindestellen des jeweiligen Transkriptionsfaktors im gesamten Genom von *P. furiosus*.

Im ersten Teil der Arbeit konnte durch ChIP-Seq nachgewiesen werden, dass der globale Regulator des *P. furiosus* Zuckerstoffwechsels, TrmBL1, nicht nur wie bekannt die Expression von Genen reprimiert und aktiviert, die Zuckeraufnahme-, Glykolyse und Glukoneogenese-spezifische Enzyme kodieren, sondern auch von Genen, die Proteine kodieren, die an der Proteolyse, dem Aminosäurestoffechsel und der Degradierung von Nukleotiden beteiligt sind. Zudem beeinflusst TrmBL1 durch Repremierung der Genexpression des bis jetzt noch nicht intensiv charakterisierten Regulators PF1476 die Transkription von Genen, die an funktionell-verbundenenen metabolischen Prozessen beteiligt sind. Eine Analyse der vermuteten Ko-Regulation des Zuckerstoffwechsels durch die zwei paralogen *P. furiosus* TrmB Proteine TrmB und TrmBL1 konnte sowohl aufgrund der Kreuzreaktivität der TrmB spezifischen Antikörper mit TrmBL1 als auch aufgrund der genomischen Deletion des Genes, das für TrmB kodiert, nicht durchgeführt werden.

Im zweiten Teil der Arbeit wurden die ChIP-Seq Experimente auf eine Untersuchung der Funktion eines dritten paralogen TrmB Proteins, TrmBL2, ausgeweitet. Diese Versuche deckten zahlreiche potentielle Funktionen dieses Transkriptionsfaktors als sowohl strukturgebendes chromosomales DNA Bindeprotein als auch als globaler Transkriptions-regulator auf. Zudem wurde eine bis *dato* unbekannte und unerwartete Beteiligung von TrmBL2 an der Hitzeschockantwort von *P. furiosus* gefunden.

Zusätzlich ermöglichte die erstmalige Kombination von ChIP-Seq Experimenten mit dem kürzlich entwickelten *in vivo* Überexpressionssystem in *P. furiosus* eine Charakterisierung von TFB-RF1 *in vivo*. Dieser fungiert als Aktivator der Transkription durch Stimulierung der Rekrutierung von TFB zum Promotor-TBP-Komplex. Unter normalen Wachstums-bedingungen ist TFB-RF1 nur schwach bzw. gar nicht exprimiert, was eine Induktion der Expression erforderte. Durch die Kombination beider Ansätze konnte die Bindung von TFB-RF1 an die *in vitro* bestimmte Bindestelle im Genom *in vivo* bestätigt werden. Es konnten jedoch keine weiteren Bindestellen nachgewiesen werden, die zur Aufklärung der physiologischen Funktion dieses außergewöhnlich archaeellen Aktivators beitragen könnten.

Die in dieser Arbeit vorgestellten Ergebnisse tragen zu einem besseren Verständnis der durch genregulatorische Netzwerke vermittelten Transkriptionsregulation bei, die ein optimales Zellwachstum unter sich verändernden Umweltbedingungen ermöglicht.

Table of content

1. Introduction

In 1990 Woese et al. proposed the Archaea as third domain of life in addition to the Eukarya and Bacteria. Beside unique features as the ether linked membrane lipids with isopreonid chains, Archaea show a mosaic of typical bacterial-like and eukaryotic-like properties (Langworthy et al., 1974; Kandler and Hippe, 1977; Zillig et al., 1979; Huet et al., 1983). Archaea are unicellular organisms. With Bacteria they share the lack of organelles, cell size and cell shape. Moreover, the archaeal genome is a single circular chromosome, which can exist as single copy in the Crenarchaeota or as multiple copies in the Euryarchaeota (Soppa, 2011). In addition, functionally and physically linked genes are often co-transcribed and co-regulated as polycistronic operons (Bell and Jackson, 1998b; Yoon et al., 2011). In contrast, most archaeal mechanisms involved in information processing often represent simplified versions of their eukaryotic counterparts, e.g. the translation (Bell and Jackson, 1998a) and the DNA replication (Grabowski and Kelman, 2003). This is also true for transcription, which represents the first step in the flow of genetic information to biological systems. During the past 20 years, the hyperthermophilic euryarchaeon *P. furiosus* became one widely used organism to study the mechanisms of transcription especially *in vitro* (Hausner et al., 1996; Hethke et al., 1996; Goede et al., 2006; Grünberg et al., 2007; Naji et al., 2007; Kostrewa et al., 2009; Fouqueau et al., 2013).

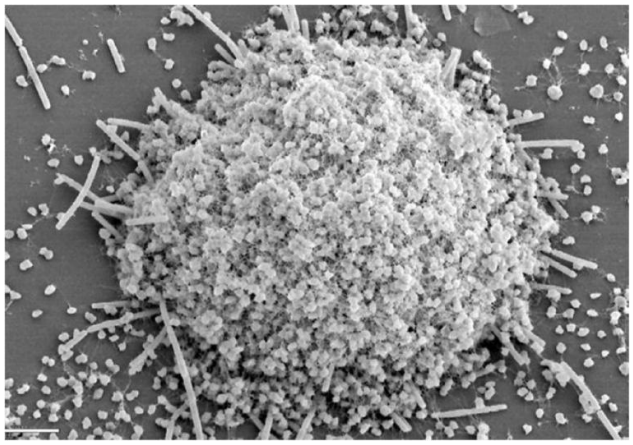


Figure 1. *P. furiosus* (cocci) in co-culture with *M. kandleri* (rods) adhering onto glassy carbon, size bar 5 µm (Schopf et al., 2008).

P. furiosus posses up to 50 appendages on the cell surface designated as flagella, which originate from one pole of the organism and are used for building cell-cell as well as cell-surface contacts and for swimming (Näther et al., 2006; Herzog and Wirth, 2012).

1.1. The hyperthermophilic model organism *P. furiosus*

Type strain and species *P. furiosus* DSM3638, 'the rushing fireball', was originally isolated from geothermally heated marine sediments at the beach of Porto di Levante, Vulcano, Italy (Fiala and Stetter, 1986). The organism belongs to the archaeal phylum Euryarchaeota and represents one widely used hyperthermophilic model organism beside *T. kodakarensis* within the order *Thermococcales* (Leigh et al., 2011). One of the most remarkable benefits of *P. furiosus* as hyperthermophilic model organism is its doubling time of 37 min exhibited at the optimal growth temperature of 100°C. Growth can be observed from 70°C to 103°C. *P. furiosus* grows on starch, maltose, peptone and complex organic substrates forming acetate, CO_2 and H_2 or H_2S in the presence of elemental sulfur (S^0) respectively (Fiala and Stetter, 1986; Thorgersen et al., 2012).

1.2. Glycolytic and gluconeogenic growth of *P. furiosus*

Metabolic pathways present in the Archaea are often unique like the methanogenesis or represent unusual and modified versions of the classical pathways known for Bacteria and Eukarya (Schönheit and Schäfer, 1995). For glycolytic growth on carbohydrates *P. furiosus* uses a modified Embden-Meyerhof-Parnas (EMP) pathway (Schäfer and Schönheit, 1992). In general, glycolysis is the conversion of glucose to two molecules of pyruvate producing ATP and reduction equivalents. Therefore, availability of a sufficient amount of glucose in the cells is an important prerequisite for glycolytic growth. In *P. furiosus* sugar uptake is accomplished by the interplay of various (poly-) saccharide-degrading enzymes and sugar-specific trans-membrane transport systems (Figure 2).

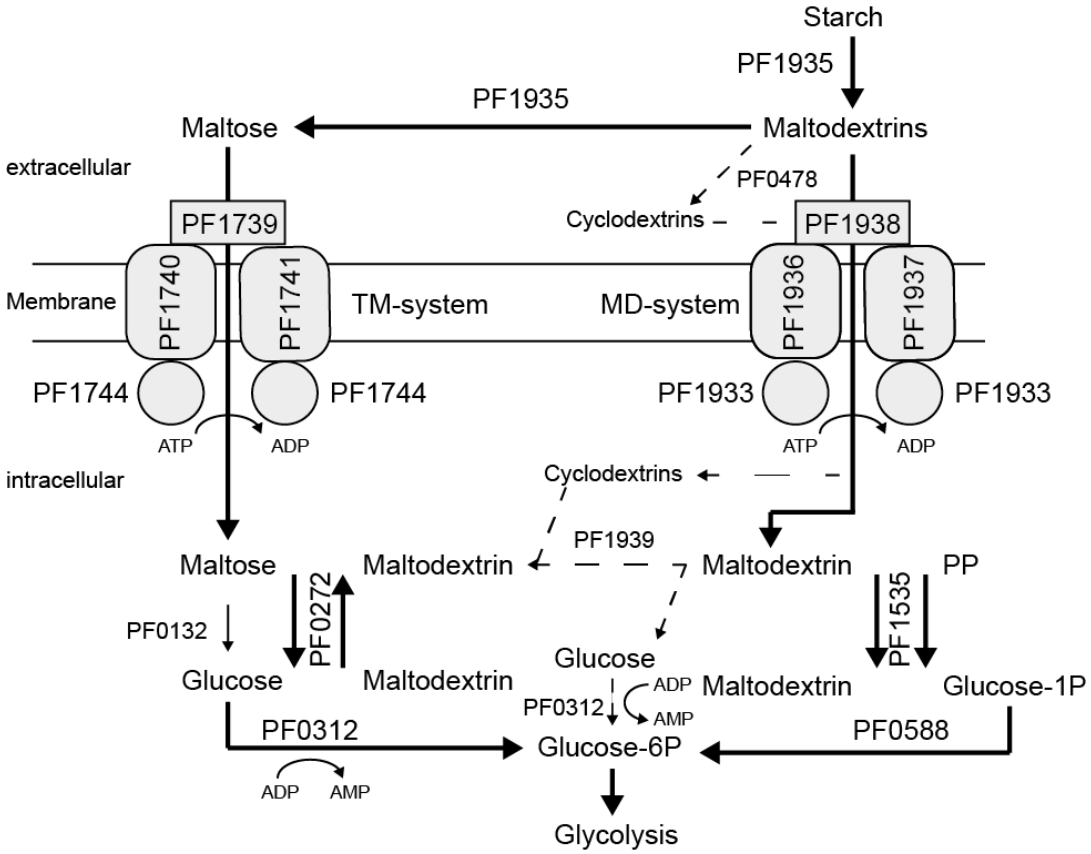


Figure 2. Proposed pathway for the starch and maltose uptake of *P. furiosus*.

The bold arrows indicate the major pathway. Dotted arrows indicate possible minor or alternative pathways. The following transporters and enzymes are invovlved in the process: PF0132, proposed α-glucosidase; PF0272, proposed 4-α-glucanotransferase; PF0312, ADP-dependent glucokinase; PF0478, proposed extracellular cyclomaltodextrin glucanotransferase; PF0588, phosphoglucose mutase; PF1535, maltodextrin phosphorylase; PF1739, 1740, 1741 and 1744, trehalose/maltose-specific ABC transporter (TM-system); PF1935, amylopullulanase; PF1933, 1936, 1937 and 1938 maltodextrin-specific ABC transporter (MD-system); PF1939, proposed maltodextrinase/ cyclodextrinase/pullulanase; ATP, adenosine triphosphate; ADP, adenosine diphosphate; AMP, adenosine monophosphate; PP, pyrophosphate. Modified from Lee et al., 2006a.

For utilization of starch, at first, polysaccharides are extracellularly hydrolyzed yielding maltodextrins and maltose (Brown et al., 1990; Lee et al., 2006a). Uptake of maltose and maltodextrins is mediated by two distinct ATP-binding cassette (ABC) transport systems, which are named regarding their substrate specificity as trehalose/maltose-specific ABC transport system (TM-system) and maltodextrin-specific ABC transport system (MD-system) (Koning et al., 2002). After transport in the cells maltose and maltodextrins are further degraded by intracellular enzymes to obtain glucose for glycolysis (Lee et al., 2006a; Vanfossen et al., 2008).

The modified EMP pathway of *P. furiosus* is characterized by the presence of several unique enzymes (Figure 3) (Siebers and Schönheit, 2005). The two sugar phosphorylation steps are carried out in *P. furiosus* by ADP-dependent enzymes, which are structurally unrelated to the ATP-dependent enzymes of the classical EMP pathways (Kengen et al., 1994; Kengen et al., 1995; Tuininga et al., 1999). Moreover, irreversible conversion of glyceralaldehyde-3-phosphate (GAP) to 3-phosphoglycerate is catalyzed by a glyceralaldehyde 3-phosphate ferredoxin oxidoreductase (GAPOR) in a ferredoxin-dependent manner (Mukund and Adams, 1995; VanDerOost et al., 1998). The pyruvate generated by glycolysis is further oxidatively decarboxylated to acetyl-CoA through a pyruvate ferredoxin oxidoreductase (POR) (Blamey and Adams, 1993). Acetyl-CoA can be used directly for energy conservation via acetyl-CoA synthetases (Schäfer et al., 1993). Moreover, the reduced ferredoxin can serve as an additional energy source using a membrane-bound hydrogenase complex and an ATP synthase (Sapra et al., 2003; Schut et al., 2013).

In addition to carbohydrates, *P. furiosus* is also able to use peptone, peptides (e.g. hydrolyzed casein) as well as pyruvate as primary carbon sources and in the presence of S^0 it even prefers peptides compared to maltose (Schäfer and Schönheit, 1991; Adams et al., 2001). Peptide fermentation in *P. furiosus* relies on various distinct ferredoxin-dependent 2-keto acid oxidoreductases like the keto-valine-ferredoxin oxidoreductase (VOR) that converts transaminated amino acids into their corresponding CoA-derivates (Heider et al., 1996). Moreover, during growth on non-carbohydrate carbon sources *P. furiosus* uses gluconeogenesis to synthesize glucose-6-phosphate from pyruvate (Figure 3). Like in Bacteria and Eukaryotes, gluconeogenesis proceeds by the reversible reactions of the modified EMP pathway (Schäfer and Schönheit, 1993). However, three enzymatic steps of glycolysis are irreversible and have to be catalyzed by gluconeogenic-specific enzymes (Verhees et al., 2003). The first step is the phosphorylation of pyruvate to phosphoenolpyruvate, which is suggested to be carried out by a phosphoenolpyruvate synthase (PEPS) (Imanaka et al., 2006). In addition, the unique and irreversible conversion mediated by the GAPOR has to be bypassed during gluconeogenic growth. This is achieved via the classical enzyme couple composed of GAP dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK). They function as solely gluconeogenic enzymes in *P. furiosus* and most other Archaea (Matsubara et al., 2011). The third gluconeogenic specific reaction is carried out by a bi-functional fructose 1,6-bisphosphate aldolase/phosphatase (FBPA/ase) (Verhees et al., 2002; Sato et al., 2004; Say and Fuchs, 2010). The bi-functionality of the enzyme relies on a single catalytic domain that catalyses two chemically distinct reactions: the reversible aldol condensation of dihydroxyacetone phosphate and glyceraldehydes 3-phosphate to fructose 1,6-bisphosphate as well as the dephosphorylation of fructose 1,6-bisphosphate to fructose 6-phosphate (Du et al., 2011; Fushinobu et al., 2011). This ensures that heat-labile triosephosphates are quickly converted to stabile fructose 6-phosphate.

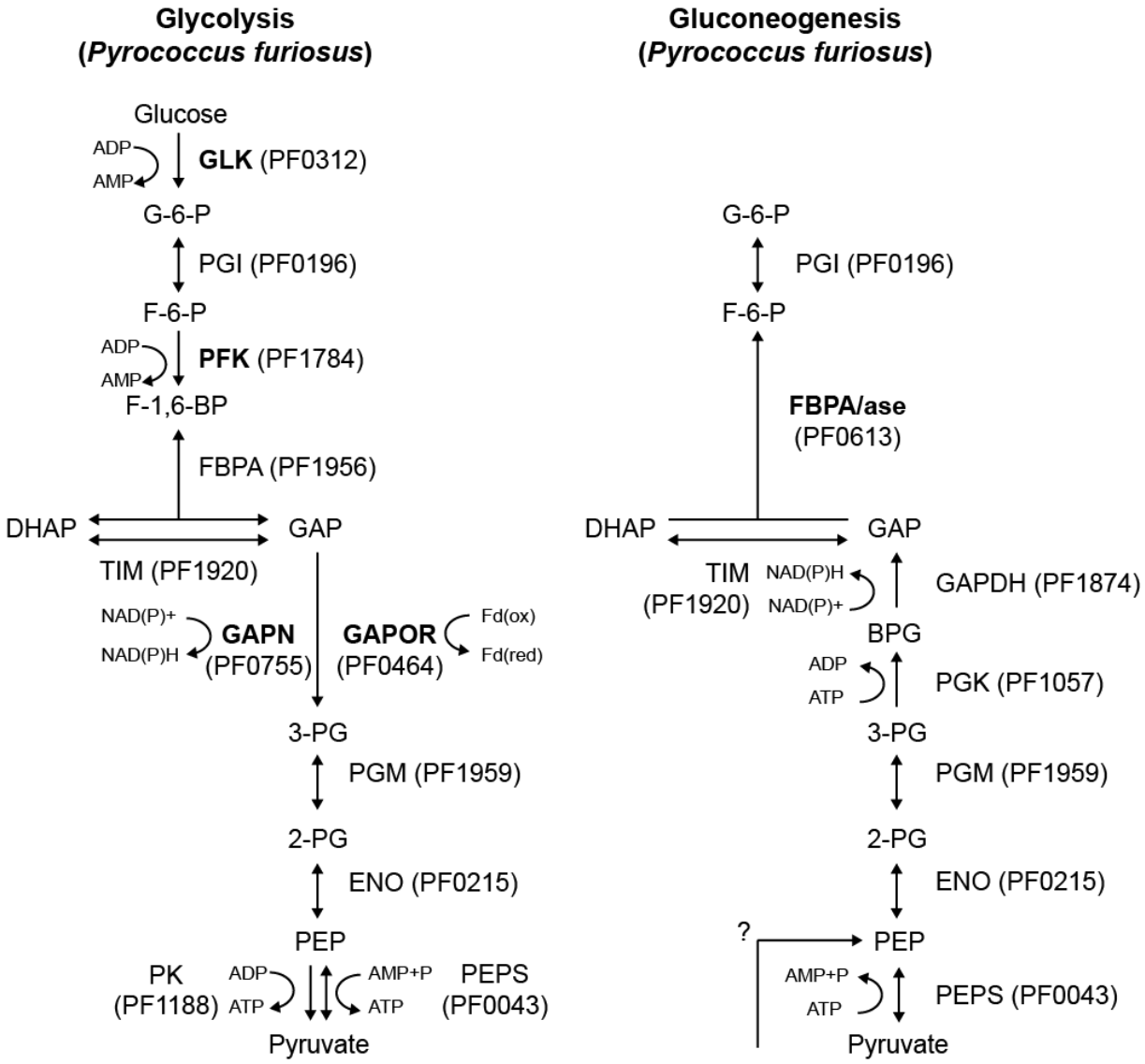


Figure 3. Schematic overview of glycolysis and gluconeogenesis in *P. furiosus*.

For glycolysis (left) *P. furiosus* uses a modified EMP pathway. Gluconeogenesis (right) of *P. furiosus* proceeds by the reversible steps of glycolysis and irreversible steps are carried out through gluconeogenesis-specific enzymes. Unusual archaeal enzymes are depicted in bold. Modified from Verhees et al., 2003; Siebers and Schönheit, 2005; Bräsen et al., 2014. Abbrevations: ENO, enolase; FBPA, F-1,6-BP aldolase; FBPA/ase bifunctional F-1,6-BP aldolase/phosphatase; GAPDH, GAP dehydrogenase; GAPN, nonphosphorylating GAPDH; GAPOR, GAP:Fd oxidoreductase; GLK, ADP-dependent glucose kinase; HK, hexokinase; PEPS, PEP synthetase; PFK, ADP-dependent phosphofructokinase; PGI, phosphoglucose isomerase; PGK, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PK, pyruvate kinase; TIM, trioseisomerase; G-6-P, glucose 6-phosphate; F-6-P, fructose 6-phosphate; F-1,6-BP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; BPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate; Fd, ferredoxin; Fd(ox), oxidized ferredoxin; Fd(red), reduced ferredoxin; ATP, adenosine triphosphate; ADP, adenosine diphosphate; AMP, adenosine monophosphate; P, phosphate; NAD(P)H, Nicotinamide adenine dinucleotide (phosphate).

Regulation of the classical glycolysis is a very complex process, which usually takes place on both the DNA level (gene expression) and the enzyme level (enzymatic activity) (Verhees et al., 2003). In all Archaea, the classical points of allosteric control like inhibition of the GLK by glucose 6-phosphate are absent (Cardenas et al., 1998; Dorr et al., 2003). In contrast, the activities of two other enzymes, the GAPOR and the GAPN, were shown to be influenced by the presence of specific effectors (Mukund and Adams, 1995; Matsubara et al., 2011). No additional allosteric inhibition has been reported so far for the *Thermococcales*. Furthermore, a whole genome microarray analysis of *P. furiosus* grown on maltose (glycolytic growth) or peptides (gluconeogenic growth) revealed differential expression of genes encoding proteins and enzymes involved in glycolysis or gluconeogenesis (Schut et al., 2003). Genes encoding proteins and enzymes that are mainly involved in sugar transport and glycolysis are downregulated under gluconeogenic growth conditions, whereas genes encoding gluconeogenic enzymes are upregulated during gluconeogenic growth. Thus, it is assumed that regulation of glycolysis and gluconeogenesis in *P. furiosus* is mainly regulated on the DNA level by transcriptional control of gene expression.

1.3. Transcription and gene regulation in *P. furiosus*

In all three domains of life transcription is carried out by DNA-dependent multi-subunit RNAPs, which are derived from a common ancestral enzyme (Grohmann et al., 2009; Lane and Darst, 2010). While Bacteria and Archaea contain only one single type of multi-subunit RNAP, Eukaryotes use three (animals) to five (plants) distinct multi-subunit RNAPs (RNAP I, II, III, IV and V). They differ in subunit composition as well as in their specific functions in transcription (Darst, 2001; Wierzbicki et al., 2008; Werner and Grohmann, 2011; Vannini and Cramer, 2012). Archaeal RNAPs are able to synthesize all kinds of RNAs and their subunit composition varies from 11 to 14 in the diverse described phyla of Archaea (Langer et al., 1995; Grohmann and Werner, 2011). Moreover, archaeal RNAPs show a high similarity to their eucaryotic counterpart, the RNAP II, regarding subunit composition as well as overall structural architecture (Goede et al., 2006; Koonin et al., 2007; Hirata et al., 2008b; Korkhin et al., 2009; Wojtas et al., 2012).

Beside the RNAPs, closeness of the archaeal and eukaryotic (RNAPII system) transcription machineries arises from the eukaryotic-like promoter elements and general transcription factors (GTFs) participating in transcription initiation in Archaea. The archaeal core promoter consists of the TATA-box, an A/T-rich region located approximately 26 to 27 bp (center) upstream of the transcription start site (TSS), the TFB response element (BRE), which is located directly upstream of the TATA-box, and the initiator element (INR), which is composed of a pyrimidine-purine-dinucleotide and is located in the range of the TSS (Soppa, 1999b, 1999a, Hausner and Thomm, 2001; Zhang et al., 2009). For promoter-dependent transcription initiation *in vitro,* only two GTFs are essential in Archaea (Qureshi et al., 1995b; Qureshi et al., 1995a; Hausner et al., 1996; Hethke et al., 1996). *P. furiosus* contains one TATA binding protein (TBP), which corresponds to the TBP subunit of the TFIID complex in Eukarya. Furthermore, *P. furiosus* has two types of TFB, designated as TFB1 (PF1377) and TFB2 (PF0687). TFB1 is highly related to eukaryotic TFIIB and it is assumed to function as standard TFB (Hausner et al., 1996; Kostrewa et al., 2009). In contrast, TFB2 lacks a region located in the TFB B-finger/B-linker domain and it functions poorly in promoter-specific transcription *in vitro* (Micorescu et al., 2008). The presence of additional copies of TBP and TFB is a common feature of Archaea. It is assumed that various subsets of different TBP-TFB pairs play a role for specific physiological functions and stress responses in Archaea (Baliga et al., 2000; Facciotti et al., 2007; Lu et al., 2008; Brooks et al., 2011). A whole

genome microarray analysis revealed increasing TFB2 transcript levels upon heat shock. This suggests that the additional TFB version of *P. furiosus* is involved in the response to heat stress (Shockley et al., 2003). In addition to TBP and TFB, a third GTF, TFE, can be found in Archaea, which is homologous to the α-subunit of TFIIE of the RNAPII system (Hanzelka et al., 2001; Meinhart et al., 2003). Whereas TFE is not essential for promoter-dependent transcription initiation *in vitro*, presence of TFE appears to be essential *in vivo* for *T. kodakarensis* (Bell et al., 2001a; Grünberg et al., 2007; Naji et al., 2007; Hirata et al., 2008a).

While the basal archaeal transcription machinery represents a simplified version of the eukaryotic RNAPII system, transcriptional regulation of gene expression in Archaea is mediated by mainly bacterial-like transcription factors (TFs) (Bell and Jackson, 1998a; Bell et al., 2001b). Approximately 53% of all identified TFs in archaeal genomes have at least one homologue in Bacteria and only 2% have a eukaryotic homologue (Perez-Rueda and Janga, 2010). The bacteria-like TFs of Archaea have been extensively studied during the past 15 years (Peeters et al., 2013). The most abundant DNA-binding domain (DBD) of archaeal/bacterial TFs exhibits the helix-turn-helix (HTH) motif (~45%) and a common ancestry is postulated for the TFs of Prokaryotes (Aravind and Koonin, 1999; Aravind et al., 2005; Perez-Rueda and Janga, 2010). The DBD recognizes a specific DNA binding motif in the promoter region for site-directed binding of a TF. Typical archaeal binding motifs have a size between 11 and 17 bp and are often semi-palindromic with a number of less- or non-informative base pairs in the center (Aravind et al., 2005; Peeters et al., 2013).

Most studied archaeal regulators represent repressors and two major repression mechanisms were described: Inhibition of TBP/TFB binding or blocking of RNAP recruitment (Bell and Jackson, 2000; Dahlke and Thomm, 2002; Vierke et al., 2003; Lee et al., 2005). It is assumed that both repression mechanisms correspond to a different mode of derepression: Through inhibition of TBP/TFB binding the entire initiation complex needs to be formed before transcription is initiated, whereas the second mechanism allows the pre-bound TBP and TFB to rapidly recruit RNAP in the absence of the blocking TF, leading to a faster derepression. Transcriptional activation is accomplished by stimulation of TBP binding or TFB binding to imperfect promoter elements (Ouhammouch and Geiduschek, 2001; Ouhammouch et al., 2003; Ouhammouch et al., 2005; Ochs et al., 2012). It is assumed that stimulatory effects of both activation mechanisms are mediated via protein-protein interactions. Beside TFs, which have a dedicated function as repressor or activator, additional archaeal regulators were described exhibiting dual functionality. For the first time, dual functionality as repressor and activator in Archaea was proven for the global regulators of sugar metabolism *Thermococcus* glycolysis regulator (Tgr) from *T. kodakarensis* and TrmB-like protein 1 (TrmBL1) from *P. furiosus* (Kanai et al., 2007; Lee et al., 2008). The dual functionality of these regulators relies on binding downstream (blocking RNAP recruitment) or upstream (transcriptional stimulation) of the BRE and TATA-box (Kanai et al., 2007; Lee et al., 2008). Figure 4 summarizes proved molecular regulation mechanisms and binding locations of archaeal TFs. The main function of TFs *in vivo* is to link diverse stimuli like nutrient availability, presence of specific ion metals or heat shock to appropriate transcriptional responses. This allows the cells to obtain optimal growth under changing environmental and intracellular conditions. All these regulatory interactions between transcription factors and their target genes are often also referred as gene regulatory network (GRN) (Babu et al., 2004). One well-studied example of a GRN in *P. furiosus* represents the regulation of its sugar metabolism.

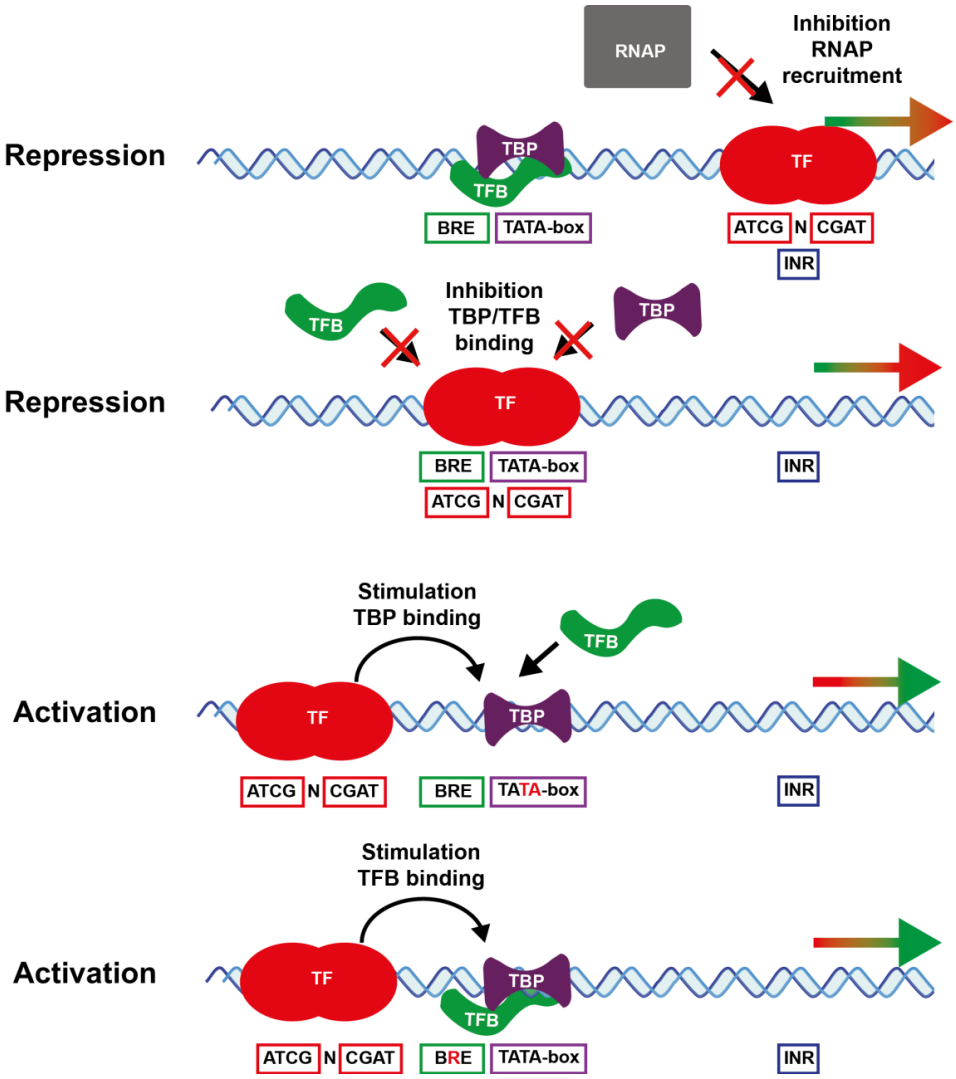


Figure 4. Proved mechanisms of transcriptional regulation mediated by archaeal TFs.

Repression can be achieved via inhibition of RNAP recruitment or inhibition of TBP/TFB binding. Transcriptional activation relies on stimulation of TBP or TFB binding to imperfect promoter elements (BRE or TATA-box), depicted by red letters. ATCGNCGAT represents a `typical´ semi-palindromic DNA binding motif of archaeal TFs. According to Dahlke and Thomm, 2002; Ouhammouch et al., 2003; Lee et al., 2005; Peeters et al., 2013.

1.4. The role of TrmB and TrmB-like regulators in sugar metabolism of *P. furiosus*

The first described TF participating in regulation of sugar metabolism of Archaea was identified in *T. litoralis*. This regulator was named transcriptional regulator of mal operon (TrmB) and it functions in a trehalose/maltose-dependent manner as repressor of an operon encoding a trehalose/maltose-specific ABC transporter (TM-operon/system) (Lee et al., 2003). The *trmb* gene itself is co-transcribed within this gene cluster and the transcriptional control by TrmB represents a mechanism of autoregulation. In the genome of *P. furiosus* a 16 kb DNA region highly similar to the TM gene cluster was identified. It is assumed that this is a result of a recent lateral gene transfer between both organisms (DiRuggiero et al., 2000). Moreover, a bacterial origin for this transport system was suggested (Horlacher et al., 1998).

The corresponding TrmB protein of *P. furiosus* represses not only transcription of the TM-system, it also regulates transcription of an operon encoding a separate maltodextrin specific ABC transporter (MD-operon/system) (Lee et al., 2005). The N-terminal DBD of TrmB resembles a winged helix-turn-helix (wHTH) structure followed by an additional helix, which promotes dimerisation of the protein (Krug et al., 2013). The wHTH DBD of TrmB presumably binds to a pseudo-palindromic DNA motif with the consensus 5'-**TACT**NNN**AGTA-**3', which was identified in the TM-promoter. This motif overlaps the promoter elements of the TM-promoter and it was shown by footprint analysis that TrmB protects the BRE and TATA-box of the TM-system promoter (Lee et al., 2003; Lee et al., 2005; Krug et al., 2013). An additional TrmB binding site could be detected upstream of the promoter elements, but the relevance of this site for transcriptional control remained unclear (Surma, 2011). Moreover, the C-terminal domain of TrmB functions as sugar binding responsive element, which is able to bind various sugars like maltose, sucrose, maltotriose and trehalose (Lee et al., 2005; Krug et al., 2006). Transcriptional inhibition of the TM operon was only released by maltose and trehalose, whereas repression of the MD operon was only released by maltotriose, maltodextrins and sucrose (Lee et al., 2005). This discrepancy of TrmB mediated transcriptional response induced by distinct sugars indicated a differential regulation mechanism functioning for both promoters. This assumption was also encouraged by the finding that the TrmB binding site in the promoter region of the MD-system overlaps the TSS and not the BRE and TATA-box. In addition, the MD-system promoter lacks the pseudo-palindromic TrmB DNA binding motif found in the TM-promoter (Lee et al., 2005).

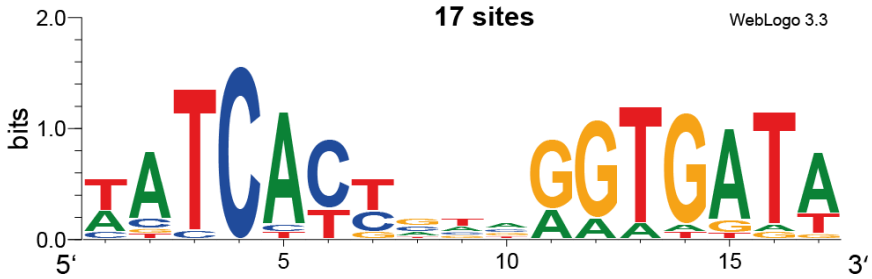


Figure 5. *Thermococcales* Glycolytic Motif (TGM).

The TGM is present in 17 promoter regions of *P. furiosus* genes encoding proteins and enzymes involved in sugar uptake and glycoylsis as well as gluconegenesic-specific enzymes (vandeWerken et al., 2006). The logo was created using WebLogo 3.3 (http://weblogo.threeplusone.com).

In contrast, a different palindromic DNA motif was identified in the MD-promoter and 16 additional promoter regions of genes encoding proteins catalyzing sugar uptake and glycolysis of *P. furiosus* (VanDeWerken et al., 2006). This sequence motif was termed as *Thermococcales* glycolytic motif (TGM) and it has the consensus 5'-**TATCAC**NNNNN**GTGATA**-3' (Figure 5). The TGM is also present in *T. kodakarensis* (29 genes), but not in *P. hirokoshii* and *P. abyssi*, which are not able to ferment sugars (VanDeWerken et al., 2006). The TrmB binding site in the MD promoter partially overlapped the TGM; however, as true transcriptional regulator, which recognizes the TGM, a TrmB paralogue in *P. furiosus* was identified designated as TrmBL1 (Lee et al., 2007b). Deletion of the corresponding TrmBL1 orthologue Tgr in *T. kodakarensis* led to significant decrease in growth rate of the deletion strain under gluconeogenic conditions compared with the wild-type strain, whereas comparable growth rates were observed under glycolytic conditions (Kanai et al., 2007). Furthermore, whole genome microarray analysis revealed that Tgr acts

in a maltotriose-dependent manner both as repressor of genes encoding proteins and enzymes involved in sugar uptake and glycolysis and as activator of gluconeogenic enzymes-encoding genes. The promoter regions of all Tgr-regulated genes contained the TGM (Kanai et al., 2007). It was shown *in vitro* that *P. furiosus* TrmBL1 can bind TGM containing promoters of genes encoding proteins involved in sugar uptake, e.g. MD-system, and glycolysis, e.g. PFK (Lee et al., 2007b). In the corresponding promoters the TrmBL1 binding site overlaps the TSS (binding downstream of the BRE and TATA-box) and TrmBL1 functions as transcriptional repressor (VanDeWerken et al., 2006; Lee et al., 2007b; Lee et al., 2008). Moreover, TrmBL1 is able to bind its own promoter region, which does not contain the TGM in contrast to the *tgr* gene of *T. kodakarensis*. Autorepression of TrmBL1 was proven by cell-free transcription *in vitro* (VanDeWerken et al., 2006; Lee et al., 2007b; Lee et al., 2008). Like TrmB, TrmBL1 is able to bind sugars. In the presence of maltotriose (1 mM) or maltose (5 mM) TrmBL1 mainly exists as hepta/octamer, whereas in the absence of sugars it forms a tetramer. In addition, TrmBL1 mediated transcriptional repression of the *pfk* promoter is inhibited in the presence of maltose, maltotriose and fructose (Lee et al., 2007b). The TGM is not only present in the promoter regions of genes involved in sugar uptake and glycolysis, it can also be found in the promoter regions of genes encoding unique gluconeogenic enzymes. There, it is located upstream of the BRE and TATA-box indicating transcriptional activation (VanDeWerken et al., 2006). This activation was shown *in vitro* for the *fbpase* promoter of *P. furiosus* (Lee et al., 2008). Moreover, the majority of genes, which contain the TGM in their promoter regions, were differentially expressed during growth of *P. furiosus* under glycolytic or gluconeogenic conditions (Schut et al., 2003). Taken all these results together, it is suggested that TrmBL1 from *P. furiosus* as well as Tgr from *T. kodakarensis* function as global regulators with dual functionality for sugar uptake, glycolysis and gluconeogenesis through binding to TGM containing promoters in a sugar-dependent manner. Moreover, in *P. furiosus* it was shown that some genes can be regulated by both TrmB and TrmBL1. TrmB binds with higher affinity to the TM-promoter than to the MD-promoter, whereas TrmBL1 recognizes with a higher affinity the MD-promoter than the TM-promoter. Transcriptional repression of TM- and MD-promoter by both regulators in a sugar-dependent manner was demonstrated *in vitro*. Thus, it is assumed that at some promoters cross-regulation mediated by both regulators plays an important role *in vivo* to obtain optimal growth of the cells depending on the varying intracellular amounts of glucose, maltose, maltotriose and larger maltodextrins (Lee et al., 2008; Bräsen et al., 2014). The proposed GRN of TrmB and TrmBL1 determined by *in vitro* and *in silico* analysis is summarized in Table 1.

In the genome of *P. furiosus* two additional TrmB-like proteins termed as TrmBL2 and TrmBL3 were identified (Lee et al., 2007b). The function of TrmBL3 for transcriptional control in *P. furiosus* is still unclear. Analysis of the evolutionary conservation of these four *P. furiosus* TrmB/TrmBL proteins within the *Thermococcales* revealed that only TrmBL2 is present in the genomes of *P. furiosus*, *T. kodakarensis*, *P. hirokoshii* and *P. abyssi* and evolutionary conserved (Lee et al., 2007b). It was shown that TrmBL2 is able to bind TGM-containing promoters like the MD-promoter, but not exclusively. However, transcriptional repression or activation by TrmBL2 of these bound promoters was not observed. Furthermore, a sequence alignment of TrmB, TrmBL1 and TrmBL2 revealed that the protein lacks the C-terminal part of the sugar binding domain of TrmB and TrmBL1 (Lee et al., 2007b; Lee et al., 2008). Thus, the physiological function of TrmBL2, e.g. in sugar metabolism, due to its belonging to the TrmBL proteins remained speculative.

Table 1. Proposed gene Regulatory Network of TrmB and TrmBL1 from *P. furiosus*.

Gene	Product	TGM	gly/glu	TrmB		TrmBL1	
P. furiosus	Protein/ Enzyme	rel. to BRE/TATA	expression (log_2ratio)	EMSA binding	cell-free transcription	EMSA binding	cell-free transcription
PF0043	PEPS	down	0,18	n.d.	n.d.	n.d.	n.d.
PF0124	TrmBL1	-	0.40	-	n.d	+	**repression**
PF0132	GLS	down	0,71	n.d.	n.d.	n.d.	n.d.
PF0196	PGI	down	**2,26**	n.d.	n.d.	n.d.	n.d.
PF0215	ENO	down	**1,32**	n.d.	n.d.	n.d.	n.d.
PF0272	AMY1	down	**4,7**	n.d.	n.d.	n.d.	n.d.
PF0312	GLK	down	**1,09**	+	n.d.	++	n.d.
PF0464	GAPOR	down	**2,54**	+	n.d.	++	n.d.
PF0477	AMY2	down	**-2,45**	n.d.	n.d.	n.d.	n.d.
PF0478	AMY3	up	0,89	n.d.	n.d.	n.d.	n.d.
PF0588	PGM	down	**0,69**	n.d.	n.d.	n.d.	n.d.
PF0613	FBPA/ase	up	**-3,95**	n.d.	n.d.	n.d.	**activation**
PF1057	PGK	up	**-2,87**	n.d.	n.d.	n.d.	n.d.
PF1109	SB-protein	down	**3,36**	n.d.	n.d.	n.d.	n.d.
PF1739	TM-system	-	**2.58**	++	**repression**	+	**repression**
PF1784	PFK	down	**2,54**	+	**repression**	++	**repression**
PF1920	TIM	down	**2,24**	n.d.	n.d.	n.d.	n.d.
PF1938	MD-system	up	**2,13**	+	**repression**	++	**repression**
PF1956	FBPA	down	0,33	n.d.	n.d.	n.d.	n.d.
PF1959	PGM	down	**1,75**	n.d.	n.d.	n.d.	n.d.

Enzymes and transporters: PEPS, PEP synthetase; GLS, proposed α-glucosidase; PGI, phospho-glucose isomerase; ENO, enolase; AMY1, proposed 4-α-glucanotransferase; GLK, glucose kinase; GAPOR, GAP:Fd oxidoreductase; AMY2, proposed extracellular α-amylase; AMY3, proposed extracellular cyclomaltodextrin glucano-transferase; PGM, phosphoglycerate mutase; FBPA/ase bifunctional F-1,6-BP aldolase/phosphatase; PGK, phosphoglyceratkinase; SB-protein, starch binding protein (PF1109-1110); TM-system, trehalose/maltose-specific ABC transporter (PF1739 to PF1744); PFK, phosphofructokinase; GAPDH, GAP dehydrogenase; TIM, trioseisomerase; MD-system, maltodextrin-specific ABC transporter (PF1938 to PF1933); FBPA, F-1,6-BP aldolase. TGM rel. to BRE/TATA, position TGM relative to the BRE and TATA-box; down, downstream; up, upstream (VanDeWerken et al., 2006). gly/glu, differential expression (log_2ratio) after growth of *P. furiosus* under glycolytic (gly) and gluconeogenic (glu) condition. Significant values are depicted in bold (Schut et al., 2003). EMSA, electromobility shift assay; binding ++, high affinity; binding +, low affinity, binding -, no affinity; n.d. not determined. EMSA and cell-free transcription results were obtained from Lee et al., 2003; Lee et al., 2005, Lee et al., 2007c, 2007b, Lee et al., 2008. Transcriptional activation (blue) and repression (red) is depicted by colors.

Maruyama et al. (2011) reported that the orthologous TrmBL2 protein, TK0471, from *T. kodakarensis* is a novel type of abundant chromosomal binding protein, which can form thick fibrous structures with DNA. In the isolated chromatin fraction of *T. kodakarensis* TrmBL2 represents one of the highest enriched proteins beside the archaeal Histones A and B, Alba and RNAP subunits. Moreover, sucrose gradient centrifugation of the partially digested chromatin fraction allowed separation of three fractions: A, a high density-fraction containing mainly TrmBL2; B, a middle density fraction containing mainly RNAP subunits but additionally TrmBL2 as well as histone proteins; C, a low-density fraction containing only histone proteins. High-throughput sequencing of the DNA isolated from a TrmBL2 enriched fraction indicates that TrmBL2 binds to non-coding (9.4%) as well as coding (90.6%) regions. This distribution does not differ from the predicted distribution of non-coding (~8%) and coding (~92%) regions in the *T. kodakarensis* genome. In total, 449 TrmBL2 enriched genomic regions could be identified. Therefore, abundant and random TrmBL2 binding to the whole genome was suggested. Nevertheless, whole genome microarray analysis of a TrmBL2 deletion strain compared with the wild-type strain showed that TrmBL2 mainly acts as transcriptional repressor of genes, where it was located in the non-coding promoter region. In conclusion, TrmBL2/TK0471 is assumed to function both as an architectural protein for the chromosome as well as global transcriptional repressor (Maruyama et al., 2011).

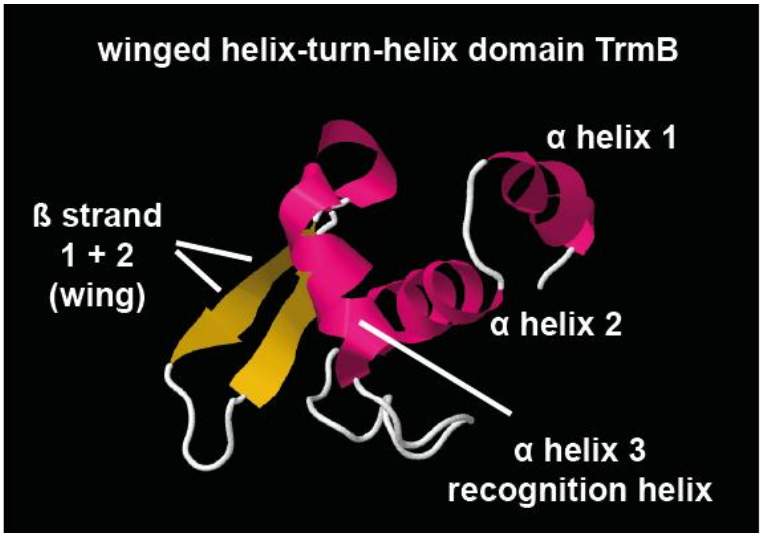


Figure 6. Winged helix-turn-turn (wHTH) domain of TrmB.

The crystal structure of TrmB (Pdb-ID 3QPH) revealed a wHTH structure of the N-terminal DNA binding domain (residues 11 to 78). The HTH structure resembles a trihelical motif composed of α1, α2 and α3. The α3 helix functions as DNA recognition helix. The wing is built by the two β-strands, β1 and β2. According to Krug et al., 2013.

The common feature of all above mentioned TrmB and TrmBL proteins from the *Thermococcales* is the TrmB_domain (Pfam ID: PF01978; Finn et al., 2006), which is a member of the HTH DNA-binding domains (Figure 6). Proteins of the TrmB_domain family are widely distributed within the Bacteria, gram-positive as well as gram-negative, and universally distributed within the Archaea, including the phyla Euryarchaeota, Crenarchaeota, Thaumarchaeota, Korarchaeota and Nanoarchaeota. Therefore, it is suggested that the TrmB domain family together with only three additional TF families might belong to the ancestral core of TFs in Archaea (Perez-Rueda and Janga, 2010). Moreover, TrmB, TrmBL1

and TrmBL2 contain an additional domain designated as Regulator_TrmB domain (Pfam ID: PF11495). About 60% of all archaeal and bacterial TrmB TF family proteins exhibit this domain architecture composed of both TrmB and Regulator_TrmB domain (http://pfam.sanger.ac.uk). In contrast to the universal distribution of TrmB_domain proteins within the Archaea (100 euryarchaeal and 47 crenarchaeal species), proteins with the combined architecture are mainly found in euryarchaeal species (60 species) compared to crenarchaeal species (5 species) (http://pfam.sanger.ac.uk; Maruyama et al., 2011).

So far, only few members of the TrmB TF family with or without combined architecture from not-*Thermococcales* species within the Archaea were characterized. A TrmB protein (combined architecture) from the euryarchaeon *H. NRC1* was shown to regulate evolutionarily diverse but functionally linked metabolic pathways in response to changes in carbon source availability (Schmid et al., 2009). Especially during nutritional limitation, this TrmB protein binds to 113 promoter regions, where it activates or represses genes encoding enzymes involved in glycolysis, tricarboxylic acid cycle, and amino-acid biosynthesis pathways of *H. NRC1*. These results together with a dynamic perturbation approach to elucidate the topology of the TrmB metabolic GRN lead to the conclusion that the TrmB protein *from H. NRC1* is an important component for mediating metabolic modularity, integrating nutrient status and regulating gene expression dynamics alone and in concert with secondary regulators (Todor et al., 2013). Furthermore, in the genome of the euryarchaeon *M. acetivorans* a protein of the TrmB TF family was identified, designated as MreA, which has single domain architecture and lacks the sugar-sensing domain (Reichlen et al., 2012). MreA functions in the global regulation of methanogenic pathways and acts both as repressor of methylotrophic-specific genes and activator of acetate-specific genes. Recently, the first TrmB protein, MalR, from a crenarchaeal species (*S. acidocaldarius*) was characterized, which contains both the TrmB as well as the Regulator_TrmB domain (combined architecture) (Wagner et al., 2014). In contrast to the two above mentioned TrmB/TrmBL proteins with dual functionality, MalR solely acts as transcriptional activator. This was demonstrated for the maltose regulon including an ABC transporter, an α-amylase and an α-glycosidase (Wagner et al., 2014).

Studying the function of TrmB and TrmBL proteins from diverse archaeal species revealed the many faces of these transcriptional regulators. They can function as solely repressors (e.g. TrmB from *P. furiosus*) or solely activators (e.g. MalR from *S. acidocaldarius*) as well as regulators with dual functionality (e.g. MreA from *M. acetivorans*). Moreover, transcriptional control can be restricted to a few genes/operons (e.g. MalR) or can affect up to more than one hundred promoter regions of functionally-linked genes (e.g. TrmB from *H. NRC1*). In addition, certain TrmB/TrmBL proteins contain a sensor domain to respond to e.g. nutrient availability (e.g. TrmBL1), whereas other TrmB/TrmBL proteins presumably lack this sensor domain (e.g. TrmBL2). Therefore, studying the function of these types of TFs in transcriptional regulation *in vitro* and *in vivo* in more detail will not only contribute to a better understanding, how the bacterial-like TFs from Archaea interact with their eukaryotic-like basal transcription machinery. It will also help to understand the diverse mechanisms of transcriptional regulation mediated by GRNs to obtain optimal growth of the cells in response to varying environmental as well as intracellular conditions.

1.5. The function of TFB-RF1 as transcriptional activator of *P. furiosus*

Comparative genomic approaches for predictions of regulons and their cis regulatory elements are widely used in Bacteria (Novichkov et al., 2010). In contrast, only few studies are reported for Archaea (Bell and Jackson, 2000; VanDeWerken et al., 2006; Novichkov et al., 2010). Validation of predicted TF binding sites regarding recognition through corresponding putative regulators in *P. furious* lead to the discovery of a novel type of transcriptional activator termed TFB-recruiting factor 1 (TFB-RF 1). This TF (PF1088) is likely co-transcribed with the hypothetical gene PF1089. Gel shift assays confirmed binding of TFB-RF1 upstream of the BRE and TATA-box in the promoter region of PF1089 and *in vitro* transcription assays showed activation of transcription. Moreover, analysis of the activation mechanism in more detail gave evidence that this TF works as the first described TFB recruitment factor in Archaea (Ochs et al., 2012).

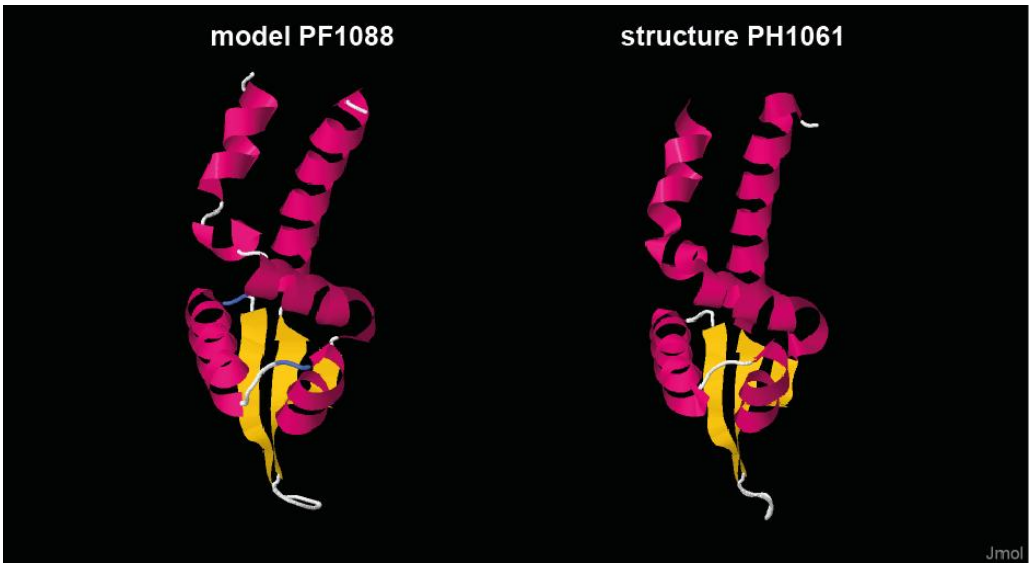


Figure 7. Calculated structure model of PF1088 (left) and crystal structure of PH1061 (right).

Both illustrate the high structural similarity of the proteins. The PF1088 (TFB-RF1) model was obtained from ModWeb (https://modbase.compbio.ucsf.edu/modweb/) (Ochs, 2009). The crystal structure of PH1061 (Pdb-ID 1UB9) was reported by (Okada et al., 2006).

For a PF1088 homologue from *P. horikoshii*, PH1061, the crystal structure was published (Okada et al., 2006). Crystallography of PH1061 (Pdb-ID: 1ub9) revealed a wHTH domain structure of the protein, but the function of the protein remained unclear. Furthermore, comparative modeling showed a high structural similarity of TFB-RF1 and PH1061 (Figure 7) (Ochs, 2009). PH1061 exists as dimer in both, the crystal and in solution (Okada et al., 2006). In addition, it was shown by gelfiltration experiments that the recombinantly expressed TFB-RF1 exists as a dimer in solution (Winter, 2012). So far, the physiological function of TFB-RF1 in *P. furiosus* is still unclear. The co-expressed hypothetical protein PF1089 belongs to the uncharacterized conserved proteins in Archaea and gives no evidence for the physiological function of TFB-RF1. Protein domain data base analysis showed that TFB-RF1 belongs to the family of ArsR transcriptional regulators. Members of this family encode for e.g. zinc-dependent transcriptional repressors, but no co-repressors or co-activators of TFB RF1 were found yet (Okada et al., 2006; Ochs et al., 2012).

1.6. ChIP-seq for genome wide TF binding site mapping

So far, characterization of bacterial-like TFs in *P. furiosus* is mainly based on *in vitro* analysis. On the one hand, this primarily originates from the availability of numerous well-established *in vitro* approaches like cell-free transcription, EMSA (electromobility-shift-assay) or footprinting (Hausner et al., 1996; Hethke et al., 1996; Dahlke and Thomm, 2002; Lee et al., 2007b; Ochs et al., 2012). On the other hand, the *P. furiosus* transcription model system lacked for a long time the disposability of a genetic system for analysis of TF mediated gene regulation *in vivo*. Whereas the first available genetic system of *T. kodakarensis* was described in 2003, the genetic system of *P. furiosus* was published recently in 2010 and optimizations are still in progress (Sato et al., 2003; Waege et al., 2010; Lipscomb et al., 2011; Atomi et al., 2012; Kreuzer et al., 2013; Waege, 2014). So far gene disruption of a *P. furiosus* TF gene combined with analysis of growth behavior and differential expression was only used once for studying the regulation of iron metabolism in the natural competent *P. furiosus* strain COM1 (Zhu et al., 2013).

Beside gene disruption analysis, which relies on the availability of a genetic system, Chromatin Immunoprecipitation (ChIP) represents a powerful tool for studying GRNs *in vivo*. ChIP enables detection of the association of individual proteins with specific genomic regions *in vivo*. In the first step, living cells are treated with formaldehyde, which efficiently generates protein-DNA, protein-RNA and protein-protein crosslinks in the cells. This prevents any large-scale redistribution of these components. In the next step, cells are lysed by sonication, which additionally shears the chromatin to small-sized DNA fragments. Then chromatin fragments containing the protein of interest are selectively immunoprecipitated with an antibody raised against this protein. DNA sequences that directly or indirectly crosslink with the protein of interest are specifically enriched in the immunoprecipitated sample. Immunoprecipitations can also be done with antibodies against a modified (e.g. phosphorylated, actetylated, methylated) peptide of the protein or against an epitope, if the protein of interest is epitope-tagged. Reversal of the formaldehyde-induced crosslinks by heating permits recovery, isolation and quantitative analysis of the immunoprecipitated DNA (Solomon and Varshavsky, 1985; Solomon et al., 1988; Orlando and Paro, 1993; Hecht et al., 1996; Orlando et al., 1997; Aparicio et al., 2005).

So far only a limited number of publications exists using ChIP for TF binding site mapping in Archaea *in vivo*. Within the Euryarchaeota the halophilic aerobic strain *H. NRC1* became a widely-used model organism for studying GRNs *in vivo*. On the one hand this was aided by the presence of a well-developed genetic system from the early stage (Cline et al., 1989; Lam and Doolittle, 1989; Mankin et al., 1992; Leigh et al., 2011). On the other hand especially its optimal growth under mesophilic conditions (42°C), which is close to the optimal growth temperature of several bacterial and eukaryotic model organisms, enabled simple adaptation of biochemical approaches like ChIP for this system (Facciotti et al., 2007; Schmid et al., 2009; Schmid et al., 2011; Sharma et al., 2012). In the Crenarchaeota analysis of GRNs of *S. solfataricus* TFs *in vivo* by ChIP raised the past years due to establishment of a Nanobody®-based ChIP protocol (Duc et al., 2012; Nguyen-Duc et al., 2013; Song et al., 2013). Most ChIP studies in archaeal systems used ChIP combined with whole genome microarray analysis (ChIP-chip) for the analysis of genome wide protein occupancies. ChIP-chip allows a genome-wide detection of DNA–protein interactions via hybridization of the ChIP-enriched fragments to a microarray (Blat and Kleckner, 1999; Ren et al., 2000; Lee et al., 2006b). During the past years ChIP combined with high-throughput sequencing (ChIP-seq) became a extensively used approach for quantitative mapping of protein-DNA binding

events in a genome wide manner in eukaryotic and bacterial systems (Johnson et al., 2007; Kahramanoglou et al., 2011). In ChIP–seq, the immunoprecipitated DNA fragments are sequenced directly (~30-50 bp read length). ChIP–seq shows higher resolution (20 to 50 bp), fewer artifacts, greater coverage and a larger dynamic range than ChIP–chip (Barski et al., 2007; Johnson et al., 2007; Park, 2009). Especially, the precise mapping of protein-binding sites by ChIP–seq enables a more accurate identification of binding targets for transcription factors and enhancers as well as of the corresponding DNA binding motifs (Visel et al., 2009). A critical step in the ChIP-seq approach represents the identification of ChIP-enriched regions. Several so-called `peak-callers´ for genome wide detection of TF binding sites using ChIP-seq are available. The resulting output of these algorithms commonly ranks enriched-regions by absolute signal intensity (read number) or by computed significance of enrichment (e.g., *P*-values and false discovery rates) (Zhang et al., 2008; Rozowsky et al., 2009; Landt et al., 2012; Bailey et al., 2013). Selection of a specific peak-caller is also based on the distinct modes of interaction of a protein of interest with the genome. The following protein classes can be distinguished: A, *point-source* factors are localized at specific positions that generate highly localized ChIP-seq signals (e.g. bacterial-like TFs); B, *broad-source* factors are associated with large genomic domains (e.g. histone proteins); C, *mixed-source* factors can bind in point-source fashion to some locations of the genome, but form broader domains of binding in others (e.g. RNAPs) (Landt et al., 2012). Whereas the majority of peak-callers is mainly suitable for analysis of ChIP experiments studying point-source or broad-source factors, binding site identification of mixed-source factors is still challenging (Rashid et al., 2011; Landt et al., 2012).

Recently, a workflow for genome-wide mapping of archaeal transcription factors by ChIP-seq was reported for *H. NRC1*, which includes the bioinformatic package Pique for identification of binding events (Wilbanks et al., 2012). Pique is adapted to the requirements of a peak calling algorithm suitable for binding site identification in the compact archaeal genomes. Moreover, Pique is able to operate on systems that have genomic complexities such as IS elements, gene dosage polymorphisms and accessory genomes that cause coverage variations unrelated to ChIP or in cases where the organism is not identical to the reference genome (Wilbanks et al., 2012). An additional application of ChIP-seq within the Archaea was described by Wojtas et al. (2012). They studied the general transcription machinery of *S. acidocaldarius* and could prove that the novel identified RNAP subunit Rpo13 is associated with RNAP during transcription elongation *in vivo.* This proved that Rpo13 represents a bona-fide RNAP subunit. Similar applications were widely used in bacterial and eukaryotic transcription systems. Transferring them to archaeal systems will contribute to a better understanding of general transcription mechanisms as well as the function of associated proteins, factors and complexes from initiation to elongation to termination in Archaea *in vivo* (Jasiak et al., 2008; Mooney et al., 2009; Werner, 2013). Figure 8 shows a schematic overview of the whole ChIP-seq workflow.

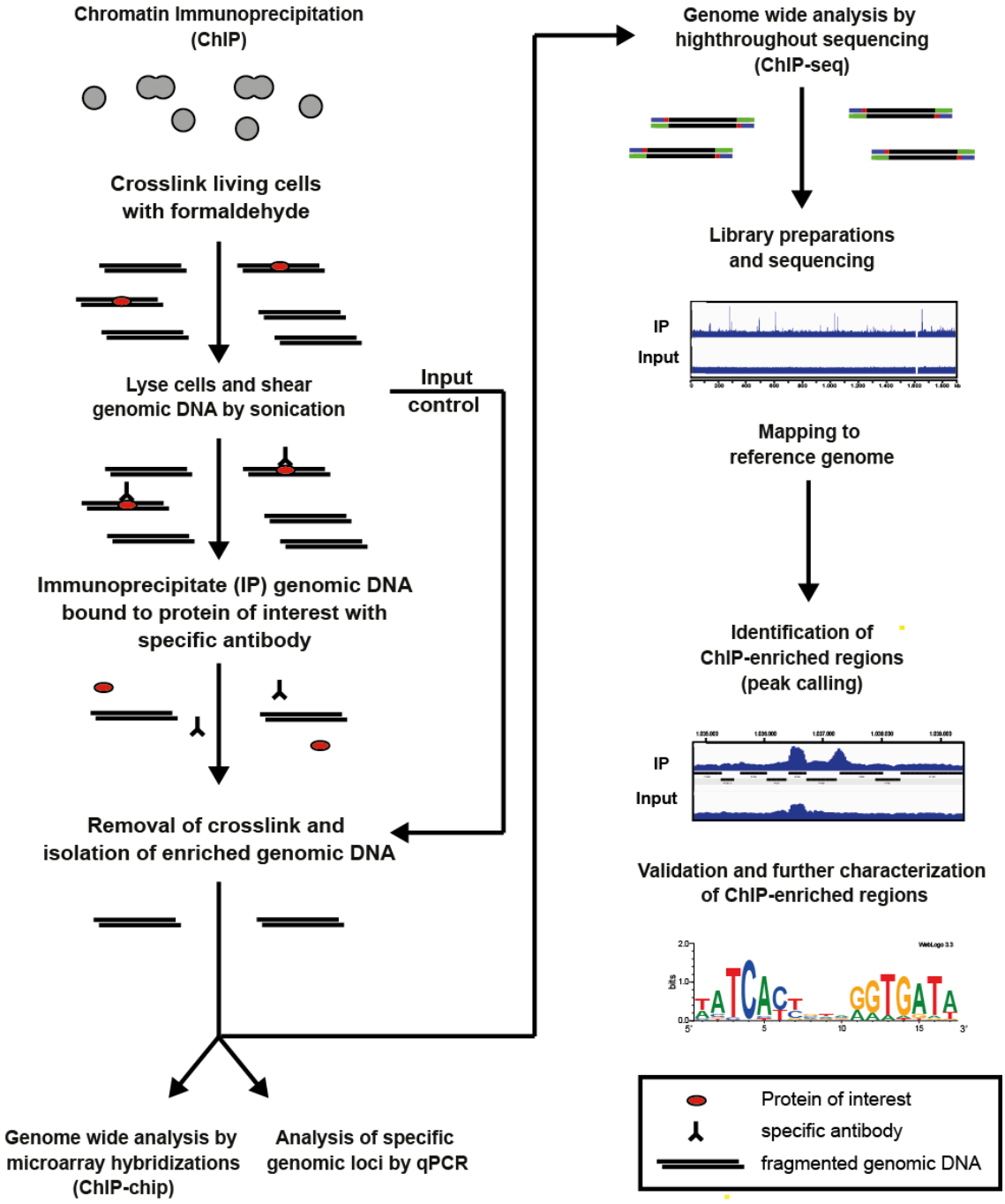


Figure 8. Schematic overview of the ChIP-seq workflow.

Modified and extended from Aparicio et al., 2005. After applying the ChIP procedure enrichment of genomic DNA can be analyzed via quantitative real-time PCR (qPCR) for specific genomic loci (ChIP-qPCR) or by microarray hybridizations for genome-wide binding sites mapping (ChIP-chip). For genome-wide indentification of TF binding sites by high-throughput sequencing (ChIP-seq), at first, immunoprecipitated DNA fragments have to processed and ligated with sequence platform-specific adaptors followed by size-selection and PCR amplification (library preparation). One of the most used sequencing platforms for ChIP-seq is the Illumina HiSeq platform (Illumina). Sequencing generates several millions to several ten millions short reads and more per ChIP sample. These DNA fragments have to be mapped to the reference genome by short reads alignment programs (Langmead et al., 2009). Identification of ChIP-enriched regions is accomplished via peak-calling algorithms (Zhang et al., 2008; Wilbanks et al., 2012). Identified ChIP-enriched regions have to be subsequently validated *in vivo* as well as *in vitro*. Additionally, DNA binding motifs present in the ChIP-enriched region can be identified *in silico* (Bailey et al., 2009).

1.7. Aim of the thesis

In *P. furiosus* ChIP was successfully used to show that the global repressor *Pyrococcus* heat shock regulator (Phr) dissociates from the DNA upon heat shock (growth temperature higher than 105°C), whereas TBP remains bound to the promoter and transcription is initiated (Vierke et al., 2003; Liu et al., 2007; Vierke, 2007). However, no additional applications for this technique in *P. furiosus* are reported so far. The aim of this thesis was to adapt the existing ChIP procedure of *P. furiosus* for genome-wide TF mapping *in vivo* by ChIP-seq. Initial work was done during my diploma thesis 2009 entitled `Untersuchung zur Funktion der TrmB-Familie in *Pyrococcus furiosus* anhand von Chromatin Immunpräzipitations experimenten´ (Reichelt, 2009). During this work the ChIP procedure was adapted to analyze binding of the global regulator TrmBL1 to specific genomic regions like the *pfk* promoter after glycolytic and gluconeogenic growth of *P. furiosus*. ChIP-enrichment was shown by qPCR analysis. Additional optimizations and first tests of library preparations for high-throughput sequencing were done by Stefan Eisenschink during his diploma thesis entitled `Analyse von DNA-Protein-Wechselwirkungen verschiedener Transkriptionskomponenten mittels Chromatin Immunpräzipitationsexperimenten in *Pyrococcus furiosus*´ (Eisenschink, 2010).

In the first part of this thesis ChIP-seq was used to dissect the specific functions of TrmB and TrmBL1 in the regulation of sugar metabolism of *P. furiosus*. After successful adaptation of the ChIP-procedure, the analysis was extended to study the genome-wide occupancy of the abundant chromosomal binding protein TrmBL2, which is assumed to act as global transcriptional repressor. Moreover, combination of shuttle-vector based protein over expression in *P. furiosus* and ChIP-seq should be used to unravel the physiological function of the unique transcriptional activator TFB-RF1, which is able to recruit TFB to an imperfect BRE. These studies led to the development of complete ChIP-seq workflow for the hyperthermophilic euryarchaeon *P. furiosus*, which can be easily adapted to any protein of interest. Moreover, studying TF binding sites by ChIP-seq in a genome-wide manner for various DNA binding proteins will contribute to a better understanding of the function of GRNs in regulation of archaeal transcription.

2. Material and Methods

2.1. Material

2.1.1. Strains

Pyrococcus furiosus DSM3638	DSMZ
Pyrococcus furiosus MurPf5	Kreuzer, 2009
One Shot® TOP10 Chemically Competent *E. coli*	Life Technologies (Invitrogen™), Darmstadt
One Shot® BL21(DE3) Chemically Competent *E. coli*	Life Technologies (Invitrogen™), Darmstadt

2.1.2. Kits, sets and mixes

Kits, sets and mixes	**Supplier**
Qiaquick PCR purification Kit	Qiagen, Hilden
Wizard SV Gel and PCR Purification Kit	Promega, Mannheim
Wizard Plus SV Miniprep DNA Purification	Promega, Mannheim
ReliaPrep™ gDNA Tissue Miniprep System	Promega, Mannheim
NEBNext ®-ChIP-Seq library prep reagent set for illumina protocol	New England Biolabs, Ipswich (MA, USA)
NEBNext® Multiplex oligos (Index Primers Set 1)	New England Biolabs, Ipswich (MA, USA)
NEBNext® Multiplex oligos ((Index Primers Set 2)	New England Biolabs, Ipswich (MA, USA)
Champion™ pET151 Directional TOPO® Expression Kit with BL21 Star™ (DE3) One Shot® Chemically Competent *E. coli*	Life Technologies (Invitrogen™), Darmstadt
SensiMix™ SYBR® No-ROX Kit	Bioline, Luckenwalde
Phusion High fidelity DNA polymerase	Biozym Scientific, Hessisch Oldendorf
NEBNext®High Fidelity Master Mix	New England Biolabs, Ipswich (MA, USA)
dNTP mix (10mM)	Peqlab, Erlangen

2.1.3. (Bio-) chemicals and reagents

All not extra-denoted chemicals and biochemicals were obtained from the common distributors of (bio-) chemicals and reagents in Germany: LifeTechnologies (Darmstadt), Merck (Darmstadt), Roche (Mannheim), Roth (Karlsruhe), Sigma Aldrich (Steinheim) and VWR (Darmstadt).

(Bio-)chemicals and reagents	**Supplier**
Agencourt AMpure XP	Beckmann Coulter, Krefeld
Bacto peptone	BD, Heidelberg
Bacto yeast extract	BD, Heidelberg
BamHI	New England Biolabs, Ipswich (MA, USA)
BSA	Roche Diagnostics, Mannheim
cOmplete ULTRA Tablets	Roche Diagnostics, Mannheim
DNase I	Sigma Aldrich, Steinheim
Dynabeads®Protein G for Immunoprecipitation	Life Technologies, Darmstadt
GeneRuler 100bp DNA ladder	Fisher Scientific (Fermentas), Schwerte
GeneRuler 1kb DNA ladder	Fisher Scientific (Fermentas), Schwerte
Leupeptin	Sigma Aldrich, Steinheim
Maltose ≥95%	Roth, Karlsruhe
PageRuler Prestained Protein Ladder	Fisher Scientific (Fermentas), Schwerte
PageRuler Unstained Protein Ladder	Fisher Scientific (Fermentas), Schwerte
Pepstatin A	Sigma Aldrich, Steinheim
PMSF	Roth, Karlsruhe
poly-dIC	Sigma Aldrich, Steinheim
Proteinase K	Fisher Scientific, Schwerte
RNase A	Fisher Scientific, Schwerte
SmaI	New England Biolabs, Ipswich (MA, USA)
Sodium-pyruvate	Applichem, Darmstadt
Starch	Merck, Darmstadt

2.1.4. DNA Oligos

All DNA oligos were obtained from Eurogentec (Köln) or Eurofins Genomics (Ebersberg).

* marks Cy5 labeled primers. Primers are written from 5' to 3' and primers used in multiple experiments are only noted once.

ChIP-qPCR analysis TrmBL1

PF0272FW	TGTCCATGCTAACACCCTGT
PF0272RW	TATCTCCCACTTCCGTCCAC
PF0287FW	AGGATTTGAACCTCGGACCT
PF0287RW	GACCAGAGAGCGCAAACTTTA
PF0478FW	CTTGCGACCTGGGAAAAATA
PF0478RW	GGTGATTCTTAATCTTTCTCTTCTTT
PF0648FW	CATCCCCCTTTTCTGACTTG
PF0648RW	CGGAAGCCATTTTGATCACT
PF1025FW	AAACGCTCGCCTGGTCTAAT
PF1025RW	AAACTCTTTCTAACCTTGTGAAGTCT
PF1085.1nFW	TTTTGCTCTCATACCTCTCCA
PF1085.1nRW	CCCACGCCTGACTCTTTTCT
PF1476FW	GCATACATTGGCCTGTCCTT
PF1476RW	CGAAGGTGATAAAAATGGAGGA
PF1761FW	GGGGGTCTATGAGGAGAACG
PF1761RW	GCCCCAGGCTCTTATGTATG
pf1784_F_new	CTAGGCATCTAAATTGAAAAAGTTT
pf1784_R_new	ATTATAGGCAGTATATATTCCGAGC
PF1874FW	TTCCTAATCAGACAATGCCAGT
PF1874RW	GCGTAGGCTACCCTCTTTCC
PF2047FW	GCAAGATCATTGGGAGAAGC
PF2047RW	TCATAAATCTCCTGGCCACA
atp_new-F (PF1882)	TTCAAAATCCTTGGATCATAACC
atp_new-R (PF1882)	TGCCCCTACCAACATCTCTC
gdh_new-F (PF1602)	TTGAAAATGTTTGAGGAACACC
gdh_new-R (PF1602)	TTGGGCAGCTCTTTCAAGTT

Southern blot analysis

PF1743probe_F	CCCCAGAAATTTCACATGCT
PF1743probe_R	CGTCAAAACTCCTGCTCCTC
PF1753probe_F	CCGAAGAGTTTGAGGCAAAG
PF1753probe_R	CGTTGGCATTCTTTTCCAGT

Copy number analysis

PF1743coding1F	ATGGAGATTCCCCCAGAAAT
PF1743coding1R	TACTGGGCCCATAAACCAAG
PF1753coding1F	ATTTGCTCCAAATCCAGACG
PF1753coding1R	GATCCGTGTCATCGAACCTT

Amplification of PF1784 see ChIP-qPCR analysis TrmBL1

ChIP-qPCR analysis TrmBL2

PF0496200F	GGGACAGTTGTTTATTTACTTTACC
PF0496upR	CCTTACCACCACCACAAACTTGTC

Amplification of PF1602 and PF1882 see ChIP-qPCR analysis TrmBL1

ChIP-qPCR analysis TFB-RF1 (PF1088)

PF1089FW	CCTCCATAAATCCCACCAAA
PF1089RW	CCCATTAATGAGGATGTTGTAAA

Amplification of PF1602 and PF1784 see ChIP-qPCR analysis TrmBL1

EMSA TFB-RF1 (PF1088)

PF1089up_ Cy5_F *	CAAACCTAAGCTCTGAACTACAGAGTTTG ACCTGAGTATAAATACCTCTACCTCCATT TAATTTTTGGTG
PF1089up_Bio_R	CACCAAAAATTAAATGGAGGTAGAGGTAT TTATACTCAGGTCAAACTCTGTAGTTCAG AGCTTAGGTTTG
PF1529Cy5FW*	GTGAAGAGCCATTACTGCTACAGC
PF1529RW	GGTAGAAAAGCTGCTATGGAAATAC

PF1476 TOPO®Cloning

PF1476topofw	CACCATGGAGGAGCCTATACTCAGAC
PF1476toporw	TTACGACTTCAGCATGTTTAAAAC

EMSA PF1476

PF0531FW *	TACAAATACGACAATATAATCTCACAAA
PF0531RW	CTT CTT TTG CCC TCT CTA TAA GTG
PF0972FW *	TTTTTGAATTTAATTACGGGCAT
PF0972RW	CATTGGCACGTATGCTCC
PF1784FW *	TCTTGCGACAAAATCTAACG
PF1784RW	AATTGAAAAAGTTTTTTAAATAATCTCAT
PF1911FW *	TATACCAGGTATTTCTCATTGC
PF1911RW	TCTTCTCCTTCTACTATCTCAT

2.1.5. Software tools and websites

	website/source	reference
Aida Image Analyzer 4.27	Raytest, Straubenhardt	
Bioedit	http://www.mbio.ncsu.edu/bioedit/bioedit.html	Hall, 1999
Bowtie	https://github.com/BenLangmead/bowtie	Langmead et al., 2009
Bowtie2	https://github.com/BenLangmead/bowtie2	Langmead and Salzberg, 2012
Clustal omega	https://www.ebi.ac.uk/Tools/msa/clustalo/	Sievers et al., 2011
Fasttree	http://www.microbesonline.org/fasttree/	Price et al., 2010
Galaxy server	https://usegalaxy.org/	Blankenberg et al., 2010
GATK	http://www.broadinstitute.org/gatk/	VanDerAuwera, Geraldine A. et al., 2002
Genetrack	https://github.com/ialbert/chipexo/tree/master/genetrack	Albert et al., 2008
Gibbs Sampler favorov	http://favorov.bioinfolab.net/SeSiMCMC/	Favorov et al., 2005
HHpred	http://toolkit.tuebingen.mpg.de/hhpred	Soding et al., 2005
IBM SPSS Statistics 21	University of Regensburg software center	
IGV	https://www.broadinstitute.org/igv/home	Robinson et al., 2011
Interpro	http://www.ebi.ac.uk/interpro/	Hunter et al., 2012
Jalview	http://www.jalview.org/	Waterhouse et al., 2009
Meme suite	http://meme.nbcr.net/meme/	Bailey et al., 2009
MACS	https://github.com/taoliu/MACS/	Zhang et al., 2008
NCBI	http://www.ncbi.nlm.nih.gov/	
Oligocalc	http://www.basic.northwestern.edu/biotools/oligocalc.html	Kibbe, 2007
Pfam	http://pfam.xfam.org/	Finn et al., 2006
Pique	https://github.com/ryneches/pique	Wilbanks et al., 2012
Primer 3	http://primer3.ut.ee/	Untergasser et al., 2012
Rotorgene software	rotor-gene_1_7_87	
Samtools	http://samtools.sourceforge.net/	Li et al., 2009
Serialcloner	http://serial-cloner.softonic.de/	
SigmaPlot 10.0	University of Regensburg software center	
Splitstree	http://www.splitstree.org/	Huson and Bryant, 2006
T- coffee	http://www.tcoffee.org/Projects/homepage.html	DiTommaso et al., 2011
UCSC archaeal genome browser	http://archaea.ucsc.edu/	Chan et al., 2012
Tree view	http://taxonomy.zoology.gla.ac.uk/rod/treeview.html	Page, 1996

2.2. Methods

All not extra-denoted methods including standard buffer or media formulations were applied according to Current Protocols in Molecular Biology (Ausubel et al., regularly updated) or according to the manufactures instructions of the applied Kits. All DNA concentrations were measured using the NanoDrop-1000 system (Peqlab, Erlangen).

2.2.1. Cultivation of *P. furiosus*

P. furiosus type strain DSM3638 was obtained from the DSMZ in 2009 and after growth in rich-medium the strain was prepared for long time storage. Cells were grown under anaerobic conditions in nutrient rich medium based on SME and supplemented with different organic substrates (Fiala and Stetter, 1986). Detailed descriptions for cultivations of *P. furiosus* cells can be found in the PhD thesis of Waege (2014) and additional numerous PhD, diploma, master and bachelor theses of the department of microbiology and Archaeenzentrum Regensburg. Rich-medium contained 0.1 % starch, 0.1 % peptone and 0.1 % yeast extract. Minimal-starch medium contained 0.1 % starch and 0.025 % yeast extract. For minimal-pyruvate medium starch was replaced by 40 mM pyruvate and for minimal-maltose starch was replaced by 5 % maltose. After inoculation with *P. furiosus* cells (1:100 dilution) cultivation was done at 75°C to 95°C overnight or until the appropriate cell density was reached.

After recultivation in rich-medium, *P. furiosus* DSM3638 cells were adapted to glycolytic or gluconeogenic growth conditions using minimal medium supplemented with either starch or sodium pyruvate. While the cells grown under gluconeogenic conditions reached a cell density of $1*10^8$ cells/ml after 8 to 10 h, the cells grown under glycolytic conditions showed an extended lag phase and reached a cell density of $1*10^8$ cells/ml after 16 to 18 h. Moreover, *P. furiosus* cells grown under gluconeogenic conditions were inoculated additional three times in medium supplemented with pyruvate. These three cell cultures were used as inoculation culture for cultivation of *P. furiosus* in a 15 L bio-fermenter containing the appropriate media (glycolytic after two inoculations, gluconeogenic after two inoculations and gluconeogenic after five inoculations). For induction of heat shock response *P. furiosus* 15 L bio-fermenter culture (minimal starch-medium; glycolytic after two inoculations) was heated after reaching a cell density of $1*10^8$ cells/ml to 107°C. *P. furiosus* cells were further incubated at 107°C for 30 min (glycolytic heat shock).

P. furiosus strain MURPf5 containing the over expression plasmid pYS5a was generated by Kreuzer (2009) and stored. Adaptation of the cells to gluconeogenic conditions was archived by two inoculations on minimal pyruvate-medium. After adaptation, cells were used for cultivation in a 15 L bio-fermenter using minimal pyruvate-medium. All media used for cultivation of strain MURPf5 contained the antibiotic simvastatin (Waege et al., 2010).

2.2.2. Purification of total genomic DNA of *P. furiosus*

Total *P. furiosus* genomic DNA was isolated using the ReliaPrep™ gDNA Tissue Miniprep System. 1-2 ml cell culture were harvested and processed according to the Standard Protocol for Animal Tissue (3.B.) described in the manual.

2.2.3. Standard PCR reactions

All standard PCR reactions were carried out using the Phusion High fidelity DNA polymerase according to the manufactures instructions. The all-four dNTPs for the PCR-reactions were obtained from Peqlab. The amount of template DNA varied from 0.5 ng per reaction for plasmid DNA to 5 ng per reaction for genomic DNA. Optimal primer annealing temperatures were determined by gradient PCRs (55°C to 65°C) using the Doppio Cycler (VWR). The results of the PCR reactions were analyzed using agarosegel electrophoresis.

2.2.4. Quantitative real-time PCR (qPCR)

qPCR reactions were assembled as duplicates or triplicates in a total volume of 10 µl using the SensiMix™ SYBR® No-ROX Kit. Primers were added to a final concentration of 0.3 µM and the total volume of the DNA samples in each reaction was 4 µl. No-template-control using EB buffer (QIAquick PCR purification kit) was included for every primer pair. qPCR reactions were run on a Rotorgene6000 platform (Qiagen, Hilden) using a three step protocol. Data analysis was done using the corresponding Rotorgene software package. Only qPCR reactions with PCR efficiency of 0.8 to 1.2 for the corresponding primer pair were considered. Moreover, duplicate or triplicate reactions with a standard deviation (SD) > 0.5 Ct values were excluded from analysis.

2.2.5. Southern blot analysis

Southern blot analysis was done as described previously (Waege et al., 2010). Total genomic DNA was digested with BamHI and SmaI. For probe labeling DNA fragments were amplified by PCR. One Probe specifically detects a 7 kb fragment harboring the gene PF1743 and the other a 3 kb fragment with the gene PF1753. Molar ratios of both probes were adjusted to achieve comparable signal intensities.

2.2.6. Copy number analysis

Genomic DNA was extracted from the following *P. furiosus* cell cultures: recultivation culture (rich-medium) and from three biological replicates after two or five inoculations in minimal-starch, -pyruvate or –maltose medium. Copy number of the genes PF1753 and PF1743 was determined by qPCR and the deltadeltaC_t ($\Delta\Delta C_t$) method (Pfaffl, 2001) using the gene PF1784 as calibrator and the recultivation culture as reference for the cells grown under the three different conditions. The results are represented as mean of the three biological replicates with SD.

2.2.7. Chromatin immunoprecipitation (ChIP)

Formaldehyde crosslinking was done according to Vierke (2007) and Liu et al. (2007). After the cells reached a cell density of $1.0*10^8$ to $1.5*10^8$ cells/ml (middle to late exponential phase) fixation was carried out directly in the bio-fermenter at 95°C or 107°C (30 min) with a final concentration 0.1 % (v/v) formaldehyde. After 20-30 s the crosslinking reaction was stopped adding glycin to a final concentration of 15 mM and the bio-fermenter was straight away cooled down to 20°C for harvesting the crosslinked *P. furiosus* cells.

Formaldehyde-treated cells were resuspended in 1x PBS and disrupted by sonication using the Branson Sonifier 250 until an average fragment length of 250 bp to 500 bp was obtained. Insoluble particles were removed by centrifugation (21,000 to 100,000 rcf). After freezing with liquid nitrogen, the cell extracts were stored at -80°C. For determination of DNA concentration and fragment length 1 volume of cell extract was mixed with 4 volumes ChIP Elution Buffer (10 mM Tris/Cl pH 8.0, 1% (w/v) SDS, 0.1 mM EGTA) and incubated overnight at 65°C. After RNase treatment (add 1 µl 20 mg/ml solution) DNA was purified via the QIAquick PCR Purification Kit and concentration was measured.

Antigen	amount polyclonal antibodies coupled to 50 µl magnetic beads
TrmBL1	5 µg
TrmBL2	2.5 µg
TFB-RF1	10 µg
Phr	1 µg

For immunoprecipitation various amounts of polyclonal antibodies raised against the protein-of-interest (see above) were coupled to 50 µl Protein G magnetic beads (Dynal, life technologies) according to the manufactures instructions. Antibody coupled magnetic beads were resuspended in 500 µl *P. furiosus* cell extracts adjusted to a total DNA amount of 15 µg in PBS buffer and incubated overnight at 4°C. Immunoprecipitated complexes were washed in total five times with 500 µl of the following washing buffers: 2x Low Salt Buffer (0.1 % (w/v) SDS, 0.1 % (v/v) Triton X-100, 150 mM NaCl, 2 mM EGTA, 20 mM Tris/Cl pH 8.0), 1x High Salt Buffer (0.1 % (w/v) SDS, 0.1 % (v/v) Triton X-100, 500 mM NaCl, 2 mM EGTA, 20 mM Tris/Cl pH 8.0) 1x LiCl Detergent (0.5 M LiCl, 1 % NP-40, 1 % deoxycholic acid, 100 mM Tris/Cl pH 9.0) and 1 x TE. Elution from the beads was done in 100 µl ChIP Elution Buffer at 65°C for 10 minutes. A second elution step was done without heating using 150 µl TE buffer supplemented with 0.67 % (v/v) SDS and both eluates were combined. For the input sample 200 µl TE supplemented with 1 % (w/v) SDS was added to 50 µl not immunoprecipitated *P. furiosus* cell extract (1,5 µg total DNA, 10 % of the IP). Eluted complexes and input samples were incubated overnight at 65°C for reversal of the crosslink. After treatment with RNaseA (add 1 µl 20 mg/ml RNaseA solution and incubate 30 min at 37°C) and ProteinaseK (add 5 µl 20 mg/ml ProteinaseK and 245 µl TE and incubate for 2 h at 50°C) ChIP enriched and input DNA was purified via the QIAquick PCR purification kit and stored at -20°C. For ChIP-qPCR analysis elution volume was 100 µl EB and for ChIP-seq analysis elution volume was 50 µl EB.

At least three replicates of IP per growth condition were conducted for ChIP-qPCR analysis and for ChIP-seq experiments two replicates of IP per growth condition (pyruvate 2; pyruvate 5; starch 2; heat shock) were analyzed.

2.2.8. Library preparation and sequencing

Library preparations were done according to the NEBNext ®-ChIP-Seq library prep reagent set for illumina protocol. For multiplex sample preparation the NEBNext® Multiplex oligos (Index Primers set 1 and 2) were used and libraries were PCR amplified by the NEBNext®High Fidelity Master Mix. Libraries were pooled in equimolar ratios and sequenced using the Illumina HiSeq2000 platform (read length = 50 bp) by the KFB Regensburg. For further analysis only the demultiplexed and quality filtered reads (Eland, Illumina) were used. Additionally, quality statistics were carried out using the Galaxy server (Blankenberg et al., 2010).

2.2.9. Data processing and peak calling TrmBL1

Reads were mapped to the *P. furiosus* DSM3638 genome using Bowtie2 with default settings (Langmead and Salzberg, 2012). Aligned and unaligned reads were written to different files. The Integrative Genome Viewer (IGV) was used for visualization of short-reads genome occupancies (Robinson et al., 2011). The Bam file of aligned reads was converted to the Sam format by Samtools 1.2 (Li et al., 2009) and from each sample 10 million reads were randomly selected (usegalaxy.org; Blankenberg et al., 2010). This did not affect the overall result (tested three times for every sample; data not shown), but it allowed to set a defined enrichment ratio as general cut-off for identified binding sites in all samples. After converting back to the Bam format, samples were analyzed with the Pique software package (Wilbanks et al., 2012) using the following settings: a = 50; l = 300; ´analysis_region´: 0 1908256 (whole genome); ´norm_region´ 1494600 1495600 (*gdh* promoter region) and ´mask´ 1613140 1629427 (PF1937-PF1951). Called peaks were verified by repeating the analysis using additional genomic regions or no specified genomic region as ‘norm_region’. Downstream analysis was performed as recommended for the software. Statistical analysis was carried out using SPSS Statistics (IBM). In addition, only binding sites were considered for further analysis, which were present in both replicates of IP for each of the three analyzed conditions (after 2 inoculations on starch and after 2 or 5 inoculations on pyruvate) and which displayed an enrichment ratio > 1.5 +/- 0.05. Position of the TGM in the TrmBL1 binding sites was identified by using MEME (Bailey et al., 2009) and a Gibbs sampler for *de novo* motif discovery (Favorov et al., 2005).

2.2.10. Data processing and peak calling TrmBL2

Reads were mapped to the *P. furiosus* DSM3638 genome using Bowtie2 with default settings (Langmead and Salzberg, 2012). Aligned and unaligned reads were written to different files. Genome occupancies of short-reads were visualized by IGV (Robinson et al., 2011). Samples were analyzed using the Pique software package (Wilbanks et al., 2012) applying the following settings: a = 50; l = 300; ´analysis_region´: 0 1908256 (whole genome); ´norm_region´ 1494600 1495600 (*gdh* promoter region) and ´mask´ 1613140 1629427 (PF1937-PF1951). First, called peaks were verified by repeating the analysis using additional genomic regions or no specified genomic region as ‘norm_region’. In addition, clustering called peaks and differing called peaks (replicates of IP and different growth conditions) were

verified using the Genetrack algorithm (Albert et al., 2008). Peak calling was carried out using default settings. As initial threshold value the twice average value determined in the *gdh* promoter region (1494600-1495600) was used. In each enriched region (peak) the 20 bp contig with the highest value of the forward and reverse strand was chosen and the TrmBL2 binding site was defined in the center of both. The final threshold for Pique as well as Genetrack peak calling was determined during the combined analysis process for each ChIP experiment individually. Downstream analyses were performed as recommended for the softwares. Statistical analysis was carried out using SPSS Statistics (IBM). In addition, only binding sites were considered for further analysis, which were present (one of the two peak callers) in both replicates of IP for each of the analyzed conditions. *De novo* motif discovery using the identified TrmBL2 binding sites was conducted by using the MEME suite (MEME; MEME-ChIP; DREME; GLAM2; Centrimo; Bailey et al., 2009), a Gibbs sampler (Favorov et al., 2005) and other.

2.2.11. Data processing and peak calling TFB-RF1 (PF1088)

Reads were mapped to the *P. furiosus* DSM3638 genome using Bowtie2 with default settings (Langmead et al., 2009). Aligned and unaligned reads were written to different files. The Integrative Genome Viewer (IGV) was used for visualization of short-reads genome occupancy (Robinson et al., 2011). The Bam file of aligned reads was converted to the Sam format by Samtools 1.2 (Li et al., 2009) and from each sample 10 million reads were randomly selected (usegalaxy.org; Blankenberg et al., 2010). After converting back to the Bam format samples were analyzed with the Pique software package (Wilbanks et al., 2012) using the following settings: a = 50; l = 300; ´analysis_region´: 0 1908256 (whole genome); ´norm_region´: 1660000 1661000 (*pfk* promoter region) and ´mask´ 1613140 1629427 (PF1937-PF1952). Called peaks were verified by repeating the analysis using additional genomic regions or no specified genomic region as ‘norm_region’. Downstream analyses were performed as recommended for the software. Statistical analysis was carried out using SPSS Statistics 21 (IBM). In addition, only binding sites were considered for further analysis, which were present in the three replicates of IP and which displayed an enrichment ratio > 1.5 +/- 0.05.

2.2.12. ChIP-qPCR analysis

ChIP experiments were done as described above. ChIP enrichment was measured by qPCR using the % input method. Considering that only 10 % of the IP sample volume was used for the input sample % input was calculated by the formula: 100 * 2 ^ (Ct (adjusted Input) - Ct (IP)). % input values are shown as mean of at least three replicates of IP with SD.

2.2.13. Antibodies production and pre-purification

Polyclonal rabbit antibodies were produced by Davids Biotechnology (Regensburg) using recombinantly expressed and purified proteins (see below). Purifications of the polyclonal antibodies were done using 1 ml Protein G columns (GE Healthcare, Freiburg) according to the manufactures instructions. Antibody containing fractions were pooled and dialyzed in PBS overnight. The protein concentrations were determined by Bradford assay.

Antigen	Pre-purification	Antigen source	Western blot
TrmB	(Reichelt, 2009)	(Lee et al., 2008)	primary
TrmBL1	(Reichelt, 2009)	(Lee et al., 2008)	primary
TrmBL2	(Reichelt, 2009)	(Lee et al., 2008)	primary
TFB-RF1	(Winter, 2012)	(Ochs, 2009; Ochs et al., 2012)	primary
Phr	(Vierke, 2007)	(Vierke, 2007)	primary
Goat anti-rabbit IgG Dylight 649 Conjugate	- ; Highly Cross-Absorbed	Thermo Scientific (# 35566)	secondary

2.2.14. SDS-PAGE and western blot analysis

The recombinantly expressed and purified proteins were obtained as previously described: TrmB, TrmBL1 and TrmBL2 according to Lee et al. (2008); TFB-RF1 (PF1088) according to Ochs et al. (2012). Cell extracts were prepared from 20 ml cell cultures with a cell density of approximately 1x10*8 cells/ml. After harvesting cells were resuspended in PBS containing protease inhibitor mix (final concentration: 1x cOmplete ultra, 20 mM PMSF, 1 µg/ml pepstatin A and 1 µg/ml leupeptin) and treated with glass beads using a FastPrep24 (M.P.Biomedicals, Irvine, CA) for cell lysis. If cell lysis by bead-beating did not lyse the cells (e.g. after formaldehyde-crosslinking), sonication was used. After RNaseA (add 1µl 20mg/ml solution) and DNaseI (add 1µl 1U/µl solution) treatment, cell debris was removed by centrifugation (21,000-100,000 rcf). The protein concentrations of the supernatants were determined by Bradford assay.

SDS-polyacrylamide gelelectrophoresis (SDS-PAGE) was carried out using the Mini-PROTEAN Tetra Cell (Bio-Rad, Hercules, CA, USA) system. Sample buffer, running buffer and gel formulations were used as described in the corresponding instruction manual according to Laemmli (1970). Usually proteins were denaturated for 5 min at 95°C. After electrophoresis proteins were visualized by Coomassie brilliant blue R250 staining (Meyer and Lamberts, 1965) or silver staining (Blum et al., 1987).

For western blot experiments proteins were separated by SDS-PAGE and transferred to a PVDF membrane (Immobilon-PSQ, 0.2µm, Merck Millipore, Billerica, MA, USA) using the Mini Trans-Blot Module (Bio-Rad, Hercules, CA, USA). Transfers were carried out according to manufactures instructions. Transfer buffer (20 % MeOH, 25 mM Tris, 192 mM glycine) was prepared according to Towbin et al. (1979). After transfer overnight PVDF membranes were washed once in TBST (1x TBS supplemented with 0.1 % (v/v) Tween20) and incubated for 1 h in blocking solution (TBST supplemented with 5 % sodium casein). After incubation for 1 h in blocking solution containing the primary antibodies (diluted to the designated concentrations), membranes were washed three-times in TBST, followed by one-time in blocking solution. Finally, membrane was incubated in blocking solution containing the secondary antibodies (1:10,000 dilution), followed by washing three-times in TBST. For detection the Fujifilm FLA-5000 imager was used. If necessary, membranes were re-used via removal of the antibodies using 0.5 M NaOH.

2.2.15. Cloning and purification of PF1476

The correct ORF of PF1476 (4.2; Preliminary characterization of PF1476) was PCR amplified from *P. furiosus* genomic DNA and cloned into pET151TOPO®Vector according to the TOPO® cloning procedure described in the user manual. This enabled expression of PF1476 fused with a polyhistidine (6xHis) and V5 epitope tag on the N-terminus of the protein for purification. After performing of the TOPO® cloning reaction the construct was transformed into OneShot®TOP10 Chemically Competent *E. coli*. Positive clones were analyzed by picking five colonies and culturing overnight in LB medium containing 100 µg/L ampicillin. Plasmid DNAs were isolated and the constructs were confirmed by sequencing using standard T7 Sequencing Primer (Seqlab, Göttingen).

For expression one positive PF1476 pET TOPO® expression construct was transformed into BL21 Star^{TM}(DE3) One Shot cells supplied with the kit. Expression of PF1476 was carried out according to the pET151TOPO®Vector instruction manual in 100 ml LB containing 100 µg/L ampicillin at 37°C. Expression was induced by addition of IPTG to 0.5 mM final concentration (OD_{578} = ~0.6) followed by additional incubation for 4 h at 37°C. Success of induction was analyzed by SDS-PAGE (not shown). Cells were harvested by centrifugation and stored overnight at -20°C.

For protein purification the Äkta Purifier 10 System (GE Healthcare, Freiburg) was used. Cells were resuspended in 1x PBS containing protease inhibitor mix (final concentration: 1x cOmplete ULTRA tablets, 20 mM PMSF, 1 µg/ml pepstatin A and 1 µg/ml leupeptin) and lysozym. After incubation for 1 h at 4°C cells were treated with glass beads using a FastPrep24 (M.P.Biomedicals, Irvine, CA, USA) for cell lysis. After RNaseA and DNaseI treatment, cell debris was removed by centrifugation (21,000 rcf). Supernatant was heated to 70°C for 10 min and precipitated proteins were removed by centrifugation. 1 volume supernatant was mixed with 1 volume 2x NI-NTA binding buffer (2x: 200 mM Tris/Cl pH 8.0, 2 M NaCl, 40 mM Imidazol pH 8.0) and loaded on a HisTrap FF 1ml column (GE Healthcare, Freiburg). After extensive washing (10 column volumes) PF1476 was eluted with 1x NI-NTA elution buffer (100 mM Tris/Cl pH 8.0, 300 mM NaCl, 250 mM Imidazol pH 8.0). Elution fractions were analyzed by SDS-PAGE and PF1476 containing fractions were pooled. Further purification was carried out by size-exclusion chromatography in Superdex buffer (40 mM Na-Hepes pH 7.5, 300 mM NaCl, 2.5 mM $MgCl_2$, 20 % (v/v) glycerol) using the Superdex75 10/300 column (GE Healthcare, Freiburg). Fractions were analyzed by SDS-PAGE and PF1476 containing fractions were pooled and stored at -20°C and -80°C. After size-exclusion chromatography no additional proteins could be detected beside PF1476 in SDS-PAGE (chapter 4.2.; Figure 34, A).

The protein concentration was determined by Bradford assay and was 150 ng/µl. In total 0.5 mg PF1476 protein was purified from 100 ml *E. coli* cell culture.

2.2.16. *In vitro* validation by EMSA, cell-free transcription and footprint experiments

The EMSA, cell-free transcription and footprint experiments mentioned in chapter 3.2.4 and 3.2.5 were carried out by Antonia Gindner and will be described and presented in detail in her PhD thesis (in preparation; 2014).

All non-denaturating polyacrylamide gelelectrophoreses were carried out according the SDS-PAGE running buffer and gel formulations described in chapter 2.2.14, but lacking SDS.

For EMSA experiments, TFB-RF1 was obtained according to Ochs et al. (2012). DNA template for the PF1089 promoter was obtained by annealing (95°C for 3 min and 45 °C for 20 min) of two primers as described by Huber (2011). DNA template for the PF1529 promoter was obtained from *P. furiosus* genomic DNA by PCR amplification. In both templates one of the primers was labeled with Cy5. DNA templates were purified via the QIAquick PCR purification column and DNA concentrations were determined. 150 fmol labeled DNA and various amounts of TFB-RF1 were assembled in a 15 µl reaction volume containing the following buffer: 80 mM Na-HEPES pH 7.5, 0.5 M KCl, 5 mM $MgCl_2$, 0.2 mM EDTA, 10 % (v/v) glycerol, 40 µg/ml BSA, and 0.015 µg/µl poly-dIC as a competitor according to Ochs et al (2012). After incubation at 70°C for 15 min, protein-DNA complexes were analyzed using a non-denaturating 6 % polyacrylamide gel. DNA fragments were visualized with a Fujifilm FLA-5000.

For the PF1476 EMSA experiments DNA templates for the PF0531, PF0972, PF1911 and PF1784 promoter regions were obtained from *P. furiosus* genomic DNA by PCR amplification. One of the primers of each template was labeled with Cy5. DNA templates were purified via the QIAquick PCR purification column (Qiagen) and DNA concentrations were determined. 150 fmol labeled DNA and various amounts of PF1476 were assembled in a 15 µl reaction volume containing the following buffer: 40 mM HEPES pH 7.3, 250 mM NaCl, 2.5 mM $MgCl_2$, 0.1 mM EDTA, 10 % (v/v) glycerol and 0.1 mg/ml BSA. After incubation at 70°C for 10 min poly-dIC was added to a final concentration of 0.015 µg/µl as competitor and reaction were further incubated at 70°C for 10 min. Protein-DNA complexes were analyzed using a non-denaturating 6 % polyacrylamide gel. DNA fragments were visualized with a Fujifilm FLA-5000.

2.2.17. Multiple sequence alignments and generation of phylogenetic trees

The AA sequences of the TrmB_domain of 87 TrmB/TrmBL proteins from 13 *Thermococcales* species were obtained from the Pfam homepage (http://pfam.sanger.ac.uk/).

For the generation of multiple sequence alignments (MSAs) the program T-Coffee in the M-coffee mode was used (DiTommaso et al., 2011). M-Coffee computes its alignments by combining a collection of multiple alignments named a Library. The standard M-Coffee protocol only uses MSA methods.

The phylogenetic tree of TrmB/TrmBL proteins was generated and analyzed using SplitsTrees4 (Huson and Bryant, 2006). SplitsTree4 is an interactive and comprehensive tool for inferring different types of phylogenetic networks from sequences, distances, and trees. The MSA originating from M-Coffee was used to calculate maximum likelihood protein distance estimates based on a JTT model (Jones et al., 1992). The bio-neighbor joining approach was used to generate the tree (Gascuel, 1997). The resulting tree was analyzed by bootstrapping (1000 replications). Topology of the tree was verified using FastTree2 (Price et al., 2010). FastTree 2 enables the very quickly inference of maximum-likelihood phylogenies for huge alignments (Liu et al., 2011). The AA sequence identities of the TrmB_domains between the different TrmB/TrmBL proteins were obtained from MSA generated by Clustal Omega (Sievers et al., 2011).

3. Results

3.1. Occurrences of TrmB family proteins in 13 *Thermococcales* species

Proteins containing the TrmB_domain are universally distributed within the Archaea. Initially, the TrmB and TrmBL proteins were designated regarding their distribution in the four *Thermococcales* species *P. furiosus*, *P. hirokoshii*, *P. abyssi* and *T. kodakarensis* (Table 2) (Lee et al., 2007b).

Table 2. Occurrences of TrmB and TrmBL1 proteins. According to Lee et al., 2007b.

Organism	TrmB	TrmBL1	TrmBL2	TrmBL3	TrmBL4
P. furiosus	PF1743	PF0124	PF0496	PF0661	–
P. horikoshii	PH1034	–	PH0799	–	PH0751
P. *abyssi*	–	–	PAB0838	–	–
T. *kodakarensis*	–	TK1769	TK0471	–	–

This classification was based on a multiple sequence alignment (MSA) using the complete AA sequence of the proteins. However, these TrmB and TrmBL proteins do not share the same domain composition. Whereas TrmBL3 only has the TrmB_domain, TrmB, TrmBL1, TrmBL2 and TrmBL4 proteins consist of two domains: TrmB and Regulator_TrmB. Moreover, TrmBL2 lacks the sugar sensing part of the Regulator_TrmB domain. The only common feature of all proteins represents the TrmB_domain. Thus, here distribution of the TrmB proteins within the *Thermococcales* was studied focusing primarily on this domain. The AA sequences of the TrmB_domains of 87 TrmB/TrmBL proteins from 13 *Thermococcales* species were obtained from the Pfam homepage (http://pfam.sanger.ac.uk/) and aligned. The MSA was used to calculate a phylogenetic tree only based on the wHTH domain. The result is shown in Figure 9. In the unrooted phylogenetic tree certain TrmB/TrmBL proteins form specific clusters, which correlate with the previous classification. In addition, this clustering extended the number of TrmBL proteins from four to ten (TrmBL1 to TrmBL10, classification only proposed for this study). The bootstrap values calculated for the splits of the different clusters are higher than 70 %, which indicates high reliability of the classification (Hillis and Bull, 1993). In contrast, the weak bootstrap values of the splits in the center of the unrooted tree (not shown) suggest that the early divergence of the TrmB/TrmBL proteins cannot be illustrated by the analysis. The AA sequence identities of the TrmB_domains between the different TrmB/TrmBL clusters range from 10 % to 50 % (average of all proteins from the corresponding clusters) and within the clusters it is higher than 50 %, 70 % for the majority of clusters respectively.

Five clusters correspond to the classification of TrmB/TrmBL proteins shown in Table 3. The TrmB cluster comprises five proteins, which are all part of operons encoding sugar specific ABC-transporters analogous to the *P. furiosus* TrmB (PF1743) and the TM-system. TrmBL1 proteins can be found in 10 of 13 species and they are only missing in three not sugar-fermenting *Pyrococcus* species (Kawarabayasi et al., 1998; Cohen et al., 2003; Lee et al., 2011). Both TrmB and TrmBL1 proteins show the same domain architecture as the studied proteins from *P. furiosus* including the sugar sensing domain. Furthermore, each studied species contains one TrmBL2 protein lacking the sugar sensing domain. This proves the universal distribution of TrmBL2 proteins within the *Thermococcales*. In contrast, TrmBL3 and TrmBL4 proteins are restricted to a few species.

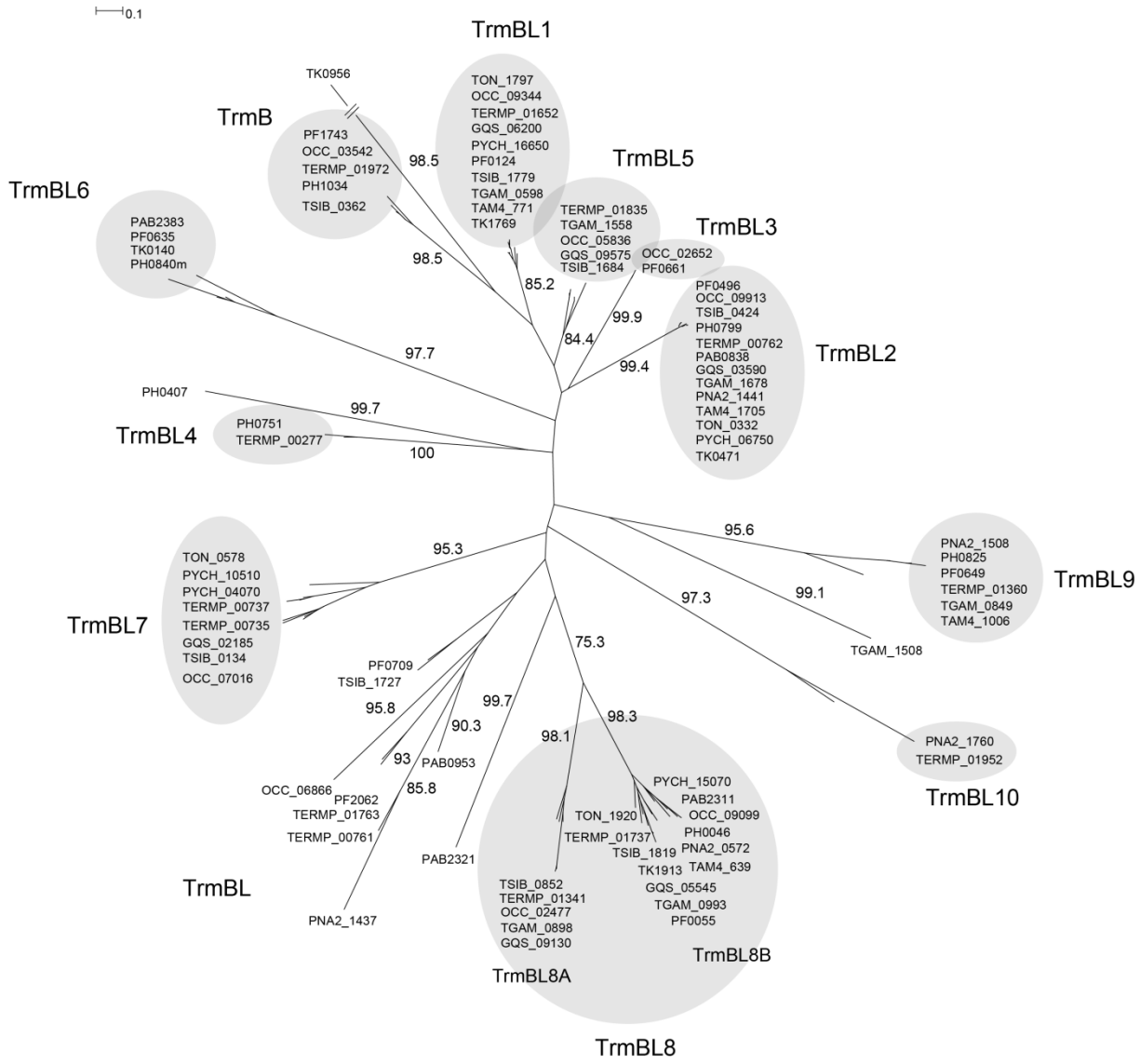


Figure 9. Phylogenetic tree (unrooted form) of TrmB/TrmBL proteins from *Thermococcales*.

The wHTH domains of 87 TrmB_domain containing proteins from the Thermococcales were aligned and the resulting MSA (0) was used to construct a phylogenetic tree. The length of the horizontal bar corresponds to 0.1 substitutions per site. Clustering proteins were labeled according to the previous classification as well as according to the classification described in the text. In total 11 TrmB/TrmBL clusters comprising 75 proteins were found. The remaining 12 proteins were termed as unclassified TrmBL proteins (TrmBL). The following proteins belong to the corresponding proposed clusters: TrmB: OCC_03542, TSIB_0362, TERMP_01972, PH1034, PF1743; TrmBL1: OCC_09344, TAM4_771, TK1769, TGAM_0598, TSIB_1779, TERMP_01652, TON_1797, GQS_06200, PF0124, PYCH_16650; TrmBL2: PNA2_1441, OCC_09913, TAM4_1705, TK0471, TGAM_1678, PAB0838, TSIB_0424, TERMP_00762, PH0799, TON_0332, GQS_03590, PF0496, PYCH_06750; TrmBL3: OCC_02652, PF0661; TrmBL4: TERMP_00277, PH0751; TrmBL5: OCC_05836, TGAM_1558, TSIB_1684, TERMP_01835, GQS_09575; TrmBL6: TK0140, PAB2383, PH0840m, PF0635; TrmBL7: OCC_07016, TSIB_0134, TERMP_00737, TERMP_00735, TON_0578, GQS_02185, PYCH_10510, PYCH_04070; TrmBL8A: OCC_02477, TGAM_0898, TSIB_0852, TERMP_01341, GQS_09130; TrmBL8B: PNA2_0572, OCC_09099, TAM4_639, TK1913, TGAM_0993, PAB2311, TSIB_1819, TERMP_01737, PH0046, TON_1920, GQS_05545, PF0055, PYCH_15070; TrmBL9: PNA2_1508, TAM4_1006, TGAM_0849, TERMP_01360, PH0825, PF0649; TrmBL10: PNA2_1760, TERMP_01952; TrmBL: PNA2_1437, OCC_06866, TK0956, TGAM_1508, PAB0953, PAB2321, TSIB_1727, TERMP_00761, TERMP_01763, PH0407, PF0709, PF2062; Abbreviations: PF, *Pyrococcus furiosus* (strain DSM 3638); PH, *Pyrococcus horikoshii* (strain DSM 12428 / OT-3); PAB, *Pyrococcus abyssi* (strain GE5); PYCH, *Pyrococcus yayanosii* (strain CH1); PNA2, *Pyrococcus sp.* (strain NA2); TK, *Thermococcus kodakaraensis* (strain KOD1); OCC, *Thermococcus litoralis* DSM 5473; TAM, *Thermococcus sp.* AM4; TGAM, *Thermococcus gammatolerans* (strain DSM 15229); TSIB, *Thermococcus sibiricus* (strain DSM 12597); TERMP, *Thermococcus barophilus* (strain DSM 11836 / MP); TON, *Thermococcus onnurineus* (strain NA1); GQS, *Thermococcus sp.* (strain 4557).

Table 3. Occurrences of TrmB, TrmBL1 and TrmBL2 in the *Thermococcales.*

Strain	TrmB_domain proteins	TrmB	TrmBL1	TrmBL2
P. furiosus	9	**PF1743**	**PF0124**	**PF0496**
P. horikoshii	7	PH1034	–	PH0799
P. abyssi	5	–	–	PAB0838
P. yayanosii	5	–	PYCH_16650	PYCH_06750
***P. sp.*(strain NA2)**	5	–	–	PNA2_1441
T. kodakaraensis	5	–	TK1769	TK0471
T. litoralis	9	OCC_03542	OCC_09344	OCC_09913
***T. sp*. (strain AM4)**	4	–	TAM4_771	TAM4_1705
T. gammatolerans	7	–	TGAM_0598	TGAM_1678
T. sibiricus	8	TSIB_0362	TSIB_1779	TSIB_0424
T. barophilus	13	TERMP_01972	TERMP_01652	TERMP_00762
T. onnurineus	4	–	TON_1797	TON_0332
***T. sp.* (strain 4557)**	6	–	GQS_06200	GQS_03590

(Complete Table is shown in Table 11). Studied proteins are depicted in bold.

The additional novel proposed TrmBL clusters 5 to 10 are also characterized by unique features. TrmBL5 proteins are present in five *Thermococcus* species and contain the Regulator_TrmB domain including the sugar-binding part. The TrmB_Regulator domain is exclusively present in the clusters TrmB, TrmBL1, 2, 4, and 5 (composed architecture). Additionally, four of five TrmBL5 proteins are co-transcribed with genes encoding putative sugar ABC transport systems similar to TrmB. In TrmBL6 proteins the TrmB-domain represents the middle domain of the proteins and it is composed together with archaeal ATPase domains (Arch_ATPase; pfam ID: PF01637). Only in one other cluster the TrmB_domain is also not located in the N-terminus of the proteins. These are the TrmBL7 proteins, where the HTH domain is located in the C-terminus. The TrmBL8 cluster contains the highest number of proteins and can be divided in the two subgroups: TrmBL8A and B. The common feature of all TrmBL8 proteins is that they are small-sized (100 to 150 AA) TFs, which only contain the HTH domain. Whereas TrmBL8A proteins are restricted to few *Thermococcus* species, TrmBL8B proteins are universally distributed within the *Thermococcales* according to TrmBL2 proteins. In addition, two other phylogenetic clusters

could be determined designated as TrmBL9 and 10. Beside these clustering proteins (n=75) some TrmBL proteins remained as unclassified TrmBL proteins (TrmBL; n=12). The repertoire of TrmB/TrmBL protein varies between the different *Thermococcales* species. Whereas *T. barophilus* contains 13 different TrmB/TrmBL proteins, only four proteins can be found in *T. onnurineus.*

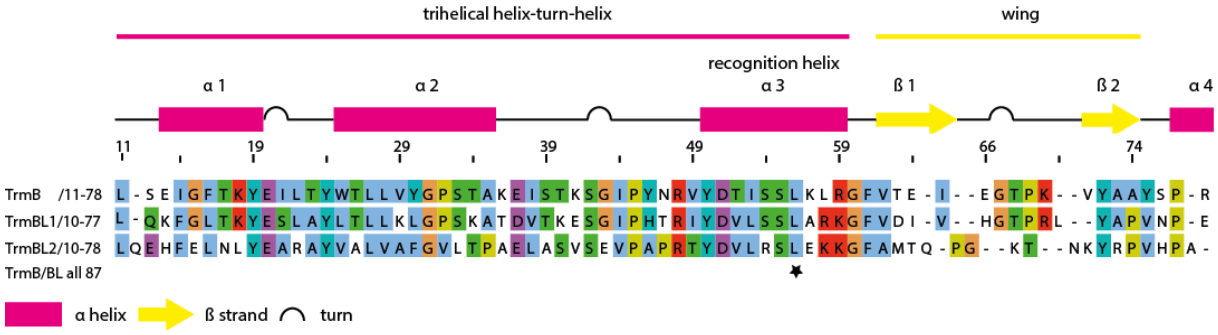


Figure 10. TrmB, TrmBL1 and TrmBL2 show a high similarity with regard to the wHTH domain.

The MSA of TrmB, TrmBL1 and TrmBL2 from *P. furiosus* based on MSA of 87 TrmB_domain containing *Thermococcales* proteins illustrates the high similarity of the proteins regarding the wHTH DBD. Residues are labeled according to the AA sequence of TrmB. The corresponding residues of the TrmB_domains from the three proteins are given next to the protein name. Line four indicates a fully conserved residue (marked by an asterisk) of all aligned 87 proteins (TrmB/BL all 87). Secondary structures are shown according to the crystal structure of TrmB (Krug et al., 2013). The α-helix 3 represents the DNA recognition helix. MSA was modified using Jalview (http://www.jalview.org/) and colors are used according to the Clustalx color scheme (http://www.clustal.org/clustal2/).

In the genome of *P. furiosus* nine TrmB TF family proteins were detected. Two proteins persist as unclassified TrmBL proteins. Furthermore, four proteins can be related to the following clusters: TrmBL3 (PF0661), TrmBL6 (PF0635), TrmBL8B (PF0055) and TrmBL9 (PF0649). This thesis focuses on the proteins TrmB (PF1743), TrmBL1 (PF0124) and TrmBL2 (PF0496). In

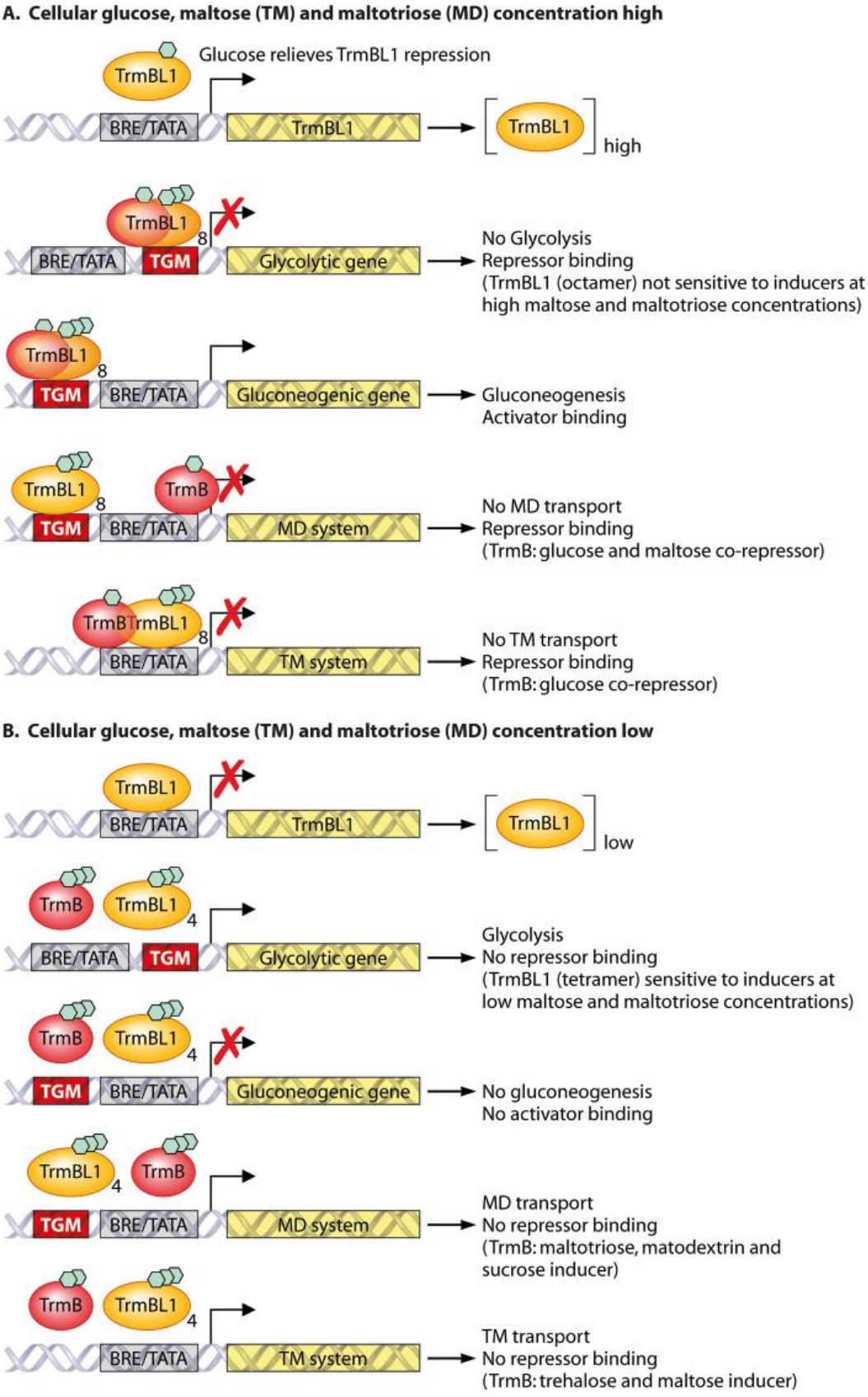
Figure 10 the MSA (based on 87 proteins) of the TrmB_domain of the three proteins is shown. It illustrates the high similarity particularly of the HTH domain of the three TFs. In the recognition helix (α 3) the two residues, tyrosine at position 50 (Y50) and aspartic acid at position 51 (D51) in TrmB, are highly conserved. These residues were shown to be essential for DNA binding of TrmB as well as TrmBL1 *in vitro* (Lee et al., 2007c, 2007b). Moreover, in the MSA of all 87 sequences one single residue is fully conserved in all proteins. This is a leucine at position 46 (L46) in TrmB, which is part of the recognition helix. Beside the similarities of TrmB, TrmBL1 and TrmBL2 regarding their AA composition and domain architecture, the function of the three proteins in transcriptional regulation was shown to be different (Lee et al., 2008; Maruyama et al., 2011). TrmBL2 is suggested to act as abundant chromosomal binding protein, which can act as global repressor. In contrast, TrmB and TrmBL1 mainly function in the sugar metabolism of *P. furiosus*. TrmBL1 represents the global regulator, which can repress or activate gene expression through binding to TGM-containing promoters. Moreover, TrmB presumably inhibits transcription of the TM-system; however it can also bind and regulate targets of TrmBL1 *in vitro*. In the following chapters genome-wide mapping of TF binding sites *in vivo* by ChIP-seq was used to dissect specific binding sites of the three proteins in the genome of *P. furiosus* under varying growth conditions.

3.2. Dissecting the function of TrmB and TrmBL1 in regulation of sugar metabolism of *P. furiosus* by ChIP-seq *in vivo*

So far functional analysis of the bacterial-like TFs TrmB and TrmBL1 from *P. furiosus* was carried out exclusively *in vitro*. This revealed that both TFs function as transcriptional regulators of genes encoding enzymes and proteins involved in sugar metabolism. Regarding some genes cross-regulation of both TFs is assumed. In the corresponding promoters transcriptional regulation depends on the concentration of certain sugars (Lee et al., 2007b; Lee et al., 2008). These sugars can function as co-repressors or inducers, which enhances or abolishes TrmB- and TrmBL1-mediated transcriptional control. Moreover, activity of these co-repressors and inducers can rely on the particular promoter region. Repression of the TM system through TrmB is abolished by its substrates maltose and trehalose, whereas the repression of the MD system is only abrogated in the presence of maltotriose, but not by trehalose and maltose. In contrast, glucose, which cannot be used as a carbon source by *P. furiosus* but is formed endogenously, was shown to act as co-repressor of the TM as well as the MD system for TrmB. For TrmBL1 three sugars can act as inducers of the *pfk* gene: maltose, maltotriose and fructose. Moreover, TrmBL1 mainly regulates genes containing the TGM in their promoter regions, whereas binding of TrmB is independent of the presence of the TGM. It is suggested that *in vivo* especially intracellular glucose may represent the essential messenger molecule, which controls regulation of glycolysis, gluconeogenesis and sugar transport by TrmB and TrmBL1 (Lee et al., 2007b, 2007c, Lee et al., 2008). A model for TrmB and TrmBL1 function in differential regulation of sugar metabolism was recently proposed by Bräsen et al. (2014) (Figure 11). This model shows numerous possible cross-regulations of both TFs. However, *in vivo* these mechanisms have to be proven yet. One first step will be to dissect the specific binding sites of both TrmB and TrmBL1 in the genome of *P. furiosus* cells grown under glycolytic or gluconeogenic growth conditions. Thus, ChIP-seq experiments were carried out to identify genome-wide TrmB/TrmBL1 binding sites under varying growth conditions. This should contribute to better understanding of the differential regulation mediated by the GRN of TrmB and TrmBL1 depending on the presence of a specific carbon-source.

Figure 11. Model for TrmB/TrmBL1 function in differential regulation of genes encoding enzymes catalyzing transport, glycolysis, and gluconeogenesis in *P. furiosus* (Bräsen et al., 2014).

Under *in vivo* conditions, the cellular glucose concentration is assumed to be the major regulatory signal (Lee et al., 2008). (A) At high cellular glucose concentrations, autorepression of the gene encoding TrmBL1 is released, and TrmBL1 is present in high concentrations. High cellular concentrations of TrmBL1 inducers (i.e., maltose and maltotriose) will cause a shift from the tetrameric regulator (inducer sensitive) to the octameric regulator (not responsive to inducers) with high DNA binding affinity. In addition, glucose serves as a co-repressor for TrmB, repressing the TM- and MD-systems. For many promoters tested, cross-regulation was observed for TrmBL1 and TrmB. In general, high cellular glucose concentrations lead to the repression of genes involved in glycolysis and transport (TM- and MD-systems) and activation of gluconeogenic genes. (B) At low cellular glucose concentrations, TrmBL1 expression is repressed. Low concentrations of the inducers maltose and maltotriose result in the formation of the inducer-responsive tetrameric form of TrmBL1 with a low DNA binding capacity. TrmB repression of the MD-system is relieved by maltotriose, maltodextrin, or sucrose (maltose serves as a corepressor), and that of the TM-system is released by trehalose and maltose, thus enhancing sugar uptake. Therefore, at low cellular glucose concentrations, expression of genes involved in transport (TM- and MD-systems) and glycolytic genes will be stimulated, whereas the expression of genes involved in gluconeogenic genes will be reduced. According to Bräsen et al. 2014.


A. Cellular glucose, maltose (TM) and maltotriose (MD) concentration high
Glucose relieves TrmBL1 repression
TrmBL1
BRE/TATA
TrmBL1
TrmBL1
high
TrmBL1
8
BRE/TATA
TGM
Glycolytic gene
No Glycolysis
Repressor binding
(TrmBL1 (octamer) not sensitive to inducers at high maltose and maltotriose concentrations)
TrmBL1
8
TGM
BRE/TATA
Gluconeogenic gene
Gluconeogenesis
Activator binding
TrmBL1
8
TrmB
TGM
BRE/TATA
MD system
No MD transport
Repressor binding
(TrmB: glucose and maltose co-repressor)
TrmB
TrmBL1
8
BRE/TATA
TM system
No TM transport
Repressor binding
(TrmB: glucose co-repressor)
B. Cellular glucose, maltose (TM) and maltotriose (MD) concentration low
TrmBL1
BRE/TATA
TrmBL1
TrmBL1
low
TrmB
TrmBL1
4
BRE/TATA
TGM
Glycolytic gene
Glycolysis
No repressor binding
(TrmBL1 (tetramer) sensitive to inducers at low maltose and maltotriose concentrations)
TrmB
TrmBL1
4
TGM
BRE/TATA
Gluconeogenic gene
No gluconeogenesis
No activator binding
TrmBL1
4
TrmB
TGM
BRE/TATA
MD system
MD transport
No repressor binding
(TrmB: maltotriose, matodextrin and sucrose inducer)
TrmB
TrmBL1
4
BRE/TATA
TM system
TM transport
No repressor binding
(TrmB: trehalose and maltose inducer)

3.2.1. Specificity of the antibodies raised against TrmB and TrmBL1

Antibody specificity plays a crucial role for efficient and reliable ChIP experiments (Orlando et al., 1997). The AA sequence identity of the paralogous TrmB and TrmB-like proteins in *Pyrococcus furiosus* is between 22 % and 30 % (Lee et al., 2007b). Therefore, a potential cross-reactivity of the antibodies raised against TrmB (anti-TrmB IgG) and TrmBL1 (anti-TrmBL1 IgG) was excluded by western blot analysis before using them for ChIP experiments. The anti-TrmBL1 IgG showed a specific reaction for recombinant TrmBL1 Figure 12, A, lane 2) and no cross-reactivity with the paralogues TrmB and TrmBL2 (Figure 12, A, lane 1 and 3). Moreover, one specific signal at 40 kDa (39 426 calculated molecular weight of TrmBL1) was detected in crude extracts from cells grown under glycolytic (starch) or gluconeogenic (pyruvate) conditions (Figure 12, A, lane 4 and 5). In contrast, the antibody raised against TrmB showed a reaction with recombinant TrmB and a cross-reaction with TrmBL1 but not TrmBL2 (Figure 12 B, lane 1 and 2). In both crude extracts (starch or pyruvate) no specific signal for TrmB was detected (Figure 12 B, lane 4 and 5).

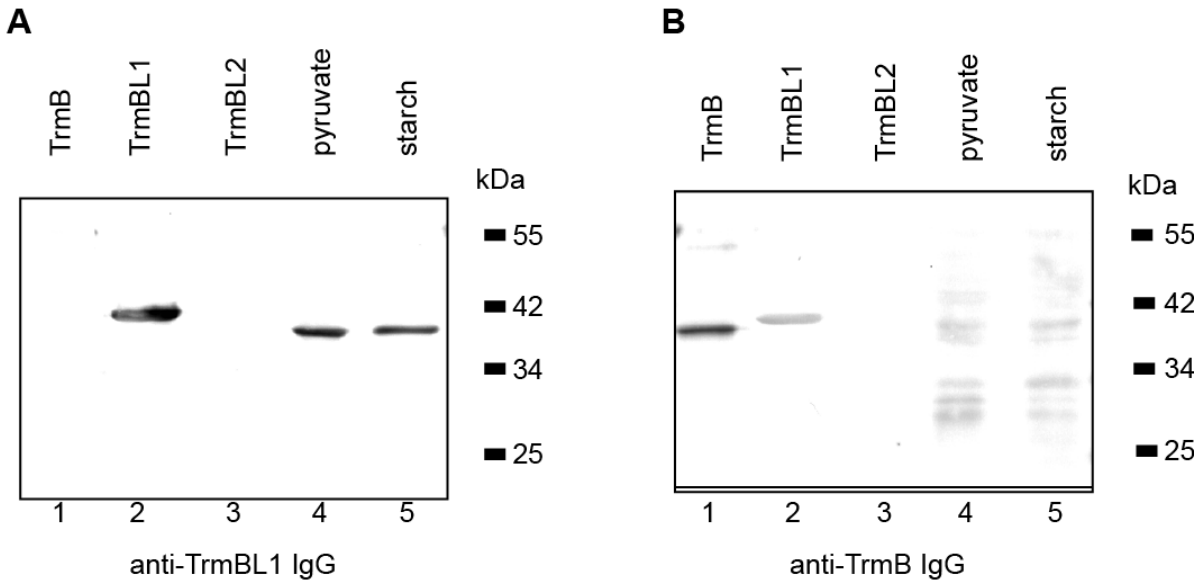


Figure 12. Western blot analysis using the anti-TrmBL1 IgG and the anti-TrmB IgG.

100 ng each recombinant TrmB (lane 1), TrmBL1 (lane 2) and TrmBL2 (lane 3) were used and 20 µg crude extract obtained from *P. furiosus* cells, which were grown under gluconeogenic (pyruvate, lane 4) or glycolytic (starch, lane 5) conditions. A, using purified antibodies raised against recombinant TrmBL1 (anti-TrmBL1 IgG, 1:2000) for detection. B, using purified antibodies raised against recombinant TrmB (anti-TrmB IgG, 1:2000) for detection.

Therefore, the anti-TrmBL1 IgG could be used for the ChIP experiments due to its high specificity, whereas the anti-TrmB IgG is not suitable for ChIP experiments because of its cross-reactivity to TrmBL1.

3.2.2. Known, predicted and novel TrmBL1 binding sites were found by ChIP-seq

TrmBL1 ChIP-seq experiments were conducted using crude extracts from formaldehyde treated *P. furiosus* cells grown under glycolytic (2 inoculations on starch) or gluconeogenic (2 or 5 inoculations on pyruvate) conditions. In total, 28 TrmBL1 binding sites were identified under both growth conditions (Table 4; Table 5; Table 12; Table 13; and Figure 13). Most binding events (n = 25) were found under gluconeogenic growth conditions, whereas under glycolytic conditions only four sites were detected. Approximately all TrmBL1 binding sites (96 %) identified under gluconeogenic conditions are located in close proximity to promoter regions. This was also true for the only overlapping TrmBL1 binding site found under glycolytic as well as gluconeogenic growth conditions. This is located in the promoter region of a gene encoding an alpha-amylase (PF0272). In contrast, the three glycolytic-specific sites were detected in coding regions of genes. Moreover, an aberration in the genomic coverage of mapped reads was found for the IP as well as input samples from cells grown on starch. It spans the genome from position 628,000 to 797,000 bp and its physiological relevance is still unclear. Additionally, in all samples almost no mapped reads were found in the *P. furiosus* chromosomal region from 1,613,140 bp to 1,629,427 bp, which indicates deletion of this region in the genome.

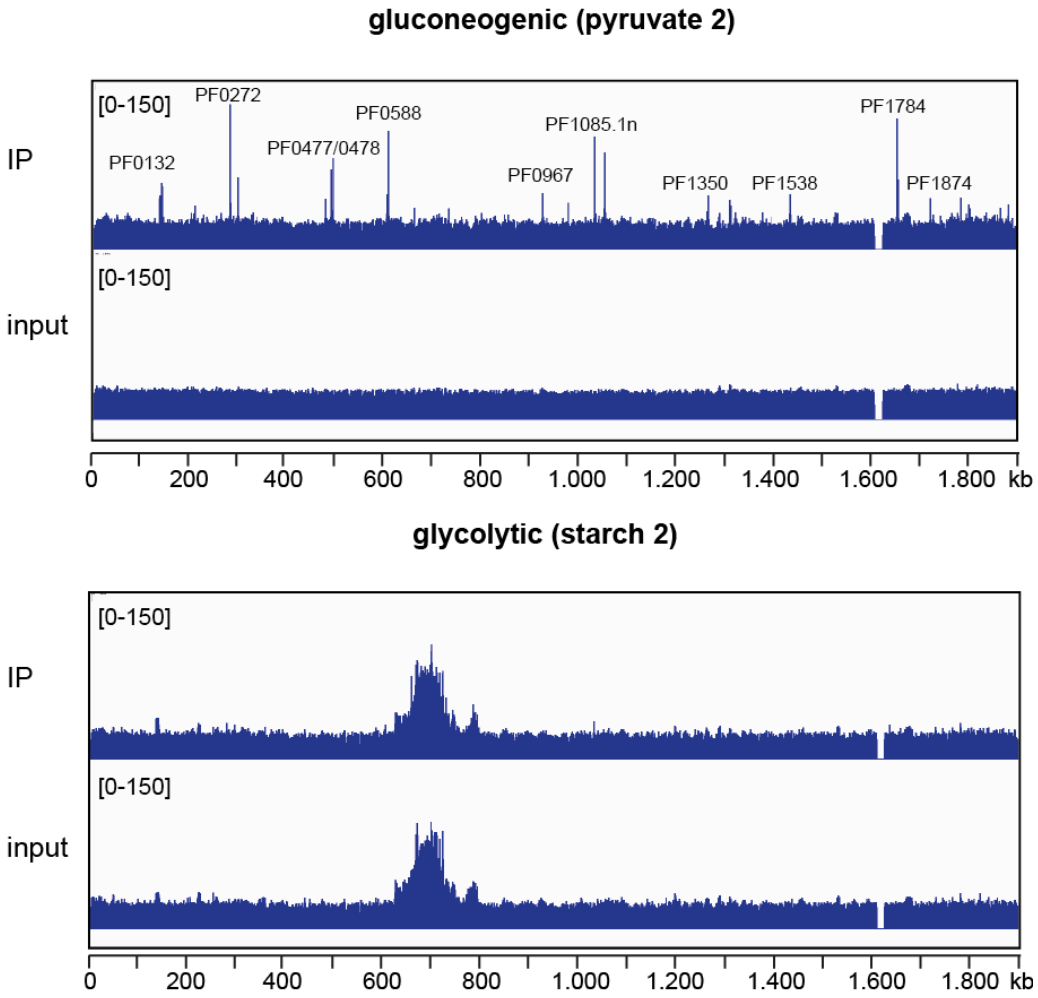


Figure 13. TrmBL1 mainly binds to the genome under gluconeogenic growth conditions.

TrmBL1 ChIP-seq experiments with *P. furiosus* cells grown under gluconeogenic (after 2 inoculations on pyruvate) and glycolytic (after 2 inoculations on starch) conditions were carried out. Mapped TrmBL1 immunoprecipitation (IP) and Input reads were visualized for the whole genome of *P. furiosus*. Prominent peaks are denoted.

Table 4. Selected known and predicted genes identified by ChIP-seq.

transcript organization [5]	gene	gene product	additional information	pathway/ biological process	ref.
operon	**PF0132**	GLS	-	starch and sucrose metabolism	[1]
	PF0133	hypothetical protein	-	-	-
singleton	**PF0196**	PGI	-	glycolysis / gluconeogenesis	[2]
singleton	**PF0272**	AMY1	-	starch and sucrose metabolism	[3]
singleton	**PF0464**	GAPOR	-	glycolysis	[2]
singleton	**PF0477**	AMY2	-	starch and sucrose metabolism	[3]
singleton	**PF0478**	AMY3	-	-	[3]
singleton	**PF0588**	PGM	-	starch and sucrose metabolism	[3]
singleton	**PF1109**/10	SB	one ORF	starch and sucrose metabolism	[3]
singleton	**PF1784**	PFK	-	glycolysis	[2]
singleton	**PF1874**	GAPDH	-	gluconeogenesis	[2]
operon	PF1933	ATPase	PF1933 and PF1934 are one gene	ABC transporters starch and sucrose metabolism	[3]
	PF1934	hypothetical protein			
	PF1935	amylopullulanase	MD system		
	PF1936	malG-like			
	PF1937	malF-like			
	PF1938	malE-like			
singleton	**PF1956**	FBPA	-	glycolysis / gluconeogenesis	[2]
singleton	**PF1959**	PGM	-	glycolysis / gluconeogenesis	[2]

Genes with an identified TrmBL1 binding site (ChIP-seq) in their promoter region are depicted in bold. Enzymes and transporters: GLS, proposed α-glucosidase; PGI, phospho-glucose isomerase; AMY1, proposed 4-α-glucanotransferase; GAPOR, GAP:Fd oxidoreductase; AMY2, proposed extracellular α-amylase; AMY3, proposed extracellular cyclomaltodextrin glucano-transferase; PGM, phosphoglycerate mutase; SB-protein, Starch binding protein (PF1109-1110; PFK, phospho-fructokinase; ; GAPDH, GAP dehydrogenase; MD-system, maltodextrin-specific ABC transporter (PF1938 to PF1933); FBPA, F-1,6-BP aldolase; POR, pyruvate ferredoxin oxidoreductase. References: [1],Comfort et al., 2008; [2], Siebers and Schönheit, 2005; [3],Lee et al., 2006a; [4],Schut et al., 2003; [5], Yoon et al., 2011.

Table 5. Selected novel genes identified by ChIP-seq.

transcript organization [2]	gene	gene product	additional information	pathway/ biological process	ref.
singleton	**PF0287**	pyrolysin	-	proteolysis and peptidolysis	-
operon	PF0965	POR subunit beta	operon PF0971 to PF0965 [5]	glycolysis / gluconeogenesis	-
	PF0966	POR subunit alpha			
	PF0967	POR subunit delta			
singleton	**PF1025**	hypothetical protein	conserved	-	-
singleton	**PF1085.1n**	hypothetical protein	-	-	-
singleton	**PF1350**	major facilitator superfamily protein	transporter	transporters	-
singleton	**PF1476**	predicted transcriptional regulator	PadR family	transcriptional regulation	-
operon	PF1535	alpha-glucan phosphorylase	same promoter region PF1539	starch and sucrose metabolism	[1]
	PF1536	hypothetical protein		-	
	PF1537	hypothetical protein		-	
	PF1538	N-ethylammeline chlorohydrolase		nucleotide degradation	
singleton	**PF1539**	dihydroorotate dehydrogenase 1B	see above	pyrimidine metabolism	-
singleton	**PF2047**	l-asparaginase	-	alanine, aspartate and glutamate metabolism	-

Genes with an identified TrmBL1 binding site in the promoter region are depicted in bold. Enzymes: POR, pyruvate ferredoxin oxidoreductase. References: [1], Lee et al., 2006a; [2]; Yoon et al., 2011.

A variety of TrmBL1 binding sites identified under gluconeogenic growth conditions were already studied by *in vitro*, *in vivo* or *in silico* analysis (VanDeWerken et al., 2006; Kanai et al., 2007; Lee et al., 2007b; Lee et al., 2008). Corresponding regulated genes mainly encode enzymes catalyzing glycolysis, gluconeogenesis and the starch and sucrose metabolism (Table 4). This includes several possible alpha-glucans hydrolyzing enzymes (PF0272, PF0477, PF0478 and PF1935), which function at various steps of starch degradation to glucose (Lee et al., 2006a; Vanfossen et al., 2008). Moreover, ChIP-seq confirmed transcriptional control of the MD-system (PF1938-PF1933) by TrmBL1 *in vivo*, which encodes for the maltooligosaccharides-specific ABC transporter in *P. furiosus* (Koning et al., 2002; Lee et al., 2007b). Specific glycolytic genes regulated by TrmBL1 encode the GAPOR (PF0464) and the ADP-dependent PFK (PF1784). Furthermore, TrmBL1 only recognized the promoter region of one specific gluconeogenic gene, which encodes the GAPDH (PF1874).

In addition to the known and/or predicted sites, the TrmBL1 ChIP-seq analysis also revealed several unknown and novel TrmBL1 binding sites in the genome of *P. furiosus* (Table 5). This includes genes encoding hypothetical (PF1085.1n) or conserved hypothetical (PF1025) proteins with unknown function; however, other genes can be associated with defined biological processes. One new identified binding site was located in the intergenic region upstream of the POR subunit delta gene (PF0967), which is part of a polycistronic operon transcribed from PF0971 to PF0965. Moreover, ChIP-seq showed TrmBL1-mediated transcriptional control of a major faciliator superfamily (MFS) protein (PF1350), which is

supposed to act as transmembrane transporter. Additional novel TrmBL1 binding sites were found in the promoter region of a predicted transcriptional regulator of the PadR family (PF1476) and upstream of a gene encoding a l-asparaginase (PF2047).

In summary, using ChIP-seq to identify TrmBL1 binding sites in a genome wide manner for *P. furiosus in vivo* revealed an extended function of TrmBL1 as global regulator of metabolism. TrmBL1 not only represses or activates genes involved in sugar uptake, glycolysis and gluconeogenesis. TrmBL1 also affects transcription of genes involved in various additional metabolic pathways and biological processes as proteolysis, transcriptional regulation, nucleotide degradation and AA metabolism.

3.2.3. Validation of identified TrmBL1 binding sites *in vivo* by ChIP-qPCR

Selected TrmBL1 binding sites identified by ChIP-seq were confirmed by a ChIP-qPCR assay. For cells grown under gluconeogenic growth conditions ChIP enrichment could be shown for all analyzed TrmBL1 binding sites in relation to the promoter region of the *gdh* gene (PF1602) (Figure 14, A). In addition, under glycolytic growth conditions ChIP enrichment was only observed for two regions (PF0272 and PF1761), which were also found during the ChIP-seq analysis (Figure 14, B). Moreover, quantification of ChIP enrichment found by ChIP-qPCR correlates well with the ChIP enrichment ratio determined by ChIP-seq ($\rho = 0.89$; $P = 0.001$; $n = 10$; Spearman`s rank correlation). Antibodies raised against the Phr served as negative control for specific enrichment using the TrmBL1 specific antibody for immunoprecipitation (Vierke et al., 2003; Liu et al., 2007). Using the anti-Phr IgG no enrichment of the promoter region of the *pfk* gene (PF1784) was detected under both growth conditions. This was one of the highly enriched regions in the TrmBL1 ChIP under gluconeogenic conditions. As expected, the previously identified Phr binding site in the promoter region of the *aaa+atpase* gene (PF1882) showed a strong ChIP-enrichment under both conditions (Figure 14, C and D). This demonstrates that the efficiency of formaldehyde crosslinking under both growth conditions was sufficient for successful ChIP experiments and that the almost complete absence of TrmBL1 binding events under glycolytic growth conditions is specific for TrmBL1.

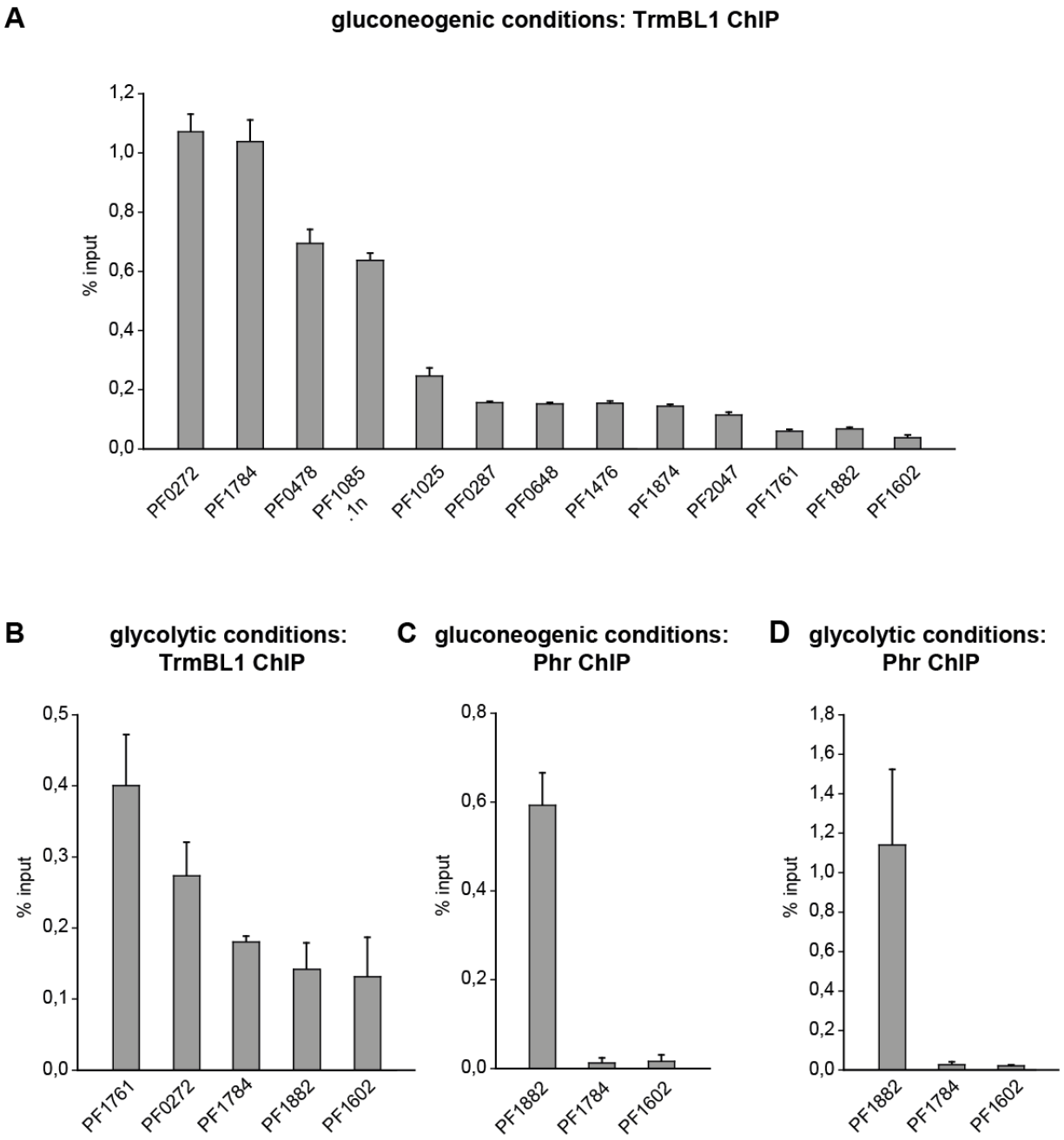


Figure 14. ChIP-qPCR validation of selected TrmBL1 binding sites identified by ChIP-seq.

ChIP enrichment is presented as % input. The mean with standard deviation of at least three replicates of immunoprecipitation is shown for all analysed genomic loci. A, TrmBL1 ChIP of cells grown under gluconeogenic condition (pyruvate, after 2 inoculations). The genes PF1882 (*aaa+atpase*) and PF1602 (*gdh*) represent the negative controls for the anti-TrmBL1 IgG. B, TrmBL1 ChIP of cells grown under glycolytic growth condition (starch, after 2 inoculations). The genes PF1882 (*aaa+atpase*) and PF1602 (*gdh*) represent the negative controls for the anti-TrmBL1 IgG. C, ChIP with a Phr specific antibody using cells grown on pyruvate (gluconeogenic conditions, after 2 inoculations). The gene PF1882 (*aaa+atpase*; Liu et al, 2007) is the positive control for the anti-Phr IgG, whereas the genes PF1784 (*pfk*) and PF1602 (*gdh*) are the negative controls for the anti-Phr IgG. D, ChIP with a Phr specific antibody using cells grown on starch (glycolytic conditions, after 2 inoculations). The gene PF1882 (*aaa+atpase*) is the positive control for the anti-Phr IgG, whereas the genes PF1784 (*pfk*) and PF1602 (*gdh*) are the negative controls for the anti-Phr IgG.

3.2.4. Spatial resolution of the ChIP-seq approach enables discrimination of TrmBL1 binding upstream or downstream of the promoter elements

Transcriptional activation or repression caused by TrmBL1 relies on binding upstream or downstream of the archaeal promoter elements. Moreover, the TGM represents the cis regulatory element for TrmBL1-mediated transcriptional control and is accordingly located upstream or downstream of the BRE and TATA-box (VanDeWerken et al., 2006; Kanai et al., 2007; Lee et al., 2008).

In all identified binding sites under glycolytic and gluconeogenic growth conditions the presence of a TGM could be detected by analysis with MEME and a Gibbs sampler for motif discovery (Favorov et al., 2005; Bailey et al., 2009) (Table 13). For analysis of the spatial resolution of the ChIP-seq approach, the identified TrmBL1 binding sites (consensus of the two replicates of immunoprecipitation for the tested conditions) were plotted regarding their distance relative to the center of the TGM. The median of 9.5 (mean = 36) is quite close to the TGM, which spans the motif center by 8 bp upstream and downstream (Figure 15, A). Outliers mostly represent binding sites with low ChIP enrichment ratios. In addition, binding of TrmBL1 to the TGM was proved by *in vitro* DNaseI footprinting of eight selected sites (experiments were carried out by Antonia Gindner and will be presented in detail in her PhD thesis (in preparation)). For the analysis five known and predicted binding sites and three novel regions were used. Moreover, two binding sites are located upstream of the promoter, while the other sites are located downstream of the BRE and TATA-box. All TrmBL1 footprints include the TGM and extend the motif upstream and downstream by 5.94 +/- 1.73 bp each. The complete TrmBL1 occupied genomic region in each binding site comprises approximately 28 bp. Seven of eight analyzed ChIP-seq TrmBL1 binding sites are located within the footprinting region and just one site is located within 10 bp upstream of the footprinting region (Figure 15, B). This proves the high spatial resolution of the ChIP-seq assay to determine protein binding sites *in vivo*. Furthermore, it allows discrimination between TrmBL1 binding upstream and downstream of the promoter and showed that positioning of the TGM *in silico* overlaps with the location of the found TrmBL1 binding sites *in vivo* (Table 6, Table 12 and Table 13)

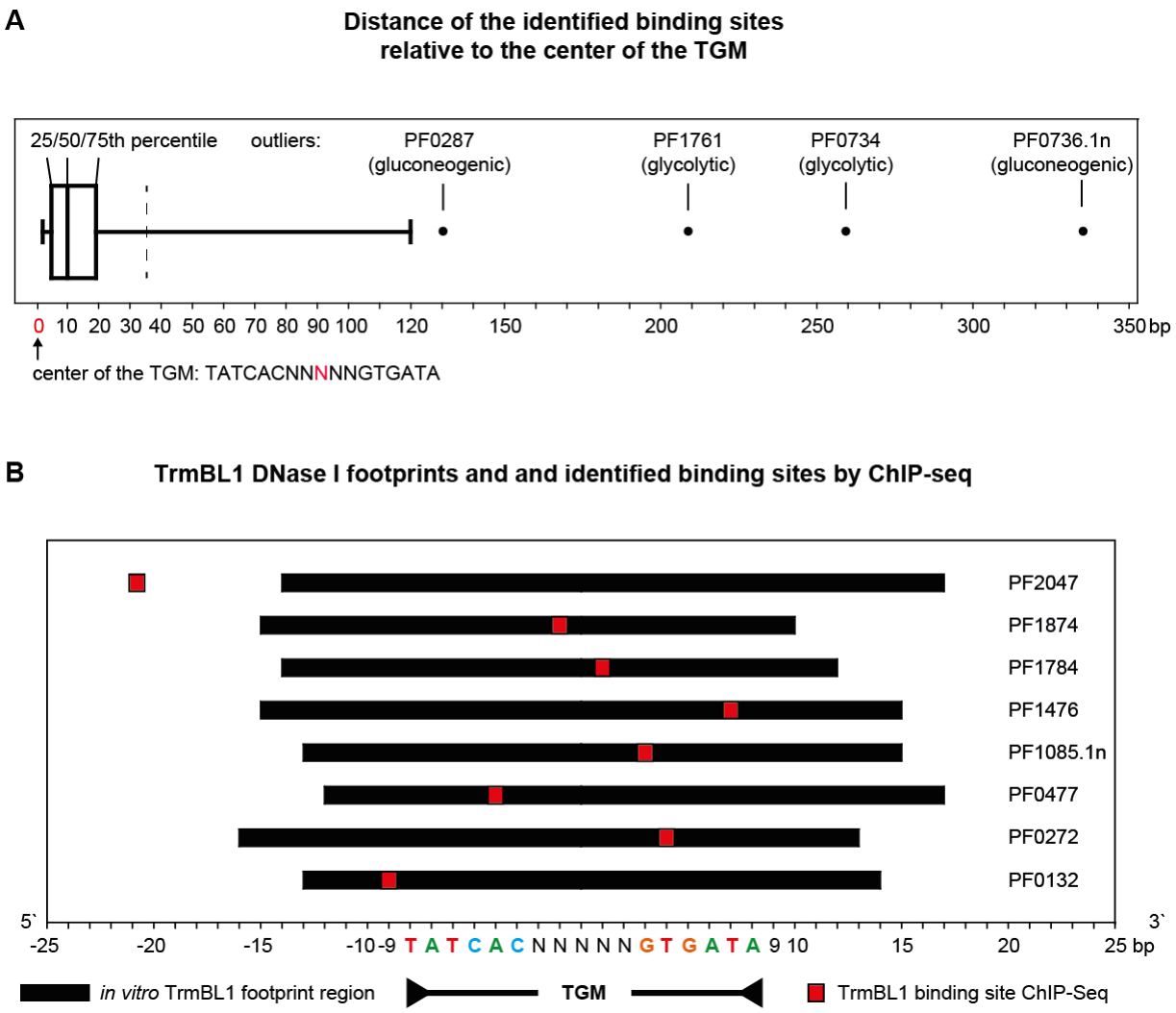


Figure 15. TrmBL1 ChIP-seq analysis displayed a high spatial resolution.

A, the distance of the identified TrmBL1 binding sites (consensus of two replicates of immunoprecipitation; n = 49) to the center of the TGM was calculated for the three tested conditions (gluconeogenic after 2 and 5 inoculations on pyruvate and glycolytic after two inoculations on starch) and is visualized by a boxplot (median = 9.5bp, box: 25th, 50th and 75th percentile; mean = 36bp, dashed line). For the outliers the binding sites and corresponding growth conditions are labeled. B, the protected regions from a TrmBL1 *in vitro* DNaseI footprint analysis are shown as black bars relative to the position of TGM of eight selected promoters. Footprint experiments were done by Antonia Gindner and will be presented in detail in her PhD thesis (in preparation). TGM containing promoters of the following genes were tested: PF0132, proposed *α-glucosidase*; PF0272, proposed *4-α-glucanotransferase*; PF0477, *α-amlyase*; PF1085.1n, hypothetical; PF1476, *padr-like regulator*; PF1784, *pfk*; PF1874, *gapdh*; and PF2047, *l-asparaginase*. The identified binding sites using ChIP-Seq (consensus of all samples derived from the three conditions) are labeled as red bars. Nucleotide positions were centered on the TGM and are represented from 5' to 3'. The TGM is depicted by coloured letters.

3.2.5. Function of TrmBL1 as global transcriptional regulator, which acts both as repressor and activator

To study the function of TrmBL1 as transcriptional regulator with dual functionality in more detail, *in vitro* transcription assays were carried out for 15 identified TrmBL1 binding sites (experiments were carried out by Antonia Gindner and will be presented in detail in her PhD thesis (in preparation)). First, binding of TrmBL1 to the used DNA templates was proved by EMSAs. These assays demonstrated specific TrmBL1 binding to all analyzed regions *in vitro*. Furthermore, *in vitro* transcription proved that binding of TrmBL1 downstream of the BRE and TATA-box leads to transcriptional repression. This includes seven known and predicted genes, but also four novel identified genes. In addition*, in vitro* transcription revealed that TrmBL1 inhibits transcription initiated from a promoter located upstream of the POR subunit delta gene (PF0967), which is predicted to be part of a polycistronic operon transcribed from PF0971 to PF0965. TrmBL1-mediated activation of transcription could be demonstrated for the *gapdh* gene (PF1874), where TrmBL1 binds upstream of the BRE and TATA-box. Additionally, for one novel identified gene, encoding an l-asparaginase (PF2047) and containing a TrmBL1 binding site upstream of the promoter, transcriptional enhancement was observed (Table 6).

Proving the effect of TrmBL1 binding to the identified binding sites by quantitative RT-qPCR analysis between gluconeogenic and glycolytic conditions *in vivo* was not possible. Using different protocols, all attempts failed to purify high quality RNA suitable for RT-qPCR analysis of *P. furiosus* cells grown under gluconeogenic conditions (pyruvate). This was presumably caused by the presence of high amounts of PCR inhibitors in the reactions (data not shown). However, comparison of the ChIP-seq data with the results derived from a microarray analysis of *P. furiosus* grown on carbohydrates or peptides verifies that TrmBL1 represents a global regulator in *P. furiosus in vivo*, which exhibits dual functionality (Table 6) (Schut et al., 2003). Transcriptional response with a reliable *P*-value could be found for 14 binding sites. Genes containing a TrmBL1 binding site downstream of the promoter are downregulated under gluconeogenic growth condition. These are especially genes encoding proteins involved in sugar uptake like the MD-system (PF1938-PF1933) and glycolysis like the PFK (PF1784). In contrast, the gluconeogenic-specific *gapdh* gene (PF1874) is upregulated under gluconeogenic growth conditions. TrmBL1 binds upstream of the BRE and TATA-box in the corresponding promoter regions. Additionally upregulation was observed for the *l-asparaginase* gene (PF2047). Only for one gene encoding a predicted extracellular α-amylase (PF0477) a discrepancy of the position of the binding site and the transcriptional response was found. Moreover, the comparison of the ChIP-seq and microarray analysis revealed that TrmBL1 may act for one single promoter region both as repressor and activator. On the one hand it inhibits reverse transcription of an operon, which encodes for a N-ethylammeline chlorohydrolase, two hypothetical proteins and the alpha-glucan phosphorylase (transcribed from PF1538 to PF1535). On the other hand it enhances forward transcription of the *dihydroorotate dehydrogenase 1b* gene (PF1539). A comparable transcriptional regulation mechanism was shown for the archaeal bifunctional regulator SurR, which controls hydrogen and elemental sulfur metabolism in *P. furiosus* (Lipscomb et al., 2009).

Table 6. GRN of TrmBL1 from *P. furiosus* based on ChIP-seq analysis (selected regions).

Gene	Product	Position TGM and TrmBL1 binding (ChIP-seq)	gly/glu	TrmBL1	
P. furiosus	Protein/ Enzyme	rel. to BRE/TATA	expression	EMSA binding	cell-free transcription
PF0132 (PF0133)	GLS hypothetical	downstream -	(0.71) 1.47	+ -	Repression -
PF0196	PGI	downstream	**2.26**	+	Repression
PF0272	AMY1	downstream	**4.7**	+	Repression
PF0287	pyrolysin	upstream	(-0.29)	+	no effect
PF0464	GAPOR	downstream	**2.54**	(+)	n.d.
PF0477	AMY2	downstream	**-2.45**	+	Repression
PF0478	AMY3	downstream[a]	(0.89)	n.d.	n.d.
PF0588	PGM	downstream	**0.69**	n.d.	n.d.
PF0967	POR	downstream	(-0.66)	+	Repression
PF1025	hypothetical	downstream	**2.47**	n.d.	n.d.
PF1085.1n	hypothetical	downstream	**-**	+	Repression
PF1109	SB-protein	downstream	**3.36**	+	Repression
PF1350	transporter	downstream	(-0.11)	+	Repression
PF1476	PadR regulator	downstream	(-0.34)	+	Repression
PF1538	N-ethylammeline chlorohydrolase	downstream	**3.20**	n.d	n.d.
PF1539	dihydroorotate dehydrogenase 1b	upstream	**-1.26**	n.d	n.d.
PF1784	PFK	downstream	**2.54**	+	Repression
PF1874	GAPDH	upstream	**-3.22**	+	Activation
PF1938	MD-system	downstream	**2.13**	(+)	Repression
PF1959	PGM	downstream	**1.75**	n.d	n.d.
PF2047	l-asparaginase	upstream	**-2.37**	+	Activation

Enzymes and transporters: GLS, proposed α-glucosidase; PGI, phospho-glucose isomerase; AMY1, proposed 4-α-glucanotransferase; GAPOR, GAP:Fd oxidoreductase; AMY2, proposed extracellular α-amylase; AMY3, proposed extracellular cyclomaltodextrin glucano-transferase; PGM, phosphoglycerate mutase; POR, pyruvate ferredoxin oxidoreductase; SB-protein, Starch binding protein (PF1109-1110); PFK, phospho-fructokinase; ; GAPDH, GAP dehydrogenase; MD-system, maltodextrin-specific ABC transporter (PF1938 to PF1933). Positions of the TGM and TrmBL1 binding sites (ChIP-seq) relative to the BRE and TATA-box are shown in supplemental table?. gly/glu, differential expression after growth of *P. furiosus* under glycolytic (gly) and gluconeogenic (glu) condition. Significant values are depicted in bold (Schut et al., 2003). EMSA, elctromobility shift assay; binding +, this study; binding (+), Lee et al. 2007a and 2008; n.d. not determined. EMSA and cell-free transcription results were obtained from Antonia Gindner. Transcriptional activation (blue) and repression (red) is depicted by colors. [a] Jorgensen et al., 1997.

3.2.6. Deletion of the 16 kb fragment encoding the TM system in *P. furiosus*

Mapping of the sequenced reads (TrmBL1 immunoprecipitations and inputs) from cells grown under gluconeogenic and glycolytic conditions to the *P. furiosus* genome revealed the deletion of a 16 kb fragment from gene PF1737 to PF1751 (Figure 16, chromosomal position 1,613,140 to 1,629,427). This 16 kb fragment contains the TM system (PF1739 to PF1744), which encodes for a trehalose/maltose-specific-ABC-transporter including TrmB (DiRuggiero et al., 2000; White et al., 2008). Southern blot analysis verified the deletion of this fragment after two inoculations (Figure 17, A). For the initial recultivation culture of *P. furiosus* type strain DSM3638 a weak signal for the region of interest could be detected. This indicates an underrepresented presence of this fragment in the genomes. To study the loss of this fragment in more detail copy number analysis was conducted after two and five inoculations in the appropriate medium (Figure 17, B). While under glycolytic conditions after two inoculations the fragment is approximately completely deleted (< 0.05), the copy number under gluconeogenic conditions varies between 0.15 after two inoculations and 0.20 after five inoculations. Even growth on maltose could not prevent deletion of this fragment.

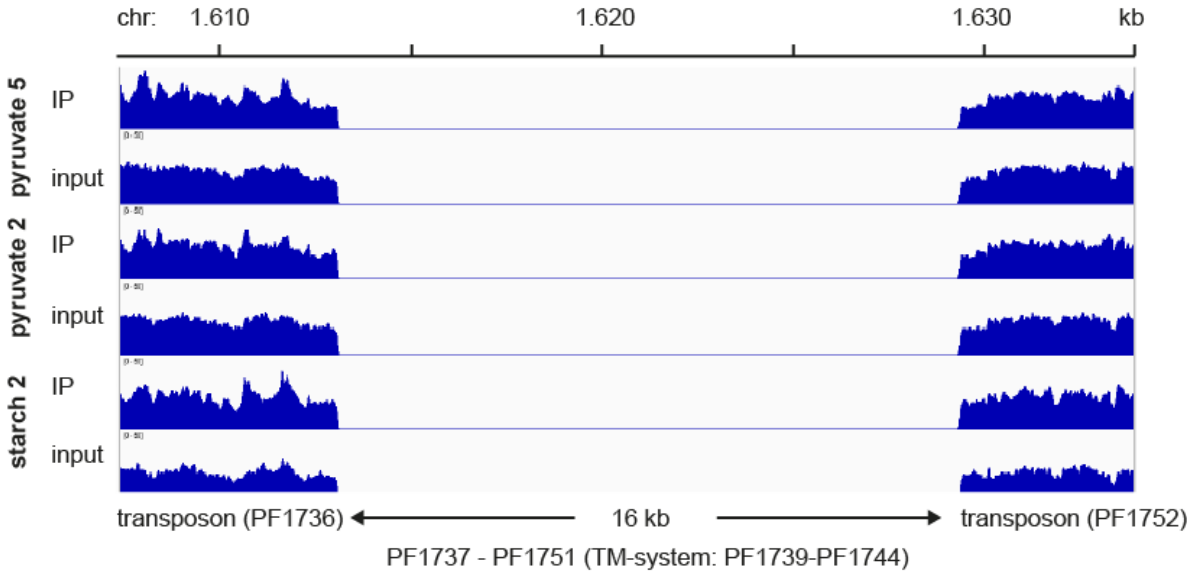


Figure 16. Deletion of a 16 kb fragment harbouring the TM-system including TrmB.

TrmBL1 ChIP-seq experiments with *P. furiosus* cells grown under gluconeogenic (after 2 and 5 inoculations on pyruvate) and glycolytic (after 2 inoculations on starch) conditions were carried out. Mapped TrmBL1 immunoprecipitation (IP) and input reads were visualized using the IGV genome browser (https://www.broadinstitute.org). Almost no mappable reads were detected for a 16 kb fragment. This comprises the genes from PF1737 tp PF1751, which includes the TM-system (PF1739-PF1744). PF1736 and PF1752 encode transposons. Data range is [1-50] for all samples.

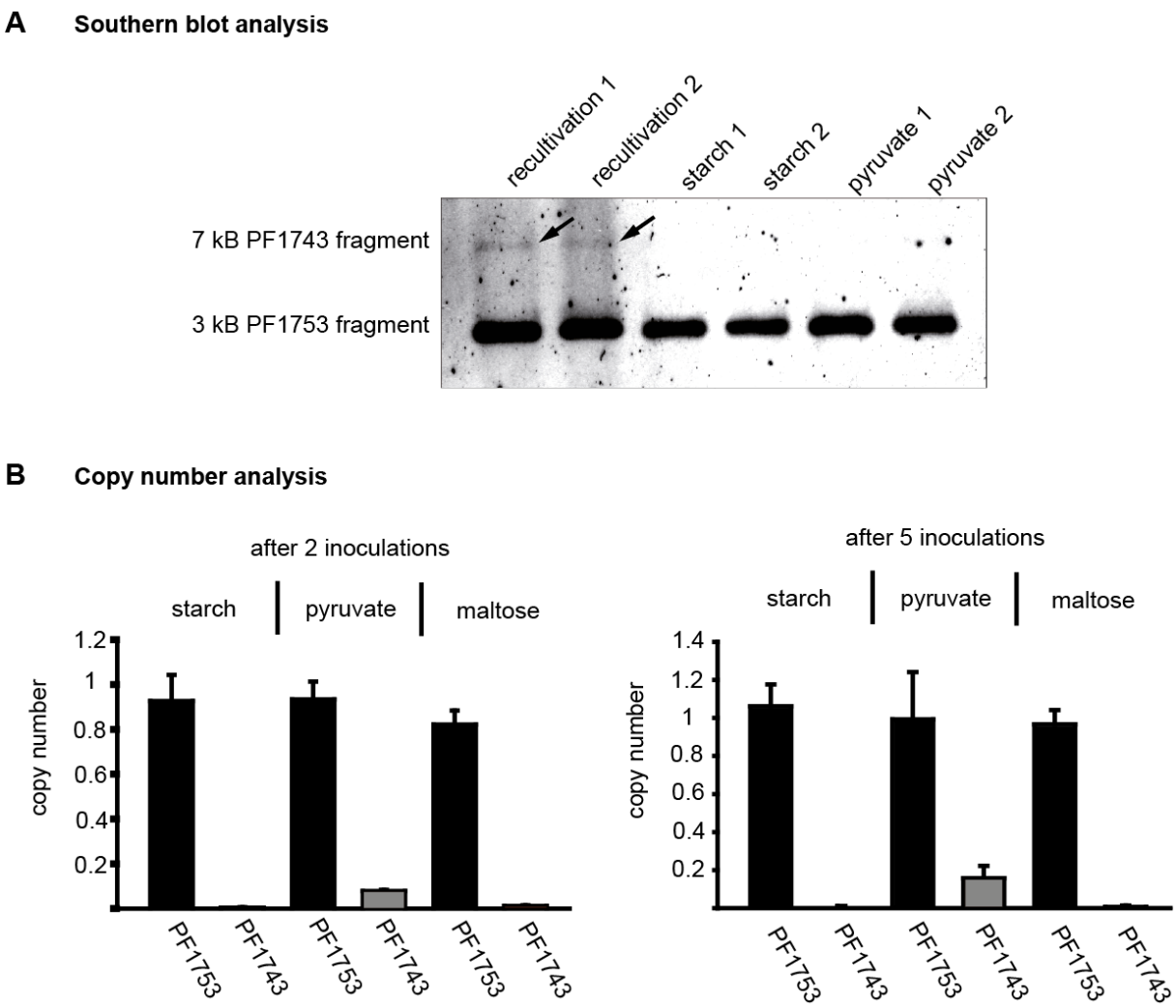


Figure 17 Validation of the deletion by southern blot and copy number analysis.

A, southern blot analysis using BamHI and SmaI digested *P. furiosus* genomic DNA. DNA was isolated from cells after recultivation and adaptation to glycolytic (starch) or gluconeogenic (pyruvate) growth conditions. The result of two independent recultivations is shown. The probe, specific for PF1753, detects a 3kb fragment, while the PF1743 specific probe binds to a 7 kb fragment. B, copy number analysis of the genes PF1743 and PF1753 in *P. furiosus* cells grown under the following conditions: glycolytic (starch and maltose) or gluconeogenic (pyruvate). The gene PF1784 was used as calibrator and the recultivation culture as reference. The mean of 3 biological replicates including SD is shown.

3.2.7. Short conclusion

ChIP-seq analysis of TrmB and TrmBL1 from *P. furiosus* lead to two major findings: A, in the used *P. furiosus* strain `DSM3638´ a 16 kb fragment harboring the TM-system including TrmB was deleted. Thus, function of TrmB in regulation of sugar metabolism could not be studied using this strain; B, In contrast, using ChIP-seq for mapping of TrmBL1 binding sites in a genome wide manner for *P. furiosus in vivo* revealed an extended function of TrmBL1 as global regulator of metabolism. TrmBL1 not only regulates genes involved in sugar metabolism, but it also controls transcription of genes involved in various additional metabolic pathways and biological processes as proteolysis or the AA metabolism.

3.3. TrmBL2 is an abundant chromosomal DNA binding protein

Analysis of the distribution of TrmB/TrmBL proteins within the *Thermococcales* confirmed that TrmBL2 is evolutionarily conserved in the genomes of all 13 sequenced species. Like TrmB and TrmBL1, TrmBL2 contains in addition to the TrmB_domain the Regulator_TrmB domain, but it lacks the sugar binding part (Lee et al., 2008). *In vitro* TrmBL2 is able to bind TGM-containing promoters like the MD-promoter, but not solely. Moreover, repression or activation of these promoters by TrmBL2 could not be shown in cell-free transcription assays. The TrmBL2 orthologue, TK0471, from *T. kodakarensis* was suggested to function as a novel type of abundant chromosomal binding protein, which can form thick fibrous structures with DNA (Maruyama et al., 2011). Furthermore, a whole genome microarray analysis of a TrmBL2 deletion strain compared with the wild-type strain revealed that TrmBL2 mainly acts as global transcriptional repressor. After successful adaptation of the ChIP-seq protocol for TrmBL1, this approach was extended to analysis of TrmBL2. The aim was to study, whether the *P. furiosus* TrmBL2 acts as an abundant chromosomal binding protein according to TK0471. Moreover, identified binding sites should be studied in more detail to confirm that TrmBL2 can function as global repressor of transcription in *P. furiosus*.

3.3.1. Specificity of the anti-TrmBL2 antibodies

Potential cross-reactivity of the antibody raised against TrmBL2 (anti-TrmBL2 IgG) with paralogous TrmB/TrmBL proteins in *P. furiosus* was tested by western blot analysis. The TrmBL2 specific antibody showed no cross-reactivity with TrmB and TrmBL1 (Figure 18, lane 1 and 2). Moreover, a specific reaction was observed for recombinant TrmBL2 (Figure 18, lane 3). In crude extracts derived from cells grown under glycolytic (starch) or gluconeogenic (pyruvate) conditions one specific signal at ~31 kDa (30,607 calculated molecular weight of TrmBL2) was detected (Figure 18, lane 4 and 5). Thus, the anti-TrmBL1 IgG could be used for the intended ChIP-seq experiments due to its high specificity.

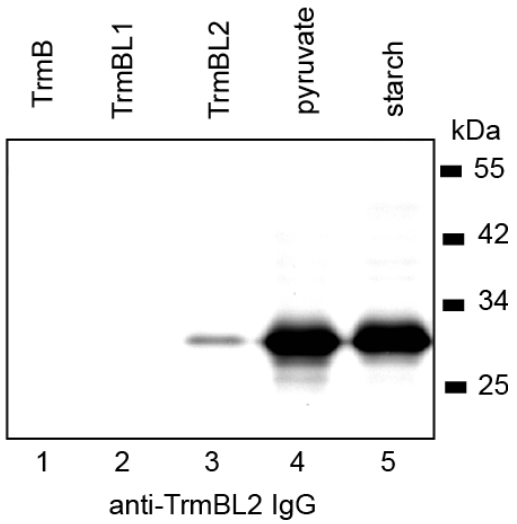


Figure 18. Western blot analysis using the anti-TrmBL2 IgG.

100 ng each recombinant TrmB (lane 1), TrmBL1 (lane 2) and TrmBL2 (lane 3) were used and 20 µg crude extract obtained from *P. furiosus* cells which were grown under gluconeogenic (pyruvate, lane 4) or glycolytic (starch, lane 5) conditions. Purified antibodies raised against recombinant TrmBL2 (anti-TrmBL1 IgG, 1:2000) were used for detection.

3.3.2. Identification of TrmBL2 binding sites in a genome-wide manner using ChIP-seq

TrmBL2 ChIP-seq experiments were carried out according to TrmBL1 experiments using cells grown under gluconeogenic or glycolytic conditions (two or five inoculations on pyruvate: pyruvate 2 and 5; or two inoculations on starch: starch 2). In total 263 TrmBL2 binding events were detected under both conditions (Figure 19, A and B). The two gluconeogenic samples overlap for the most part, but the number of called peaks is slightly lower after five (n = 227) than after two (n = 244) inoculations on pyruvate. In contrast to TrmBL1, binding of TrmBL2 is not dependent on the growth condition (overlap TrmBL2 binding sites glycolytic and gluconeogenic: n = 237). Only eight binding sites are specific for gluconeogenic growth conditions (3.4 % of all gluconeogenic sites) and 18 binding events are exclusively found under glycolytic conditions (7.0 % of all glycolytic sites). The corresponding genes are not differentially expressed in a whole genome microarray analysis using *P. furiosus* cells grown on maltose or peptides (Schut et al., 2003). Furthermore, the correlation of ChIP-enrichment between the glycolytic and gluconeogenic samples was as high as between the two gluconeogenic samples (pyruvate 2 versus starch 2: $\rho = 0.80$, $P = 8.9*10^{-47}$, n = 225; pyruvate 5 versus starch 2: $\rho = 0.83$, $P = 3.4*10^{-52}$, n = 225; and pyruvate 2 versus pyruvate 5: $\rho = 0.87$, $P = 5.0*10^{-67}$, n = 225; Spearman`s rank correlation). The small difference found between both conditions is also illustrated by visualization of the TrmBL2 ChIP-enriched regions (Figure 19, A). The peaks are distributed over the whole genome and the highest enrichment was found in the intergenic region upstream of the *pyrolysin* gene (PF0287). Moreover, the deletion of a 16 kb fragment encoding the TM system was also observed in the TrmBL2 samples. In addition, an aberration in the genomic coverage of mapped reads was detected under glycolytic growth conditions according to the TrmBL1 experiments. During his diploma thesis Stefan Eisenschink could already show that TrmBL2 binds to its own promoter region *in vivo* (Eisenschink, 2010). This indicated TrmBL2-mediated autoregulation of its own gene expression. In ChIP-seq experiments binding of TrmBL2 to its own promoter region in cells grown on pyruvate or starch could be confirmed (Figure 19, C). Moreover, the TrmBL2 binding site overlaps both the transcription and translation start site. This suggests a negative autoregulation mechanism, which would lead to less expression of TrmBL2 in the presence of high amounts of the TF.

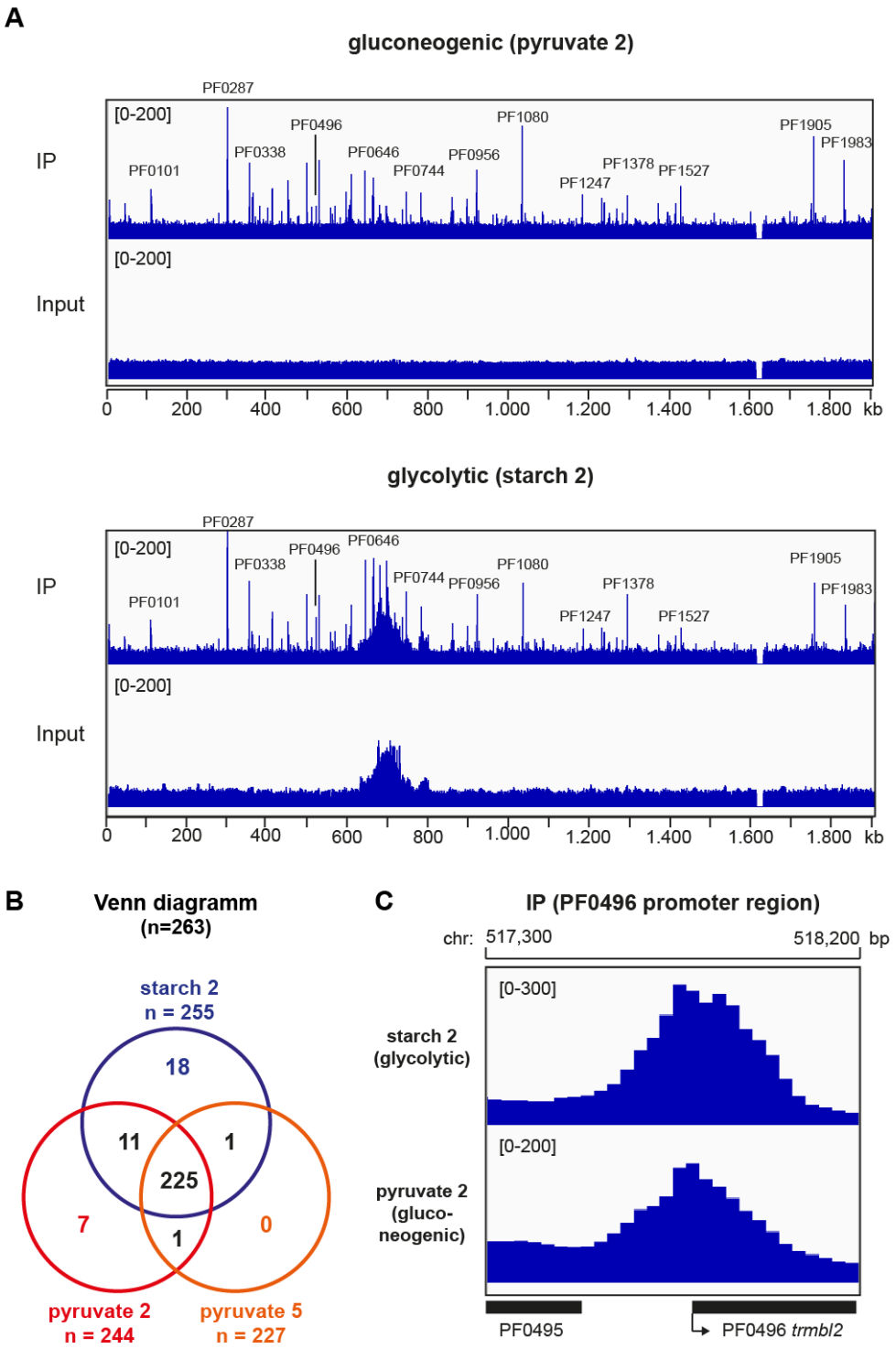


Figure 19. TrmBL2 is an abundant chromosomal DNA binding protein under gluconeogenic as well as glycolytic growth conditions.

A, TrmBL2 ChIP-seq experiments with *P. furiosus* cells grown under gluconeogenic (after 2 inoculations on pyruvate) and glycolytic (after 2 inoculations on starch) conditions were prepared. Mapped TrmBL1 immunoprecipitation (IP) and input reads were visualized. Prominent peaks found under both conditions are designated. B, peak calling was performed using the Pique software package. Peaks were verified by the genetrack software package. The numbers of overlapping and growth condition-specific enriched regions are presented in a Venn-diagram. C, detailed view of the TrmBL2 (PF0496) promoter region in the TrmBL2 ChIP-seq samples revealed enrichment in the range of the transcription and translation start site. The bent arrow indicates direction of transcription.

3.3.3. Co-occupancy of TrmBL1 and TrmBL2 in the genome of *P. furiosus*

In EMSA experiments TrmB, TrmBL1 as well as TrmBL2 could bind to the promoter region of the MD-system (Lee et al., 2007b). Thus, it was speculated that at some promoters differential regulation by all three TFs may take place *in vivo* (Lee et al., 2008). However, TrmBL2 was not able to recognize all TGM-containing promoters regulated by TrmBL1. To verify co-occupancy in specific promoter regions *in vivo*, ChIP-seq results obtained for TrmBL1 and TrmBL2 under glycolytic and gluconeogenic growth conditions were compared. Visualization of enriched regions identified under gluconeogenic conditions illustrates the major differences, especially regarding the amount of peaks, between both datasets (Figure 20, A). First, windows spanning the TrmBL1 binding sites by 500 bp upstream and downstream were scanned for identified TrmBL2 binding events. This preliminary analysis revealed seven possible regions, where co-occupancy of both proteins could appear *in vivo* (Table 8). While for these sites TrmBL1 binding can only be observed under one specific condition, binding of TrmBL2 is independent of glycolytic or gluconeogenic growth.

Table 7. Genomic regions potentially co-occupied by TrmBL1 as well as TrmBL2

	TrmBL1		TrmBL2		
gene	**gluconeogenic**	**glycolytic**	**gluconeogenic**	**glycolytic**	**distance [bp]**
PF0286/87	297340	-	297243	297218	97
PF0647/48	**661496**	-	**661499**	661497	**3**
PF0736.1n	**732475**	-	**732493**	732498	**19**
PF1025	**980987**	-	**980947**	980952	**40**
PF1761	-	1638683	1638556	1638527	157
PF1784/85	1660103	-	1660248	1660231	146
PF1997	-	1845188	1845055	1845051	137

Bold letters indicates presumably real co-occupied sites.

Analysis of the distances between the binding sites in more detail showed that four potentially co-occupied sites are separated by more than approximately 100 bp. The spatial resolution of the ChIP-seq approach is at least higher than 50 bp. Thus, these regions were classified as independent binding events, which likely cause differential effects in gene expression. In the non-coding region between PF0286 and PF0287 the TrmBL2 site is located in the termination region of PF0286, which encodes a hypothetical endonuclease, and a tRNA gene. In contrast, TrmBL1 recognizes the TGM, which is located upstream of the promoter elements of the *pyrolysin* gene (PF0287). In the PF1784/PF1785 intergenic region the TrmBL1 peak overlaps the translation and transcription start site of the *pfk* gene (PF1784). TrmBL2 binds close to the translation and transcription start site of the *pseudouridine synthase* gene (PF1785) (Figure 20, B). In the two cases, where co-occupancy is observed under glycolytic conditions, the TrmBL2 binding regions overlap the promoter regions, whereas TrmBL1 is located in the coding regions of the corresponding genes.

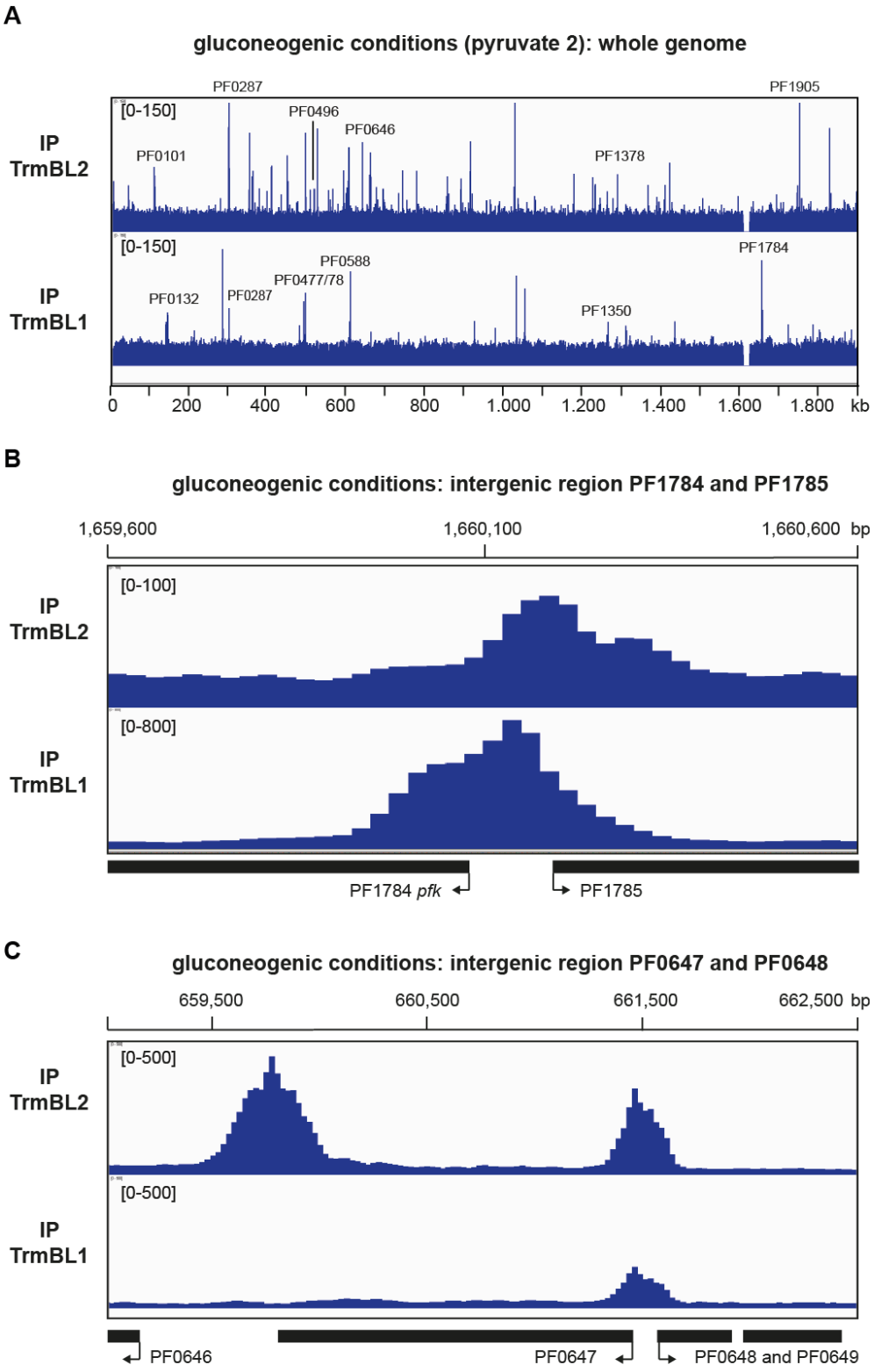


Figure 20. TrmBL1 and TrmBL2 co-occupy only a few regions in the *P. furiosus* genome.

TrmBL1 and TrmBL2 ChIP-seq results obtained from cells grown on pyruvate (gluconeogenic) were visualized. Bent arrows indicate direction of transcription. A, TrmBL1 binds to specific regions in the whole genome, whereas TrmBL2 binding sites are abundantly distributed. B, in the PF1784/1785 intergenic region TrmBL1 overlap the PF1784 TSS and TrmBL2 the PF1785 TSS. C, both proteins bind the intergenic region between PF0647 and PF0648. For TrmBL2 a second enriched region was detected in the PF0647 termination region. PF0648 is cotranscribed with PF0649, which encodes a TrmB_domain containing TF.

The remaining three promoter regions can represent sites, where both proteins bind in parallel or interchangeably *in vivo*. In these genomic regions no differentiation between the TrmBL1 and TrmBL2 peak is feasible (Figure 20, C).

3.3.4. TrmBL2 binds coding as well as non-coding regions of the genome

Maruyama et al. (2011) analyzed the genome-wide distribution of the TrmBL2 orthologue TK0471 from *T. kodakarensis* using sucrose gradient centrifugation combined with high-throughput sequencing. This demonstrated abundant binding of TK0471 to the genome. The identified distribution of TK0471 binding to non-coding (9.4 %) and coding (90.6 %) regions did not differ from the predicted distribution of non-coding (~8 %) and coding (~92 %) regions in the *T. kodakarensis* genome. The ChIP-seq analysis revealed that TrmBL2 also binds to coding (annotated ORFs and RNA genes) as well as non-coding (intergenic regions) regions of the *P. furiosus* genome (Figure 21 and Figure 22, A and B).

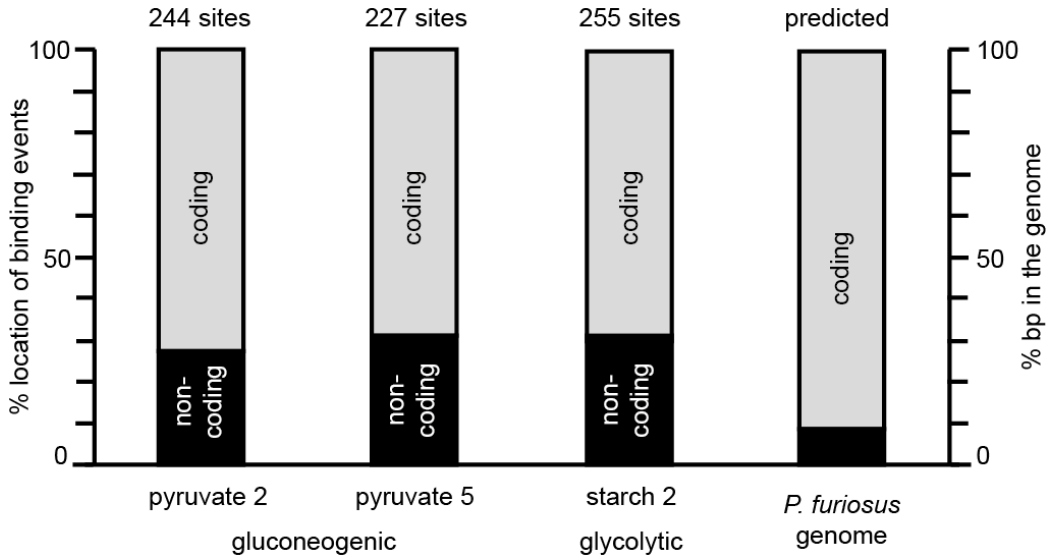


Figure 21. TrmBL2 binding sites are enriched in non-coding regions of the genome under gluconeogenic and glycolytic conditions.

TrmBL2 ChIP-enriched regions are located in coding (annotated ORFs and RNA genes) and non-coding regions (intergenic regions). Distribution of the binding events is presented as % of the detected sites located in coding (grey bar) or non-coding (black bar) regions. The predicted distribution of coding to non-coding regions of the *P. furiosus* genome is shown on the right side.

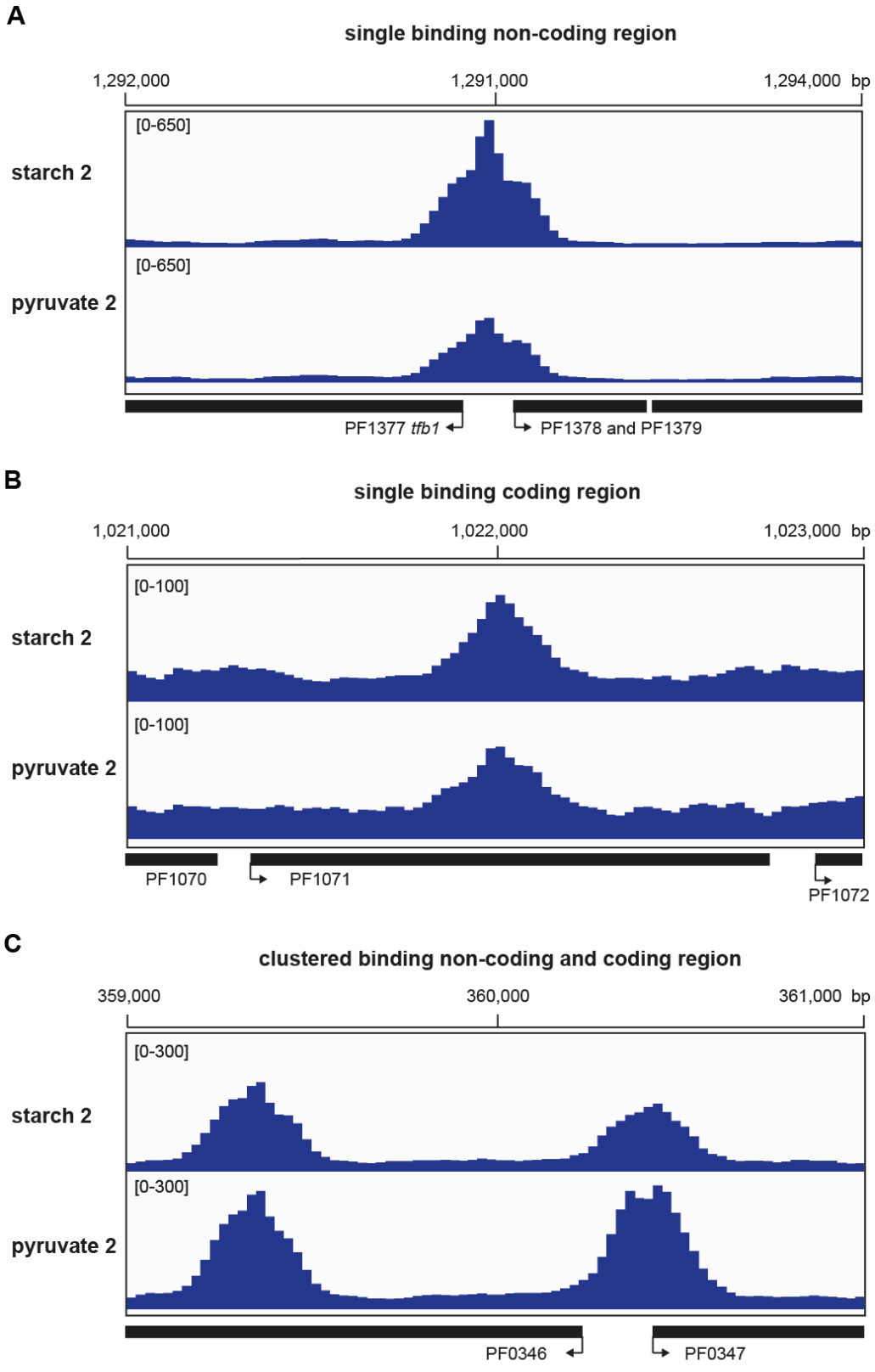


Figure 22. TrmBL2 binds to coding and intergenic regions in the *P. furiosus* genome.

TrmBL2 ChIP-seq results obtained from cells grown on pyruvate or starch were visualized. Bent arrows indicate direction of transcription. A, TrmBL2 binds the intergenic region between the genes PF1377 and PF1378. PF1377 encodes the GTF TFB1. B, for the PF1071 gene the TrmBL2 binding site is located in the coding region. C, TrmBL2 peaks can cluster within 1000 bp. For the PF0346 gene TrmBL2 binds on the one hand in the intergenic region and on the other hand in the coding region.

In contrast to TK0471, the detected TrmBL2 distribution differ from the predicted distribution of ~8 % non-coding and ~92 % coding regions in the *P. furiosus* genome. 27 % to 30 % of binding events detected under the particular condition are located in intergenic areas (Figure 21). These distributions of peaks in the ChIP-seq experiments are significantly enriched regarding TrmBL2 binding to intergenic regions compared to random distribution: ChIP-seq pyruvate 2 versus predicted random distribution: $P = 0.001$, n = 244; ChIP-seq pyruvate 5 versus predicted random distribution: $P = 0.0001$, n = 227; ChIP-seq starch 2 versus predicted random distribution: $P = 0.00005$, n = 255 (Fisher's exact test). Within the ChIP-seq samples no significant differences were observed. This indicates that TrmBL2 has a higher affinity to non-coding intergenic regions compared to coding regions and it is not as randomly distributed as expected. Moreover, in some regions TrmBL2 peaks cluster within 1000 bp (Figure 22, C). For approximately 39 % of the binding events a second binding site is located within 1000 bp upstream or downstream in the genome, whereas 61 % appear as single binding sites. The average distance of peaks within clustering regions was 579 bp (pyruvate 2: 543 bp +/-286; pyruvate 5: 624 bp +/- 265; starch 2: 568 bp +/- 289; mean with SD). This observation additionally encourages the assumption that TrmBL2 does not bind to the genome in random manner. However, this would require that TrmBL2 recognizes a cis regulatory element in the DNA according to TrmBL1 and the TGM. So far, all attempts for *de novo* discovery of a TrmBL2 binding motif were not successful. The used programs were not able to find a motif (MEME) or the identified motifs were only present in less than half of the analyzed genomic regions (MEME-ChIP or Gibbs sampler, etc.; Favorov et al., 2005; Bailey et al., 2009). If a motif was present in approximately all regions, these motifs occurred at more than 1000 to 5000 positions of the genome (MAST, FIMO; Bailey et al., 2009). This extremely large number of sites would disagree with the assumption that TrmBL2 binds specifically to the genome. Furthermore, motifs identified by the different programs differed regarding their length and consensus sequence. Also analysis of subsets of the whole dataset did not improve the results (e.g. only intergenic regions, only coding regions, glycolytic-specific or gluconeogenic-specific peaks). Thus, the mechanism how TrmBL2 binds to specific regions of the genome remains unclear and appears to be independent of a prominent DNA binding motif like the TGM. The only common feature of most *de novo* identified motifs was that they represent AT-rich regions.

.

3.3.5. The function of TrmBL2 as global repressor remains unclear

Besides representing an architectural chromosomal binding protein, TrmBL2 (TK0471) is assumed to function as global repressor during the logarithmic growth phase (Maruyama et al., 2011). Moreover, activity as transcriptional regulator correlated with presence of the protein in the promoter region. During her PhD thesis Ingrid Waege constructed a *P. furiosus* deletion strain of TrmBL2 (ΔTrmBL2) using homologous recombination (Ingrid Waege 2014). This showed that TrmBL2 is not essential for growth of cells under all tested conditions. In addition, whole genome microarray analysis of the deletion strain compared to the wildtype strain was used to analyze differential gene expression. The identified differences of mRNA-levels between both strains were much lower compared to the results obtained from TK0471/TrmBL2 deletion strain analysis from *T. kodakarensis* (Ingrid Waege 2014; Maruyama et al., 2011). In the *P. furiosus* deletion strain the signal intensities of only seven ORFs on the genome were elevated more than twofold (log-fold change) compared to those in the wild-type strain. Furthermore, no ORF was decreased more than twofold. Nevertheless, these data were used to analyze the function of *P. furiosus* TrmBL2 as global regulator in comparison with the ChIP-seq results. Focusing on the 51 ORFs, which showed the highest differential expression, encouraged the assumption that TrmBL2 functions as global repressor. 80 % of the genes were upregulated in the deletion strain compared to the wildtype strain, whereas only 20 % were downregulated. These 51 genes represent 35 single transcription units (72 single-transcribed genes or co-transcribed in an operon). 30 of them were used for a comparison of the gene expression results with TrmBL2 binding identified by ChIP-seq (Table 15). TrmBL2 binds independent from glycolytic or gluconeogenic growth to the intergenic regions upstream of 23 transcription units (77 %). For one differential expressed gene it is located in the coding region, whereas for the remaining six transcription units no TrmBL2 binding could be detected. In Table 8 selected genes are shown, which were down- or upregulated in the ΔTrmBL2 strain compared with the wildtype and which contained a TrmBL2 binding site in the intergenic region. Most identified genes encode so far hypothetical proteins with unknown function. Thus, the physiological role of TrmBL2 remained unclear. However, database search for homology detection and protein domains (HHpred and Pfam; Soding et al., 2005; Finn et al., 2006) revealed that numerous of these hypothetical proteins contain protein structures, which can be linked to membrane-associated proteins and transporters. These genes are mainly repressed in the presence of TrmBL2. In contrast, the functions of genes, which are upregulated in the presence of TrmBL2, were already annotated. This includes an aspartate kinase (PF1052), which functions in the AA metabolism. Moreover, deletion of TrmBL2 leads to downregulation of the hypothetical protein PF0648, which is co-transcribed with a TF containing the TrmB_domain (TrmBL9 cluster). The most unexpected result represents activation of expression of the flagellin-operon (*fla*-operon; PF338-PF0332; reverse transcribed) in the presence of TrmBL2. This suggests that the ΔTrmBL2 strain has less flagella and it is deficient in swimming and building cell-cell as well as cell-surface contacts. In contrast to TrmBL1, transcriptional activation or repression in the presence of TrmBL2 could not be linked to binding upstream or downstream of the BRE and TATA-box. TrmBL2 is located upstream in the promoter regions of all activated genes, but it was bound both downstream and upstream in the promoter regions of repressed genes. Thus, the mechanism of TrmBL2-mediated transcriptional control appears to be different from the mechanisms described so far in the Archaea.

Table 8. Selected genes down- or upregulated in the presence of TrmBL2.

transcript organization [1]	gene	protein	microarray (ΔTrmBL2/WT) (Ingrid Waege 2014)	TrmBL2 binding site ChIP-seq
gene expression repressed in the presence of TrmBL2				
singleton	PF0889	hypothetical protein (membrane protein)	upregulation	intergenic (-6)
singleton	PF0580	hypothetical protein (filamin domain)	upregulation	intergenic (-88)
operon	PF0570	hypothetical protein (transport)	upregulation	-
	PF0571	hypothetical protein (membrane protein)	upregulation	intergenic (-12)
operon	PF0921	ABC transporter	upregulation	-
	PF0922	hypothetical protein (membrane protein)	upregulation	-
	PF0923	hypothetical protein (dehydratase)	upregulation	-
	PF0924	hypothetical protein (PAS domain)	-	-
	PF0925	heme biosynthesis protein	-	-
	PF0926	hypothetical protein (coiled-coil domain)	-	intergenic (-44)
gene expression activated in the presence of TrmBL2				
singleton	PF1052	aspartate kinase	downregulation	intergenic (-168)
operon	PF0648	hypothetical protein	downregulation	intergenic (-83)
	PF0649	hypothetical protein (TrmB_domain)	-	-
operon	PF0332	flagellar protein FlaHc	downregulation	-
	PF0334	archaeal flagellar protein FlaF	downregulation	-
	PF0335	flagella-related protein D	downregulation	-
	PF0337	flagellin	downregulation	-
	PF0338	flagellin	downregulation	intergenic (-130)

ΔTrmBL2, TrmBL2 deletion strain; WT, wildtype strain DSM3638; protein, results from database search are shown in parenthesis; binding site, positions of TrmBL2 binding sites (consensus of all experiments) relative to translation start site are shown in parenthesis; [1], http://archaea.ucsc.edu.

However, all these results are based on rather marginal differential expression patterns derived from a TrmBL2 gene disruption whole-genome microarray analysis. Moreover, the 23 up- or downregulated transcription units, which contained a TrmBL2 binding site in their intergenic region, represent only 20 % of all indentified TrmBL2 binding events located in non-coding regions of the *P. furiosus* genome. The reason why the other 80 % did not cause transcriptional repression or activation of the corresponding genes and operons remained unclear.

3.3.6. ChIP-seq revealed an unexpected role of TrmBL2 in the heat shock response of *P. furiosus*

Liu et al. (2007) showed by ChIP that *in vivo* the global repressor Phr dissociates from the DNA upon heat shock (growth temperature higher than 105°C), whereas TBP remains bound to the promoter. Therefore, it was an interesting task whether the ChIP-seq approach will also work using *P. furiosus* cells, which were formaldehyde treated at 107°C. Before crosslinking, cells reaching the middle- late log-phase were incubated for 30 min at 107°C. The anti-TrmBL2 antibody was choose for immunoprecipitation due to its high specificity and the large amount of expected ChIP-enriched regions. When the experiments started, no link existed between the function of TrmBL2 and the response of *P. furiosus* to heat shock. Thus, it was estimated that no striking differences regarding TrmBL2 genome occupancy should be observed between optimal growth conditions (95°C) and heat shock. The obtained result from the ChIP-seq analysis is visualized in Figure 23.

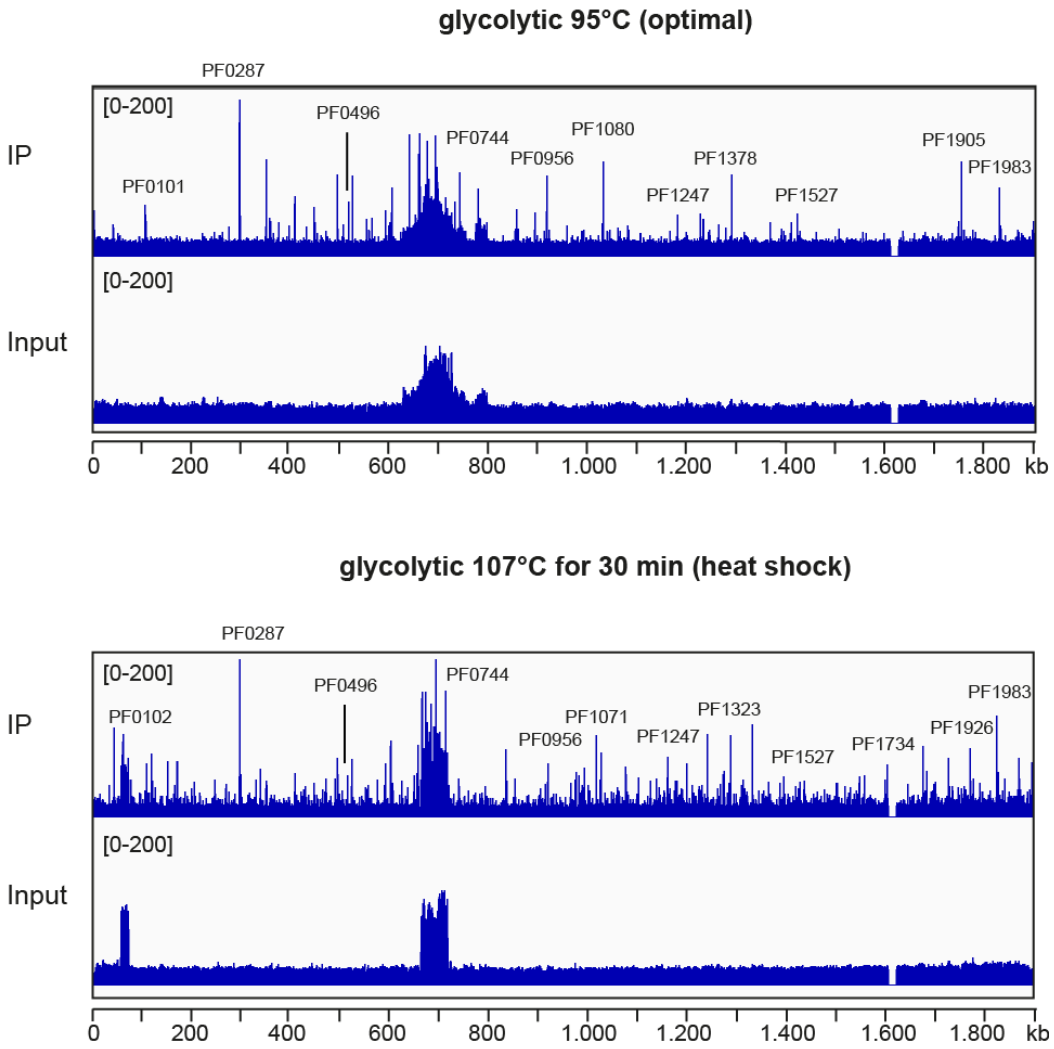


Figure 23. The number of detected TrmBL2 binding sites increased upon heatshock.

For ChIP analysis *P. furiosus* cells were treated with formaldehyde at 95°C (optimal growth condition) or after incubation at 107°C for 30 min (heat shock). The results from the TrmBL2 IP and input samples are shown. Prominent peaks are denoted.

It is obvious that the number of TrmBL2 binding sites is tremendously increased after heat shock. In total, 803 TrmBL2 binding events were identified by ChIP-seq after heat treatment compared to 255 after optimal growth (Figure 24, A). Moreover, some enriched regions can be found under both conditions (n = 175), whereas other can only be detected under one specific condition (optimal: n = 80; heat shock: n = 628). The aberration in the genomic coverage of mapped reads, which specifically appeared in the previous experiments only under glycolytic growth conditions, is also present after heat treatment of cells grown starch. However, the spanning area is more defined (chromosomal position 666,025 to 720,350). Additionally, a second aberration could be observed covering the genome from position 57,050 to 73,000 (Figure 23).

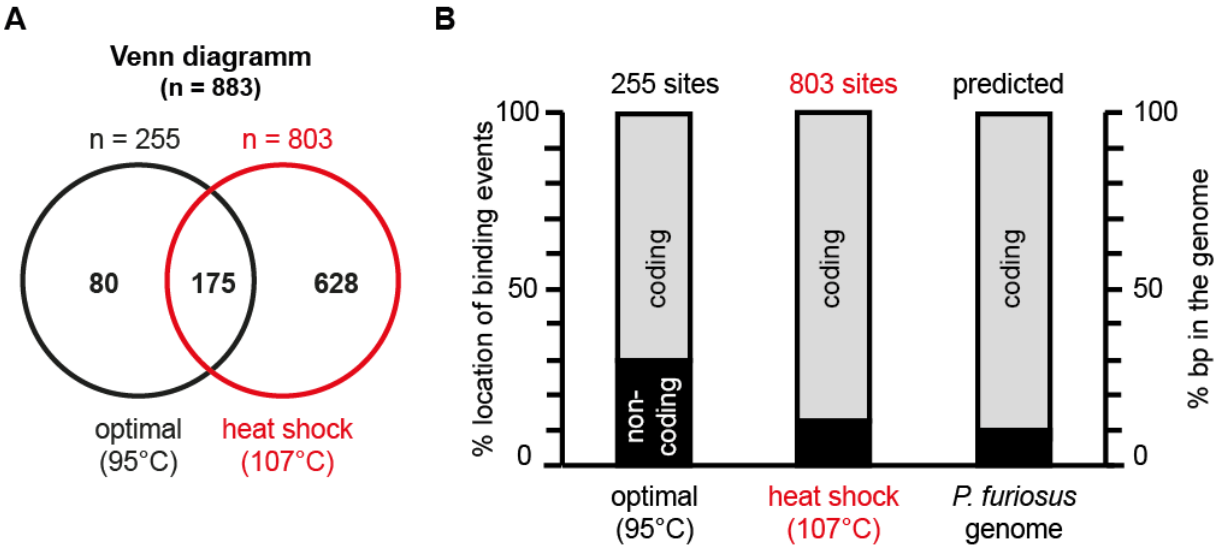


Figure 24. The number and distribution of detected TrmBL2 peaks is altered upon heat shock.

A, peak calling was performed using the Pique software package and verified by the genetrack software package. The numbers of overlapping and growth condition-specific (optimal: 95°C or heat shock: 107°C) enriched regions are presented in a Venn-diagram. B, TrmBL2 ChIP-enriched regions are located in coding (annotated ORFs and RNA genes) and non-coding regions (intergenic regions). Distribution of the binding events is presented as % of the detected sites located in coding (grey bar) or non-coding (black bar) regions. The calculated distribution of coding to non-coding regions of the *P. furiosus* genome is shown on the right side.

Analysis of the genome-wide TrmBL2 binding properties between optimal and heat shock conditions revealed mainly differential characteristics. Under heat shock conditions, TrmBL2 binds not only an increased number of genomic regions, but it also binds the overlapping enriched regions with differing intensities. By comparing the detected ChIP-enrichment between both samples, only a weak correlation was observed. ($\rho = 0.25$, $P = 0.001$, n = 175; Spearman`s rank correlation). Moreover, TrmBL2 binds non-coding (12 %) as well as coding regions (88 %) (Figure 24, B); however in contrast to optimal conditions the distribution is not biased toward the coding or intergenic parts of the genome compared to predicted random distribution of peaks along the genome: $P = 0.48$ (Fisher's exact test). Thus, after heat shock TrmBL2 appears to be genome-wide allocated in a random manner. In addition, for approximately 67 % of the binding events a second site is located within 1000 bp upstream or downstream in the genome after heat treatment, whereas 33 % appears as single binding site. The average distance of peaks within clustering regions was 565 bp +/-275 (mean with

SD). The number of clustering TrmBL2 binding sites is increased (*P* = 0.002; Fisher's exact test), whereas the determined average distance between clustered peaks is consistent with the values obtained under optimal conditions.

Figure 25 shows representative examples from the numerous differential binding patterns of TrmBL2, which were identified by ChIP-seq analysis between optimal and heat shock conditions. TrmBL2 binds to its own promoter *in vivo*. Thus autoregulation is suggested. The detected TrmBL2 binding site overlapped the transcription as well as translation start site. After heat shock the TrmBL2 peak is shifted upstream within the intergenic region (Figure 25, A). This indicates increased accessibility of the TrmBL2 TSS for the RNAP, which would lead to activation of transcription. The TrmBL2 deletion strain analysis of Ingrid Waege revealed mostly transcriptional repression caused by TrmBL2; however few transcription units were also upregulated. The highest activated ORFs were the genes PF0332 to 0338, which comprise the *fla*-operon. In the promoter region of the *fla*-operon TrmBL2 was located upstream of the BRE and TATA-box. This peak is also present in the heat shock samples, but the peak intensity is lowered (Figure 25, B). Moreover, an additional third clustering peak could be detected in the transcribed region. This suggests less transcriptional activity of the *fla* operon upon heat shock. Under optimal conditions a single TrmBL2 binding site was detected in the intergenic region of the *tfb1* gene (PF1377). This enriched region is preserved under both conditions. In contrast, the second *tfb* gene (PF0687; TFB2) in *P. furiosus* is differentially bound by TrmBL2 (Figure 25, C). Under heat shock conditions the intensity of TrmBL2 in its non-coding region is remarkably enlarged and the enriched region consists of two peaks. Additionally, the binding pattern in the transcribed region (PF0687 to PF0685) is altered. It is supposed that TFB2 may function as alternative GTF upon heat shock due to its increased expression under this condition (Shockley et al., 2003). To what extend this upregulation is affected by TrmBL2 has to be proven yet.

Analysis of genome-wide TrmBL2 binding after heat shock in *P. furiosus* showed that ChIP seq is also applicable at growth temperatures up to 107°C. The identified differential binding patterns under optimal and heat shock conditions uncovered an unexpected role of TrmBL2 as abundant chromosomal binding protein at extreme high temperatures. Additional *in vitro* and *in vivo* characterization is required to unravel the function of increased and altered TrmBL2 binding under these conditions.

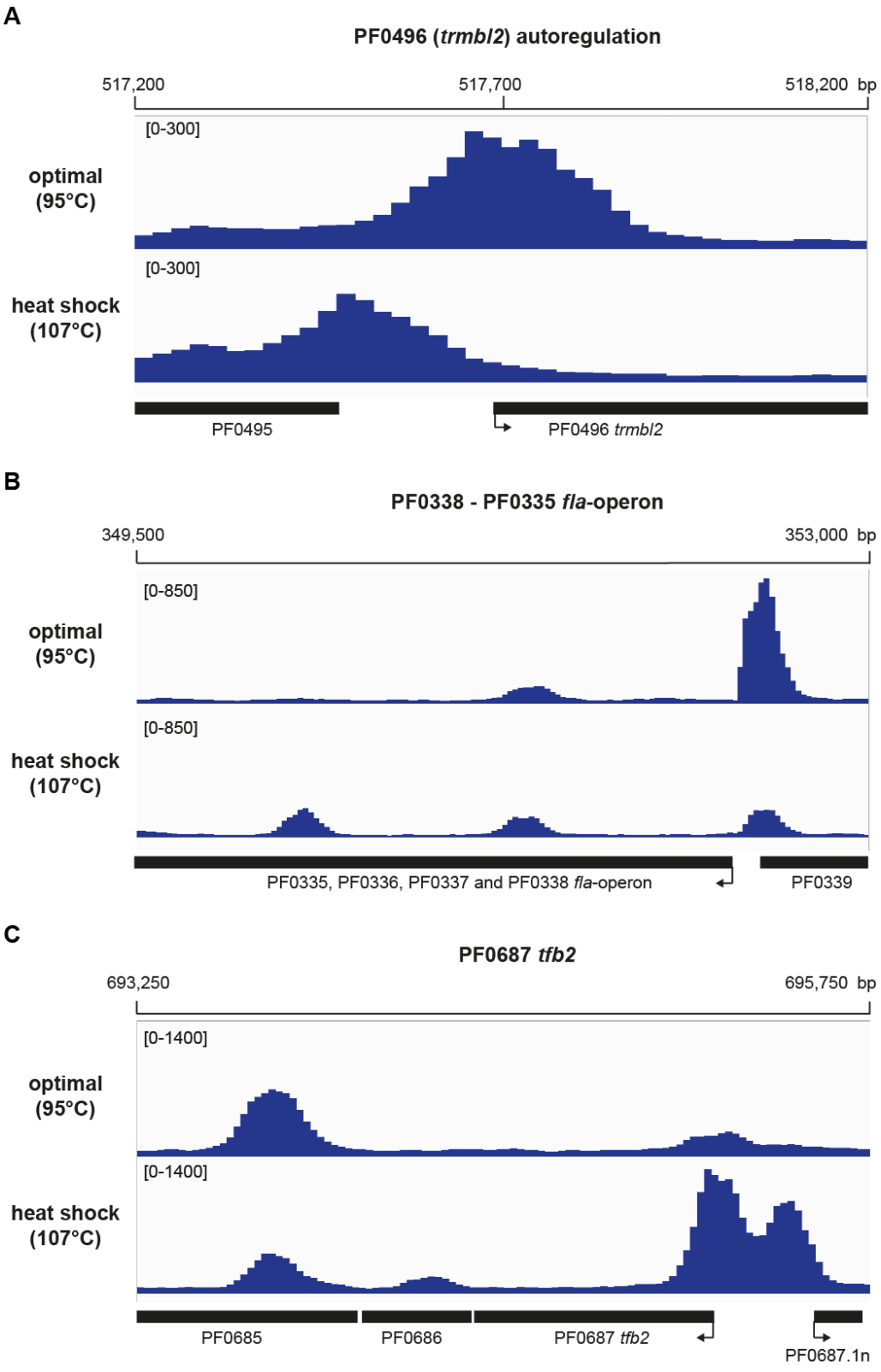


Figure 25. TrmBL2 binding patterns at specific genes vary under optimal and heat shock conditions.

TrmBL2 ChIP-seq results obtained from formaldehyde-treated cells at 95°C (optimal) or after incubation at 107°C for 30 min (heat shock) were visualized. Arrows indicate direction of transcription. A, the TrmBL1 binding site in its own promoter region (PF0496) is shifted upstream. B, the intensity of ChIP-enrichment in the *fla*-opreon (PF0338-PF0335) promoter is decreased. C, after heat shock a double signal of enrichment arises in the *tfb2* (PF0687) promoter region.

3.3.7. ChIP-qPCR analysis confirms induction of the heat shock response in *P. furiosus* cells by heating to 107°C

Heat shock response of cells after heating to 107°C was proved by ChIP-qPCR analysis. TrmBL2 ChIP-enrichment of its own intergenic region was observed under optimal conditions and upon heat shock (Figure 26, A). In contrast, no ChIP-enrichment was found under both conditions for the promoter regions of the *aaa+atpase* gene (PF1882) and the *gdh* gene (PF1602). According to Liu et al. (2007), Phr dissociates from the DNA, here shown for the *aaa+atpase* promoter (PF1882), upon heat shock (Figure 26, B). This demonstrates successful induction of the heat shock response in the *P. furiosus* cells through heating to 107°C for 30 min.

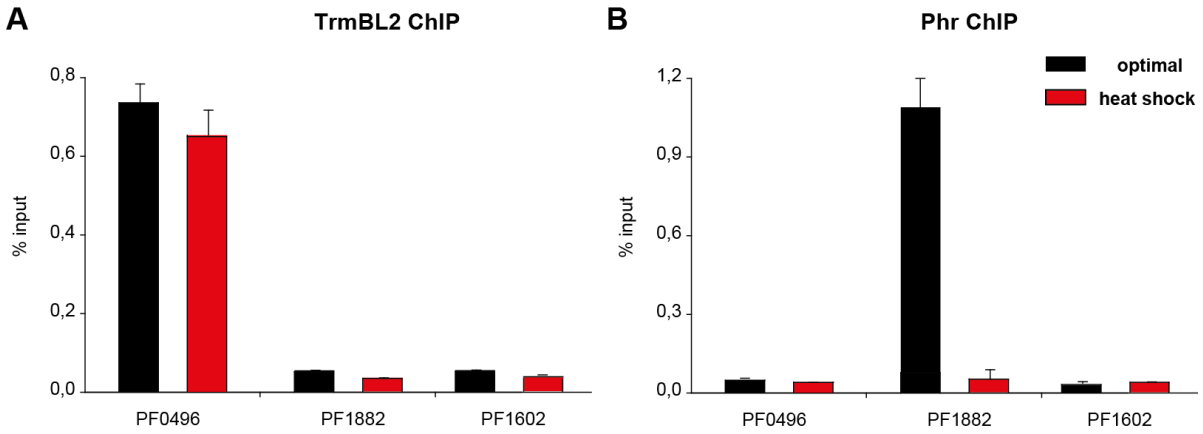


Figure 26. Phr dissociates from the DNA upon heat shock, whereas TrmBL2 remains bound to the chromosome.

ChIP enrichment is presented as % input. The mean with SD of 3 replicates of immunoprecipitation is shown for all analyzed genomic loci. A, TrmBL2 ChIP using cells grown under optimal conditions (black bars) or upon heat shock at 107°C for 30 min (red bars). The genes *aaa+atpase* (PF1882) and *gdh* (PF1602) represent negative controls for the TrmBL2 specific antibody. B, Phr ChIP using cells grown under optimal conditions (black bars) or upon heat shock at 107°C for 30 min (red bars). The gene *aaa+atpase* (PF1882) serves as positive control for the anti-Phr IgG under optimal growth conditions. The *gdh* gene (PF1602) represents the negative control for the Phr specific antibody.

3.3.8. Short conclusions TrmBL2

ChIP-seq analysis of TrmBL2 from *P. furiosus* revealed three major findings: A, *P. furiosus* TrmBL2 is an abundant chromosomal binding protein, which appears to be enriched in intergenic genomic regions in contrast to the random genomic distribution of its orthologue TK0471 from *T. kodakarensis*. B, the suggested function of TrmBL2 as global transcriptional repressor in *P. furiosus* remained unclear due to the inconsistency of the ChIP-seq and gene-deletion analysis. C, after heat shock the number of detected TrmBL2 binding sites increased as well as the binding patterns vary in contrast to optimal growth at 95°C.

3.4. Analysis of the physiological function of TFB-RF1 (PF1088) using ChIP-seq

The transcriptional activator TFB-RF1 (PF1088) was discovered by *in silico* prediction of TF binding sites in *P. furious*. TFB-RF1 binds upstream of the promoter elements of PF1089 and enhances transcription by TFB recruitment to the imperfect BRE. In a transcriptomic analysis using *P. furiosus*, expression of PF1088 and PF1089 could not be detected (Yoon et al., 2011). Moreover, western blot analysis using an antibody raised against recombinantly expressed TFB-RF1 and primer extension experiments revealed that PF1088 together with PF1089 is not transcribed and expressed in *P. furiosus* under various tested growth conditions (e.g. glycolytic and gluconeogenic). To study the function of this unusual transcriptional activator *in vivo*, Kreuzer (2009) constructed the expression vector pYS5a for over expression of TFB-RF1 in *P. furiosus* under gluconeogenic conditions (Table 9). According to the work of Waege et al. (2010) within this vector, expression of TFB-RF1 is controlled by the *fbpase* promoter (PF0613), which is highly upregulated after growth of *P. furiosus* on peptides (gluconeogenic conditions).

Table 9. Key elements of the *P. furiosus* shuttle vector pYS5a (Kreuzer, 2009).

Element	Gene	chr start	chr end	function
Promoter over expression	PF0613	633842	634045	inducible promoter (gluconeogenic growth conditions) of the *fbpase* gene for expression of TFB-RF1
TFB-RF1	PF1088	1036424	1036724	coding region TFB-RF1
Termination region	PF1831	1689565	1689648	termination region of *histone a1* used for the *tfb-rf1* as well as the *hmgcoa reductase* gene
Hmg-CoA	PF1848	1704745	1705971	HMG-CoA reductase coding region; resistance to simvastatin
Promoter constitutive	PF1602	1494934	1495218	strong promoter of the *gdh* gene used for expression of the HMG-CoA reductase

After transformation of the plasmid pYS5a in *P. furiosus* and adaptation to gluconeogenic conditions, the presence of TFB-RF1 in the crude extract was detected by western blot analysis. Moreover, primer extension analysis revealed transcription of the PF1089/PF1088 mRNA after over expression of TFB-RF1 (Kreuzer, 2009). In the western blot experiments the TFB-RF1 specific antibody detected mainly a protein in the range of 25 kDa. This size differs from the computed molecular weight of 13.8 kDa as well as observed molecular weight of recombinantly-expressed TFB-RF1 using SDS-PAGE (Ochs, 2009; Rost, 2012). The PF1088 homologue PH1061 from *P. horikoshii* exists as dimer in both the crystal and in solution (Okada et al., 2006). This may explain detection of a signal in the range of 25 kDa, representing the TFB-RF1 dimer. Moreover, Winter (2012) showed that the recombinant *P. furiosus* TFB-RF1 elutes from gelfiltration columns as dimer.

Combining over expression of TFB-RF1 and ChIP-seq will allow to prove binding of TFB-RF1 to the promoter region of PF1089 *in vivo*. Moreover, it can account for identifying novel TFB1-RF1 binding sites beyond PF1089 in a genome-wide manner. Numerous additional binding sites in the genome of *P. furiosus* were already predicted *in silico* using the software package BART (Ochs, 2009). Validation of these binding sites *in vivo* can facilitate to associate PF1088 with a specific biological process. This contributes to a better understanding of the physiological function of TFB-RF1 as transcriptional activator, which recruits TFB.

3.4.1. Over expression of TFB-RF1 in *P. furiosus* and *in vivo* formaldehyde crosslinking

P. furiosus strain MURPf5 containing the expression vector pYS5a was recultivated and adapted to gluconeogenic growth conditions. Martina Kreuzer and Mario Rost showed that in this strain the transcriptional activator TFB-RF1 (PF1088) can be specifically over expressed after growth on pyruvate (Kreuzer, 2009; Rost, 2012). *P. furiosus* cells were cultivated for crosslinking in a 15L bio-fermenter and treated with formaldehyde. Successful induction of expression of TFB-RF1 in these cells was proved by western blot analysis (Figure 27). The antibody raised against recombinantly expressed TFB-RF1 mainly detected a signal in the range of 25 kDa, which may represent a TFB-RF1 dimer. The signal for monomeric TFB-RF1 was weak and appeared as double signal. This phenomenon was already described in previous studies (Rost, 2012).

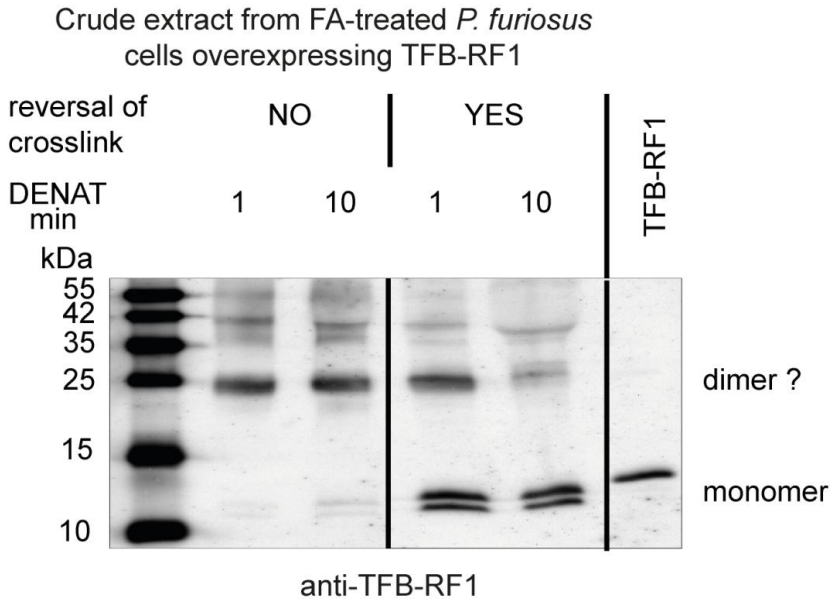


Figure 27. TFB-RF1 is expressed in the formaldehyde-treated *P. furiosus* MURPf5 cells.

15 µg of total protein crude extract derived from formaldehyde- (FA)-treated *P. furiosus* strain MURPf5 cells grown under gluconeogenic growth conditions were separated on a 15% SDS polyacrylamide gel under denaturating conditions. If indicated, reversal of the crosslink was done overnight in SDS (1% w/v final concentration) containing buffer overnight at 65°C. Before loading onto the gel samples were denaturated for 1 min or 10 min at 95°C. TFB-RF1 was detected using a rabbit antiserum raised against TFB-RF1 in a dilution of 1:1000. As positive control 10 ng recombinant TFB-RF1 was used.

Additional reversal of the crosslink overnight enhanced the signal for the monomeric state and extensive denaturating for 10 min at 95°C led to approximately complete disappearance of the signal at 25 kDa. The double signal of the monomeric form may indicate modification or degradation of the protein. Nevertheless, the western blot analysis confirmed the presence of TFB1-RF1 in MURPf5 crude extract derived from formaldehyde-treated cells.

3.4.2. TFB-RF1 binds the PF1089 promoter *in vivo*

Binding of TFB-RF1 to the promoter region of PF1089 was well studied *in vitro* (Ochs et al., 2012). ChIP-qPCR analysis demonstrated that over expressed TFB-RF1 binds to this promoter region *in vivo*. In contrast to the TrmBL1 binding site in the *pfk* promoter (PF1784), ChIP enrichment was only detected for the promoter region of PF1089 compared to *gdh* (PF1602) (Figure 28, A). In addition, using the antibody raised against TrmBL1 led to enrichment of the PF1784 promoter, but not of the PF1089 promoter (Figure 28, B). This shows specificity of the detected TFB-RF1 binding to the PF1089 promoter *in vivo*. Moreover, it demonstrates the efficient formaldehyde treatment of TFB-RF1-overexpressing cells.

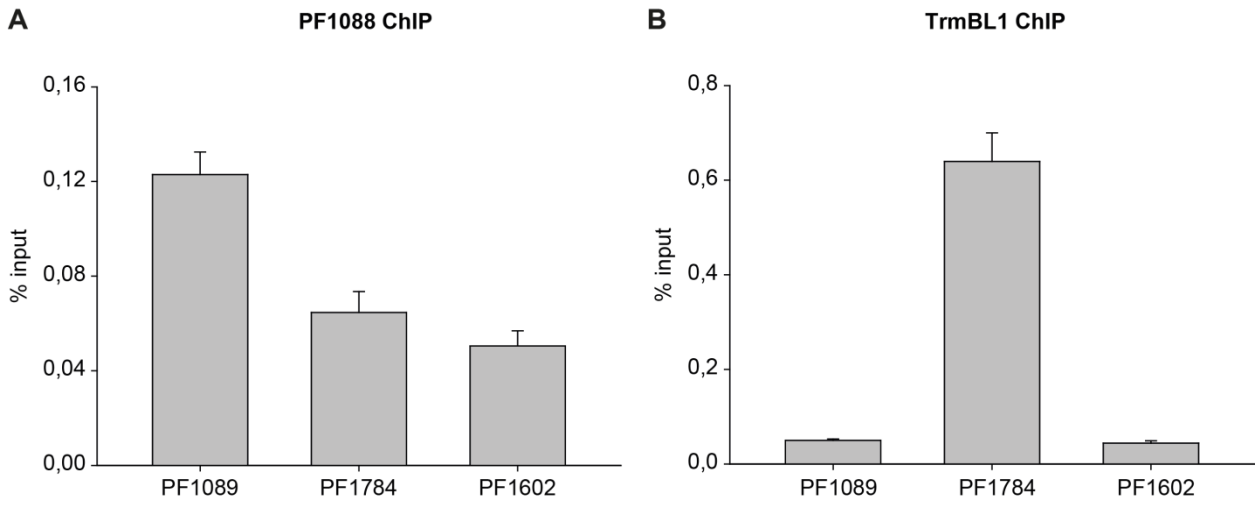


Figure 28. TFB-RF1 binds to its designated binding site in the PF1089 promoter region *in vivo*.

ChIP-enrichment is presented as % input. The mean with SD of 3 replicates of immunoprecipitation is shown for all analyzed genomic loci. A, PF1088 ChIP using MURPf5 cells grown under gluconeogenic growth condition. The genes *pfk* (PF1784) and *gdh* (PF1602) represent negative controls for the TFB-RF1 specific antiserum. B, TrmBL1 ChIP using MURPf5 cells grown under gluconeogenic growth condition. The gene *pfk* (PF1784) serves as the positive control for the anti-TrmBL1 IgG, whereas the gene *gdh* (PF1602) represents the negative control for the anti-TrmBL1 IgG.

3.4.3. TFB-RF1 exclusively binds the promoter region of PF1089 *in vivo*

Identification of additional TFB-RF1 binding sites in the genome of *P. furiosus* will raise the possibility, to elucidate the physiological function of this unusual transcriptional activator in Archaea. Applying the ChIP-seq approach to TFB-RF1 showed specific binding of the regulator to its designated binding site in the promoter region of PF1089 *in vivo* (Figure 29, B). No additional enriched region was detected, which exceeded an enrichment ratio > 1.5 (Table 10).

Table 10. Result of three TFB-RF1 ChIP-seq experiments (IP1, IP2 and IP3).

TFB-RF1 ChIP	peak start	peak stop	binds at	enrichment ratio	promoter (gene)
IP1	1036934	1037629	1037281	**1.57**	PF1089
	. . .	. . .	. . .	< 1.5	. . .
	1426665	1427355	1427010	1.22	PF1529
IP2	1036938	1037622	1037280	**1.55**	PF1089
	. . .	. . .	. . .	< 1.5	. . .
	1426651	1427351	1427001	1.23	PF1529
IP3	1036937	1037626	1037281	**1.60**	PF1089
	. . .	. . .	. . .	< 1.5	. . .
	1426657	1427353	1427005	1.23	PF1529

‘ · · · ’ represents the enriched regions identified in the coding regions of PF1602, PF1831 and PF1848.

The TFB-RF1 binding site in the promoter region of PF1089 identified by ChIP-seq was located within the TFB-RF1 DNaseI footprint region determined *in vitro* (Figure 29, C). This showed the high spatial resolution of the approach. *In vivo* TFB1-RF1 binds upstream of the BRE according to the results obtained *in vitro*. Beside enrichment in the PF1089 promoter region an additional enriched region was detected in the TFB-RF1 coding region (PF1088) (Figure 29, B). In contrast to the PF1089 promoter region, this enrichment in the coding region was also present in the input sample. These signals derived from the extra copy of the TFB-RF1 gene in the over expression plasmid pYS5a and corresponding signals were also observed for the *gdh* promoter region (PF1602), the *histona1* terminator region (PF1831) and the *hmgcoa reductase* coding region (PF1848) (Figure 29, A; Table 9; Table 10, ‘ · · · ‘).

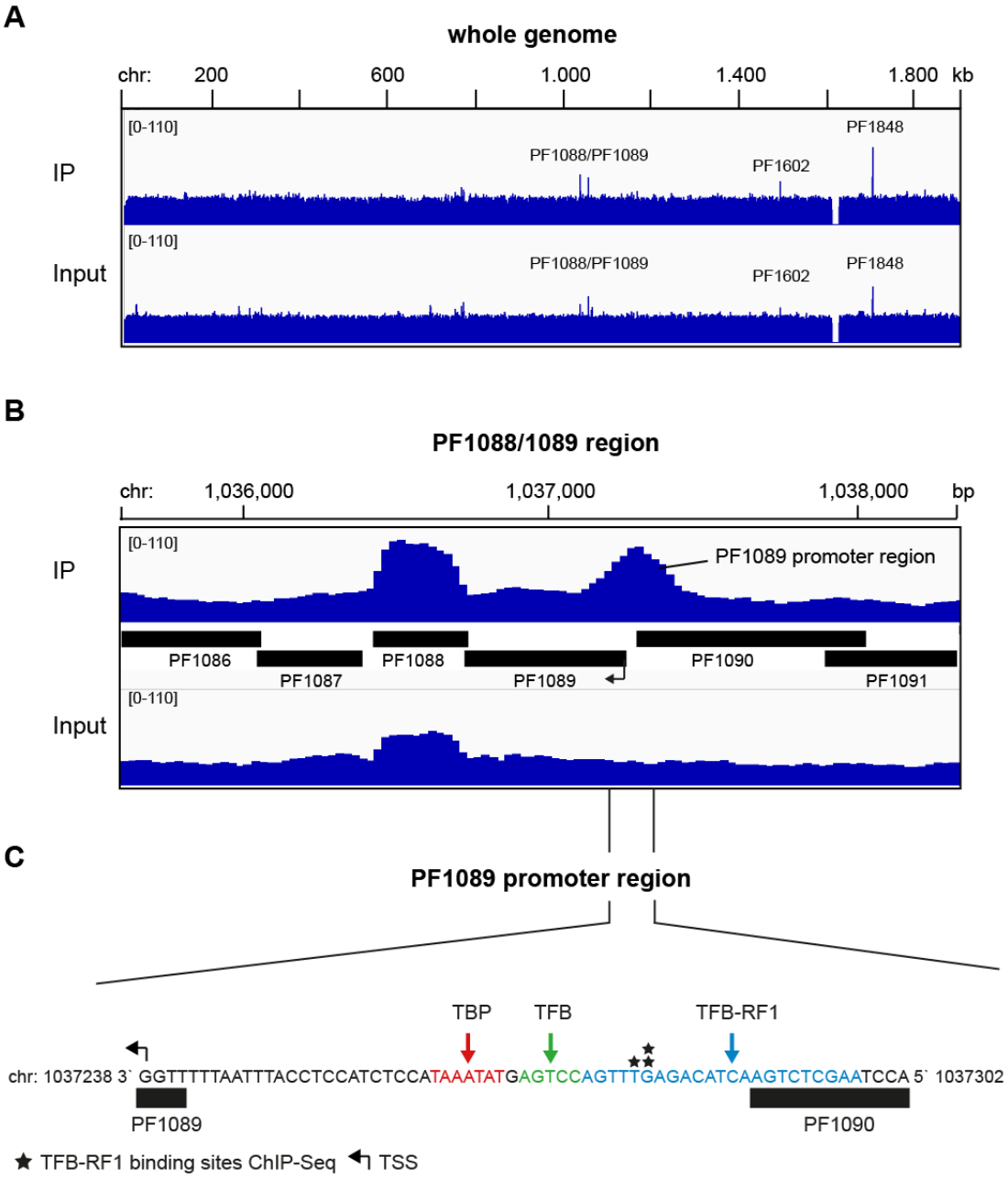


Figure 29. TFB-RF1 exclusively binds upstream of the BRE- and TATA-box of the PF1089 promoter region *in vivo*.

A, TFB-RF1 ChIP-seq results (IP3) using *P. furiosus* MURPf5 cells grown under gluconeogenic conditions were visualized. The results from the immunoprecipitation (IP) and input samples are shown. A, the genome wide view reveals only a few enriched regions. Prominent peaks are depicted. B, the detailed view uncovers specific enrichment in the PF1089 promoter. Enrichment of the PF1088 coding regions is present in the IP and input sample. C, the non transcribed strand of the PF1089 promoter region is shown. The chromosomal position in the *P. furiosus* genome is given. The promoter elements are marked with red for the TATA-box (TBP binding) and green for the BRE (TFB binding). In addition, the TFB-RF1 DNaseI footprint region is highlighted in blue (Ochs et al., 2012). Annotated coding regions of the genes PF1089 and PF1090 are shown as black bars. Transcriptional start site determined *in vitro* and *in vivo* is presented by a bent arrow (Kreuzer, 2009; Ochs, 2009). TFB-RF1 binding sites identified by ChIP Seq *in vivo* are labeled with asterisks (3 replicates of immunoprecipitation).

3.4.4. Validation of the TFB-RF1 ChIP-Seq results *in vitro* by EMSA

Exclusive binding of TFB-RF1 to the promoter region of PF1089 *in vivo*, was validated *in vitro* by EMSA experiments. As control the second highest enriched promoter region in the ChIP-seq data was included in the analysis. This is located in the promoter region of PF1529 encoding a pyridoxal biosynthesis lyase. The enrichment ratio of this region was approximately 1.23 in all three experiments (Table 10), which is lower than the defined threshold of 1.5. Unlike using the PF1089 promoter region as template, TFB-RF1 was not able to shift the promoter region of PF1529 (Figure 30). Thus, the promoter region of PF1089 remains the solely known binding site for the transcriptional activator TFB-RF1 by now.

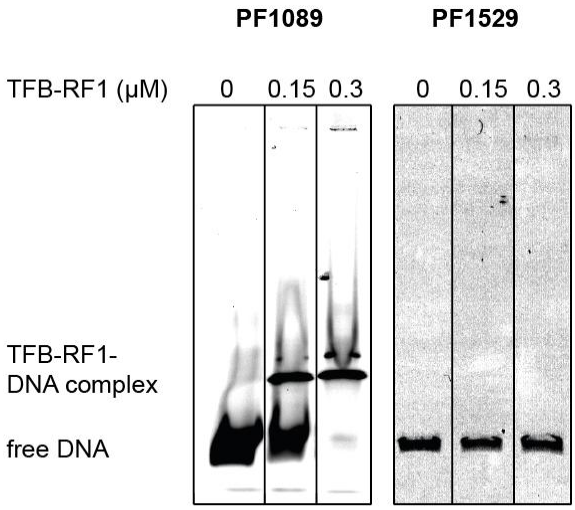


Figure 30. EMSA confirmed exclusive binding of TFB-RF1 to the PF1089 promoter region.

Cy5 labeled DNA templates were used and gelshifts were separated on a 6% non-denaturating polyacrylamid gel. Gelshifts using the PF1089 or the PF1529 promoter region as DNA template, are presented. TFB-RF1 was used in the given concentrations.

3.4.5. Short conclusion TFB-RF1

Two major findings were uncovered by combining shuttle vector based over expression of TFB-RF1 *in P. furiosus* with genome-wide TF binding site mapping by ChIP-seq: A, after over expression, TFB-RF1 binds according to the *in vitro* results upstream of the BRE in the PF1089 promoter region *in vivo*. B, the physiological function of TFB-RF1 as transcriptional activator remains unclear due to the lack of additional binding sites beyond PF1089 in the genome.

3.5. Complete ChIP-seq workflow for the hyperthermophilic euryarchaeon *P. furiosus*

Adaptation of an established ChIP protocol for ChIP-seq analysis of numerous TFs enabled the development of complete ChIP-seq workflow for the hyperthermophilic euryarchaeon *P. furiosus*. This chapter summarizes experiences and suggestions for every step of the ChIP-seq protocol including not shown experiments and data analyses.

3.5.1. Growth of *P. furiosus* cells and formaldehyde treatment

For formaldehyde treatment *P. furiosus* cells had to be cultivated in bio-fermenter scale. Culture volumes from 15 L to 100 L generated sufficient cell mass for the ChIP procedure (crosslink in the middle to late log-phase). Down-scaling the volume to 20ml or 100ml cultures using anaerobic media bottles could not be achieved. The resulting cell mass was too low for the ChIP experiments. The crosslinking protocol was carried out according to Vierke (2007) and Liu et al. (2007). Formaldehyde treatment at 85°C to 107°C for 20-30 s using 0.1 % formaldehyde (v/v) as final concentration and quenching through glycine led to an adequate preservation of protein-DNA and protein-protein interactions. Increasing the crosslinking time or concentration of formaldehyde did not improve the results and even impaired the signal-to noise ratio of ChIP. Reducing to some extend did not influence the experiments. Formaldehyde-treated *P. furiosus* cells can be stored for years at -80°C without affecting the ChIP efficiency.

During adaptation of *P. furiosus* to the varying growth conditions (glycolytic or gluconeogenic) an ambiguous growth behavior was observed. By inoculation of cells from rich medium (0.1 % starch, 0.1 % peptone and 0.1 % yeast extract) into pyruvate-containing medium (40 mM pyruvate and 0.025 % yeast extract) a cell density of > $1*10^8$ cells/ml was reached by incubation overnight (~8 h to 10 h) at 85°C to 95°C. In contrast, after inoculation in medium containing starch as sole carbon source (0.1 % starch and 0.025 % yeast extract) an extended lag-phase was observed and a cell density of > $1*10^8$ cells/ml was reached after incubation for 16 h to 18 h. Previous ChIP-seq experiments demonstrated that TrmBL1 binding to its operator regions did not occur in cells grown on rich medium according to the only-starch conditions (data not shown). This suggests that cells are able to adapt rapidly from proposed glycolytic conditions (rich medium) to gluconeogenic conditions (pyruvate medium), whereas during adaptation from rich medium to only-starch medium processes beside the glycolysis-gluconeogenesis switch play a role and cause an extended lag-phase.

3.5.2. Lysis of cells and fragmentation of genomic DNA by sonication

Depending on the number of intended immunoprecipitation experiments 0.1 to 0.5 g formaldehyde-treated *P. furiosus* cells were resuspended in 700 µl PBS buffer. In some crude extracts microscopy showed presence of numerous crystal-like structures, which interfered with the fragmentation by sonication. These crystal-like structures can be removed by washing the cells at low pH (titration with HCl) and subsequent neutralization through washing with PBS buffer (Eisenschink, 2010). Best sonication results (average fragment length 250 to 450 bp; Figure 31) were obtained using gently conditions, but increasing the total sonication time to 36 min (12 x 3 min; output control: 2 -2.5; 10 – 20 %; Branson sonifier 250, mikrotip). Sonication had to be carried out on ice to avoid heating of the samples. Increasing the sonication intensity led to strong heating as well as foaming of the samples.

This impaired with the fragmentation efficiency. After sonication, cell debris was removed by centrifugation (21,000 rcf). The used force did not sediment all debris; however, resulting crude extracts could be successfully used for ChIP. For complete sedimentation, cell debris had to be removed by ultracentrifugation (100,000 rcf). Nevertheless, this did not improve the results. Sonicated crude extract from formaldehyde-treated *P. furiosus* can be stored for up to 1 month at -20°C and for up to 1 year at -80°C without any loss of ChIP efficiency.

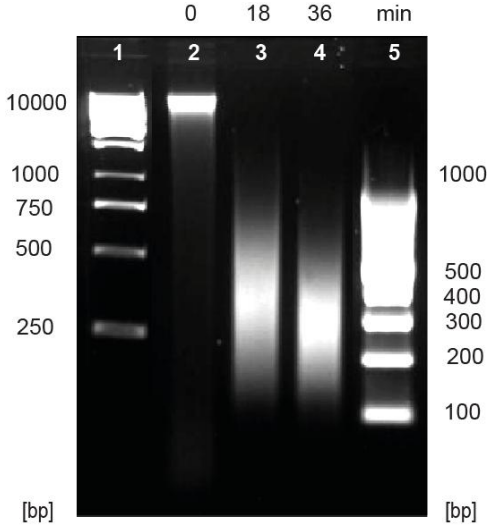


Figure 31. Fragmentation of *P. furiosus* genomic DNA by sonication.

Time course of sonication of formaldehyde treated *P. furiosus* cells, grown on rich-medium. Crosslink was reversed over night and fragmented DNA was analyzed by a 1.5% agarosegel. 1, 1kB ladder; 2, no sonication; 3, after sonication for 18 min (total time); 4, after sonication for 36 min (total time); 5, 100 bp ladder.

3.5.3. Immunoprecipitation using specific antibodies

For immunoprecipitation crude extracts were adjusted to a concentration of 30 ng/µl DNA and 500 µl aliquots (15 µg total DNA) were used for IPs. This represents the minimal amount of DNA sufficient for ChIP combined with analysis by qPCR as well as high-throughput sequencing. Protein-of-interest-specific antibodies were coupled to 50 to 100 µl magnetic Protein G beads (Dynal). Using more than 100 µl beads tremendously increased the background (for 15 µg total DNA). The optimal amount of antibodies had to be determined for every single antibody and varied between 1 to 15 µg (for 50 µl magnetic beads). It is recommended to pre-purify the IgGs via Protein G columns. Using directly the antiserum increased the background. These steps did not influence the ChIP results using high-affinity antibodies like anti-TrmBL1/BL2 IgGs. In contrast, all optimization steps had to be carried out for anti-PF1088 IgGs to maintain successful ChIP-seq experiments. Beside the mock controls additional negative controls using no antibody and pre-immunesera were carried out for all experiments (data not shown) and are recommended. All tested negative controls showed comparable low background signals. The five different denoted washing steps were sufficient to minimize background. Additional washing steps did not improve the signal-to-noise ratio. Elution of the immunoprecipitated protein-DNA complexes by 65°C in the appropriate SDS-containing buffer recovered approximately all bound complexes.

All attempts to establish an epitope-tag-based IP procedure were not successful (data not shown). The following widely used epitope tags were tested as single or in combination as tandem-affinity tags: c-myc, HA, Flag, His_6 and Strep (Evan et al., 1985; Field et al., 1988; Hochuli et al., 1988; Hopp et al., 1988; Schmidt, Thomas G M and Skerra, 2007). The most promising results were obtained using the Strep-His_6 tandem affinity tag according to the

described *in vivo* purification system of *P. furiosus* RNAP (Waege et al., 2010). However, no successful IPs could be accomplished yet and additional optimization steps are required.

3.5.4. Removal of crosslink and DNA purification

Removal of the crosslink was carried out in SDS containing buffer overnight at 65°C. Minimal incubation time at 65°C was 4 hours; however, overnight is recommended. 30 min RNaseA digestion had to be carried out before purification. Additional ProteinaseK digestion is recommended for optimal purification of immunoprecipitated DNA fragments. For purification of ChIP-enriched and input DNA the Qiaquick PCR Purification Kit (Qiagen) yielded the best amounts compared to other commercially available DNA purification systems. Purified immunoprecipitated and input DNAs could be stored for at least one year at -20°C.

3.5.5. ChIP-qPCR

Successful ChIP-qPCR experiments especially relied on the PCR efficiency of the designed primer pairs. For primer design the free-available online tool Primer3 (http://bioinfo.ut.ee/primer3-0.4.0/; (Untergasser et al., 2012) was used according to the commonly accepted recommendations for qPCR primers and corresponding amplicons. The known or anticipated binding site of the protein-of-interest should be located in the center of the qPCR amplicon. The detected spatial resolution of ChIP-qPCR was 200 to 400 bp (data not shown). This was sufficient for initial experiments and validation of ChIP-seq enriched regions.

3.5.6. Library preparation

Initially, library preparations were carried out according to a free-available protocol, which was adapted by Eisenschink (2010). Additionally, the barcoding system according to Lefrancois et al. (2009) was used for multiplex short-read DNA sequencing of different samples, which enabled simultaneous sequencing and analysis of four samples using Illumina's platform. During the thesis the library preparation protocol was replaced by a commercially available DNA Library Prep Master Mix Set for Illumina® sequencing (NEBNext®) and the corresponding NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1 and 2) were used for multiplex sequencing, which enables simultaneous sequencing and analysis of up to 24 samples. Changing the library preparation system improved the signal-to-noise ratio of the ChIP-seq approach, whereas the total number of detected peaks did not vastly differ between both protocols.

Moreover, the deletion of the 16 kb fragment as well as the observed aberrations in chromosomal occupancy (chromosomal position 628,000 to 797,000 bp) after mapping were present using both systems. This phenomenon was specifically identified in the ChIP-seq experiments using *P. furiosus* cells grown on only-starch medium (glycolytic conditions). In contrast, this aberration was not detected after growth of the cells under gluconeogenic conditions (only-pyruvate) or on the rich medium (0.1 % starch, 0.1 % peptone and 0.1 % yeast extract; glycolytic, too). The physiological relevance of the detected aberrations has not been characterized yet and is still unclear. The aberration in genomic occupancy can also represent a sequencing artifact.

3.5.7. Sequencing

Before sequencing, success of ChIP-enrichment was proved by ChIP-qPCR analysis of known or predicted TF binding sites. This step is recommended to avoid unusable ChIP-seq results. Illumina HiSeq sequencing was carried out by the KFB Regensburg (Kompetenzzentrum für Fluoreszente Bioanalytik). This included quality assessment of the ChIP-seq libraries, multiplexing, sequencing, demultiplexing and finally, initial quality assessment of resulting short reads. In the first sequencing experiment (2010) one sequencing-run yielded ~20 million short reads (50 bp read length) per sequencing lane. The last experiments, which were carried out in 2013, yielded more than 150 million short reads (50 bp read length) per sequencing lane. The total number of short reads per ChIP-sample varied between 4 million (105 x genome coverage; multiplexing of 4 samples) to 15 million (393 x genome coverage; multiplexing of 12 samples) short reads. The obtained genome coverage was by far sufficient and even impaired with the sensitivity of the ChIP-seq approach. Thus, multiplexing of 24 and more ChIP samples for parallel sequencing and analysis will be applicable for ChIP-seq experiments using *P. furiosus*.

3.5.8. Data analysis

For researchers, which are inexperienced in dealing with big-scale data sets, the public-available Galaxy server (http://galaxyproject.org) provides powerful tools for ChIP-seq analysis (Blankenberg et al., 2010). However, running these tools as stand-alone software enlarges the options to improve every step of ChIP-seq data analysis.

The first critical step in data analysis represents the short-read alignment of ChIP-seq sequences to the reference genome. At the beginning of the experiments in 2010, the available short-read alignment tools like Bowtie were able to map the majority of short reads in a correct manner (Langmead et al., 2009); however, for some regions, which contain repetitive elements and are present as multiple highly similar copies in the genome like transposases, an unspecific enrichment of short-reads was observed. Improvements of the underlying algorithm (Bowtie 2) during the past years mostly overcame these problems (Langmead and Salzberg, 2012). Besides the obvious deletion of the 16 kb fragment and aberrations in genomic occupancy of mapped short reads, numerous minor anomalies were detected. This indicates that the available genome annotation of *P. furiosus* type strain DSM3638 (NCBI RefSeq Accession: NC_003413) still contains numerous *ab initio* sequencing and assembly errors like single nucleotide mismatches, deletions or insertions, which interferes with correct gene annotation (Robb et al., 2001; Poole et al., 2005; Gao and Wang, 2012). Applying the complete Genome Analysis Toolkit (GATK) workflow on the input samples derived from cells, grown under varying conditions (pyruvate 2, pyruvate 5, starch 2 and heat shock; two replicates each), discovered 279 conserved variants (single nucleotide mismatches and deletions as well as insertions of up to five bp length) in comparison to the reference genome (data not shown; GATK analysis was carried out according to GATK best practices: VanDerAuwera, Geraldine A. et al., 2002; McKenna et al., 2010; DePristo et al., 2011). Evaluations by comparing the variant sequences with the recently published genome sequence of the *P. furiosus* COM1 strain (NCBI RefSeq Accession: NC_018092; Bridger et al., 2012) confirmed the reliability of the majority of identified variants. Furthermore, some variants could be validated by literature search. Comfort et al. (2008) showed that the PF1109-PF1110 locus represents a single ORF, which encodes a single 120 kDa starch binding protein. Originally, PF1109 and PF1110 were annotated as separate hypothetical

proteins in the *P. furiosus* genome. The GATK analysis revealed a missing C nucleotide between the chromosomal positions 1058815 (A) and 1058816 (C). This frame shift (AC to ACC) results in one single PF1109-PF1110 ORF. The GATK workflow is suitable for identification of minor sequence variants and larger deletions in the genome become obvious through reduced genome occupancies of mapped reads. In contrast, recognition of larger insertions is still challenging. The outcome of short-read mapping to a genome containing missing DNA sequences can be seen in Figure 25, B (chapter 3.3.6). The TrmBL2 ChIP-seq peak in the promoter region of the *fla*-operon appears to be truncated in downstream (reverse) direction. This derives from a missing 771 bp fragment in the PF0338/PF0339 intergenic region, which has already been described by Näther (2008). Thus, the presented peak represents only two third of the complete peak, which cannot be correctly illustrated by using the annotated *P. furiosus* reference genome. In addition, an example of miss annotated ORF is shown in the discussion for PF1476, which encodes a PadR-like transcriptional regulator (chapter 4.2). These findings demonstrate that for the correct analysis of the ChIP-seq results numerous additional factors have to be considered beside the standard operations.

The outcoming result of a ChIP-seq experiments especially relies on the chosen peak calling algorithm (Pepke et al., 2009; Malone et al., 2011; Landt et al., 2012). In this study the Pique software was applied as peak calling algorithm because of its practicability for archaeal systems (Wilbanks et al., 2012). Analysis by Pique unproblematically functioned for the TrmBL1 and PF1088 ChIP-seq experiments. An enrichment ratio threshold of 1.5 was found to give best results regarding sensitivity as well as specificity. Moreover, the results obtained from subsequent *in vitro* validation consolidated this cut off. Originally, analysis was carried out by MACS (Model-Based analysis of ChIP-seq); however the called enriched regions varied between 1000 to 5000 bp covering the binding sites (Zhang et al., 2008). This resolution was not sufficient for the highly compact archaeal genome of *P. furiosus* and presumably derived from the background signals (high genome coverage), the observed genomic occupancy aberrations and identified deletions. In contrast, the spatial resolution of the Pique software was as high as described previously in the literature for ChIP-seq (< 20-50 bp) and the Pique algorithm was able to handle the identified genomic occupancy aberrations and deletions. Furthermore, analysis of the TrmBL2 ChIP-seq results depicted the slight limitations of the Pique software. Especially close clustering TrmBL2 peaks or highly enriched regions were not correctly called by the software. Thus, differences between the replicates as well as the analyzed conditions had to be verified using the genetrack peak calling algorithm (Albert et al., 2008). In the TrmBL2 ChIP-seq experiments peak calling using only Pique produced a large amount of false negatives, whereas applying only the genetrack algorithm led to a high rate of false positives (data not shown). Combined analysis overcame these constraints of both algorithms.

3.5.9. Validation of ChIP-enriched regions *in silico*, *in vitro* and *in vivo*

It is recommended to confirm ChIP-seq-enriched regions by ChIP-qPCR analysis of selected genomic regions. In the TrmBL1 ChIP experiments this revealed a high correlation of ChIP enrichment between both approaches. Comparable results were obtained in TrmBL2 and TFB-RF1 ChIP experiments. Moreover, the transcription system of *P. furiosus* provides a variety of well-established approaches like cell-free transcription, EMSA or footprinting for studying TF binding sites and TF-mediated transcriptional regulation (repression or activation) *in vitro*. In addition, identified TF binding sites can be used for *de novo* motif

discovery. The MEME suite (http://meme.nbcr.net/meme/) offers a bulk of different software tools for further analysis of ChIP-seq results (Bailey et al., 2009). The obtained results should be verified by secondary algorithms.

3.5.10. Short conclusion

The ChIP-seq experiments presented in this thesis represent the final optimized experiments of each studied TF. All steps in between led to the establishment of a complete ChIP-seq workflow for *P. furiosus*, which can be easily adapted for analysis of any protein-of-interest providing that a highly-specific antibody is available. Beside the crosslinking and IP procedure per se, data analysis and postprocessing steps in the ChIP-seq workflow are still challenging; however, the development of improved algorithms for these steps like the peak calling software Pique aided to manage these limitations.

4. Discussion

4.1. Distribution of TrmB/TrmBL proteins within the *Thermococcales*

Analysis of the distribution of TrmB/TrmBL1 proteins within the *Thermococcales* demonstrated that they can be divided into numerous TrmB/TrmBL clusters. This classification only relied on a MSA of the TrmB_domain. Beside the already designated TrmB proteins TrmB, TrmBL1, 2, 3, and 4 six novel TrmBL protein subgroups were found (TrmBL6 to TrmBL10 and unclassified TrmBL proteins) (Lee et al., 2007b). Furthermore, each subgroup is characterized by secondary unique features in addition to the TrmB_domain. For example, the three TrmB/TrmBL proteins (TrmB, TrmBL1 and TrmBL2) from *P. furiosus* analyzed in this study contain a second domain designated as Regulator_TrmB domain. Moreover, in contrast to TrmB and TrmBL1, TrmBL2 lacks the sugar binding part of the domain (Lee et al., 2007b; Lee et al., 2008). In addition, they differ regarding their genomic context. TrmBL1 and TrmBL2 are transcribed as single ORFs, whereas TrmB is part of an operon, which encodes a trehalose/maltose-specific ABC transporter system (TM-system) (Lee et al., 2003; Lee et al., 2005). These additional features are shared with the members of corresponding TrmB/TrmBL clusters from other *Thermococcales* species. This suggests that the split of these TrmB/TrmBL subgroups already existed before the evolutionary divergence of the species. However, lateral gene transfer events should not be disregarded, as shown for TrmB and the TM-system of *T. litoralis* and *P. furiosus* (DiRuggiero et al., 2000).The early divergence of TrmB_domain containing proteins is also indicated by the weak bootstrap values in the center of the unrooted phylogenetic tree of this study.

Moreover, TrmB/TrmBL proteins containing the composed domain architecture (TrmB and Regulator_TrmB) can be found in Archaea and Bacteria (Maruyama et al., 2011). However, in the Archaea they are only widely distributed in Euryarchaeota, whereas only a few corresponding proteins can be found in Crenarchaeota. Thus, it is still controversially discussed, where this group of TrmB/TrmBL proteins derived from. It might have emerged in one of these genera and spread by horizontal gene transfer to other archaeal and bacterial species (Maruyama et al., 2011). So far, only proteins of the TrmB_domain family from Archaea were characterized regarding their function *in vivo* and *in vitro*. Comparative studies between Archaea and Bacteria would contribute to a better understanding of the specific functions of these types of proteins in an evolutionary context. Not all TrmB/TrmBL proteins can be found in all analyzed 13 *Thermococcales* species. The repertoire of TrmB/TrmBL proteins varies from only four proteins in *T. onnurineus* to 13 different TrmB/TrmBL proteins in *T. barophilus.* To some extent this may reflect the adaptation of the corresponding *Thermococcales* species to different environmental habitats. Only two clusters are universally distributed within the *Thermococcales*. For the TrmBL2 proteins it is suggested that their evolutionary conservation within the Euryarchaeota including the *Thermococcales* derive from their function as chromosomal binding proteins according to chromosomal proteins from Bacteria like H-NS in the Gram-negative Gammaproteobacteria (Maruyama et al., 2011; Dorman, 2013). Beside the TrmBL2 proteins also members of the TrmBL8B subgroup, which only contain the TrmB_domain, were found to be universally distributed. The function of theses proteins is up to date unclear. For a better characterization, the phylogenetic analysis of TrmB/TrmBL proteins has to be extended to all described archaeal phylas and finally should also contain members from bacterial species.

4.2. ChIP-seq revealed an extended function of TrmBL1 in regulation of metabolism in *P. furiosus*

The initial purpose of this study was to dissect the particular functions of the two paralogous TFs TrmB and TrmBL1 from *P. furiosus* in regulation of sugar metabolism in a genome-wide manner by ChIP-seq *in vivo*. Due to the high similiarity of both proteins regarding their AA composition, specificitis of the used antibodies represent a crucial feature for efficient and reliable ChIP experiments. While the antibody raised against TrmBL1 showed a high specificity for its antigen, the anti-TrmB antibody exhibited a cross-reactivity with TrmBL1 and could not used for the ChIP assay.

Moreover, mapping of the short-reads from the immunoprecipitation and input samples to the *P. furiosus* genome during ChIP-seq analysis revealed the deletion of a 16 kb genomic region covering the genes from PF1737 to PF1751 in the used *P. furiosus* strain DSM3638. The deletion of this 16 kb region was confirmed through southern blot and copy number analysis. It represents one of six highly variable chromosomal regions, which were previously described in a population of seven *Pyrococcus* isolates from Vulcano Island, Italy (White et al., 2008). The deleted genomic region comprises the genes encoding the TM-system (trehalose/maltose-specific ABC transporter system) including the TF TrmB. This 16 kb fragment is flanked by two insertion sequence (IS) element sequences and it is proposed as an evidence for a recent transposon mediated gene transfer between *P. furiosus* and *T. litoralis* (DiRuggiero et al., 2000; Escobar-Paramo et al., 2005). Moreover, the two ISs (PF1737 and PF1752) are identical and belong to the IS*Pfu*1 group (formal name), which is a member of the IS*6* family. They build together with the 16 kb fragment a composite transposon, which is flanked by matching direct repeats (Figure 32). This indicates insertion as a complete composite transposon (Hamilton-Brehm et al., 2005).

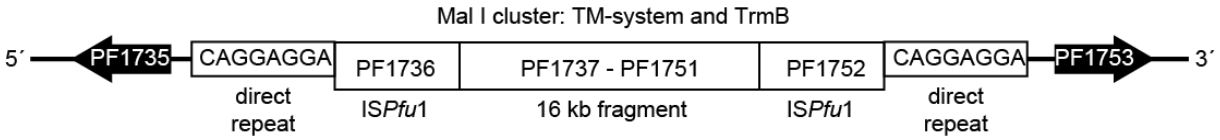


Figure 32. The 16 kb fragment represents together with the flanking IS elements a composite transposon.

The genomic region covering the genes PF1737 to PF1751 is flanked by two transposons (PF1736 and PF1752), which belong to the IS*Pfu*1 group. The whole composite transposon is encompassed by two identical direct repeats. According to Hamilton-Brehm et al. (2005).

Mapping of the ChIP-seq short reads suggests that in the *P. furiosus* strain used in this study only the fragment covering the genes PF1737 to PF1751 is deleted and the two ISs remain in the genome; however mapped reads can also derive from the six additional highly similar IS*Pfu*1 group transposons present in the *P. furiosus* genome. Members from the IS*6* family usually transpose via replicative transposition. This is accomplished by formation of cointegrates between the donor and target sites that resolve, leaving a copy of the IS in the target and the donor site (Mahillon and Chandler, 1998). Thus, deletion of the 16 kb fragment presumably does not occur because of active transposition of the whole composite transposon. In contrast, the deletion can also arise from the composition of the original culture. Two possible scenarios can not be excluded: A, the original culture is a mixed culture

containing *P. furiosus* cells with and without the deletion. B, within one *P. furiosus* cell multiple copies of the genome exist, which contain the 16 kb fragment or not. The number of chromosome equivalents per *Pyrococcus* cell can reach up to 15 during the exponential growth phase (Matsunaga et al., 2001). The *P. furiosus* type strain DSM3638 culture used in this study was obtained from the DSMZ in 2009. Southern blot analysis verified presence of the 16 kb fragment in this strain directly after reception (data not shown). Thus, the strain was cultivated in rich medium supplemented with 0.1 % starch, 0.1 % peptone as well as 0.1 % yeast extract and stored. After recultivation in the corresponding rich medium, presence of the 16 kb fragment was already underrepresented in the genomes. Furthermore, the fragment was almost completely deleted after additional cultivation steps in minimal medium supplemeted with 0.025 % yeast extract and either 0.1 % starch or 40 mM pyrvuvate. This suggests that the used cultivation condition of this study, which represents the usual conditions for cultivation of *P. furiosus* in Regensburg, promotes the deletion of the TM-system. In contrast, the conditions used from the DSMZ (0.5 % peptone and 0.1 % yeast extract; https://www.dsmz.de) appears to avoid the deletion. The key compound may represent the starch. For heterotrophic growth on starch, the polysaccharides are degraded yielding maltodextrins and maltose, which are transported in the cells by two separate ABC transporters. The TM-system is specific for maltose and the MD-system is specific for maltodextrins (Lee et al., 2006a). Moreover, it is assumed that 1 or 2 ATPs (complex contains two ATPases) are required for the uptake (Lee et al., 2007a). Thus, the cells have to spend at least 1 ATP for the uptake of maltose, which equates to 2 glucose molecules available for glycolysis, or 1 ATP for the uptake of maltodextrin, which equates to 3 to up to 20 glucose molecules available for glycolysis. In the presence of starch as carbon source, this energetic divergence may lead to preferential uptake of maltodextrins compared to maltose and benefits cells, which only contain the MD-system and not the TM-system. In contrast, presence of maltose and trehalose should prevent deletion of the TM-system. However, for the used strain even growth on maltose did not inhibit the deletion (Figure 19). One explanation for this result represents the used maltose, which can contain up to 5 % polysaccharides (Hamilton-Brehm et al., 2005). So far, the growth experiments were not repeated using ultrapure (>99 %) maltose. In conclusion, genome wide binding sites mapping of TrmB using ChIP-seq was not only impossible due to the crossreactivity of the TrmB specific antibodies, but it was also not feasible because of the deletion of the TF and the enclosing 16 kb genomic region in the *P. furiosus* strain used in this study under all tested growth conditions.

Focusing on ChIP-Seq analysis of TrmBL1 revealed in total 28 TrmBL1 binding sites under gluconeogenic and glycolytic growth conditions. The majority of binding events (~90 %) were found under gluconeogenic growth condtions, while under glycolytic growth conditions only four binding sites were detected. The function of TrmBL1 binding to three sites, which were solely found under glycolytic conditions and which are located in the coding regions of genes, remained unclear so far. No significant changes of relative transcript levels of the corresponding genes were observed in a whole genome microarray analysis of *P. furiosus* grown on carbohydrates or peptides (Schut et al., 2003). Moreover, in all identified ChIP-enriched regions presence of the palindromic DNA motif TGM was shown by computational analysis (Figure 33). This indicates that *in vivo* TrmBL1 exclusively recognizes TGM containing areas in the genome.

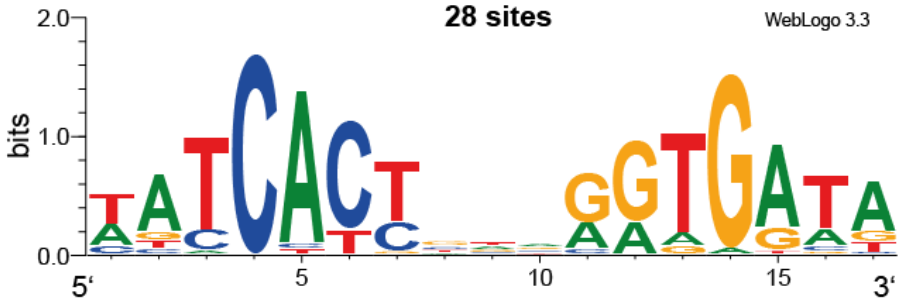


Figure 33. TGM logo based on the 28 TrmBL1 binding sites identified in this study.

The TGM was found in 28 genomic regions of *P. furiosus*, which were detected during ChIP-seq analysis of TrmBL1 under glycolytic and gluconeogenic growth conditions. The logo was created using WebLogo 3.3 (http://weblogo.threeplusone.com).

Numerous TrmBL1 binding sites found under gluconeogenic growth conditions were already known from *in vivo* (Tgr from *T. kodakarensis*) as well as *in vitro* studies or predicted by *in silico* analysis (VanDeWerken et al., 2006; Kanai et al., 2007; Lee et al., 2007b; Lee et al., 2008). The corresponding genes encode proteins and enzymes that catalyze sugar uptake (e.g the MD-system) and glycolysis (e.g. PF1784, *pfk*). Moreover, ChIP-seq confirmed transcriptional regulation of the gene PF1109/1110 by TrmBL1, which encodes one single protein (false annotated as two separate proteins). This encourages the hypothesis, that the PF1109/1110-encoded protein functions as extracellular starch protein, which likely plays a role in sequestering extracellular α-glucans for subsequent hydrolysis by amylolytic enzymes (Comfort et al., 2008). In the promoter region of these genes the identified TrmBL1 binding sites and TGMs are exclusively located downstream of the BRE and TATA-box. They overlap the TSS and TrmBL1 binding likely prevents RNAP recruitment, which leads to transcipitional repression.

The only solely-gluconeogenic gene identified by ChIP-seq in this study was the *gapdh* gene (PF1874). In the *gapdh* promoter the TrmBL1 binding site is located upstream of the promoter elements and transcription is activated. Unexpectedly, binding of TrmBL1 in the promoter region of the *fbpase* gene (PF0613) was not detected by ChIP-seq. Additionally, ChIP-qPCR analysis of this region exhibited only weak ChIP-enrichment (data not shown). In contrast, previous studies showed TrmBL1-mediated activation of this gluconeogenic-specific gene for *P. furiosus in vitro* and in *T. kodakarensis in vivo* (Kanai et al., 2007; Lee et al., 2008). Detailed analysis of the *fbpase* promoter region revealed upstream of the designated TGM a second additional BRE and TATA-box in *P. furiosus* (-100 bp of the translational start site), which is not present in *T. kodakarensis*. Primer extension analysis and cell-free transcription confirmed the presence of two RNA transcripts *in vivo* and *in vitro*, derived from initiation at both promoters (Gindner, 2009). This suggests existence of a more sophisticated transcriptional regulation mechanism for this gene in *P. furiosus in vivo*, which may explain disposability of TrmBL1 binding in the PF0613 promoter in the current ChIP-seq analysis.

Furthermore, previous *in vitro* studies showed autorepression by TrmBL1 of its own promoter (Lee et al., 2007b). This autoregulation mechanism could not be proved *in vivo*. In contrast to *T. kodakarensis*, the TrmBL1 promoter has no TGM. Thus, it appears that binding of TrmBL1 to TGM-containing promoters and its function as transcriptional regulator in *P. furiosus* is controlled independent from autoregulation.

Additionally, the ChIP approach also revealed novel, unknown target genes of TrmBL1 mediated transcriptional control. Some of them like PF1350, which encodes a putative membrane-spanning transport protein of the major facilitator superfamily, may play a role in sugar uptake. In contrast, positive regulation of the *l-asparaginase* gene (PF2047) links the regulation by TrmBL1 to the AA metabolism. The position of the TrmBL1 binding site determined by ChIP-seq *in vivo* and of the TGM was more than 50 bp upstream of the predicted corresponding BRE and TATA-box. This shows that TrmBL1 is able to enhance transcription through binding close upstream (*gapdh* promoter, PF1874) as well as far upstream (*l-asparaginase* gene, PF2047) of the promoter elements. However, the mechanism of transcriptional activation through stimulation of TBP, TFB or RNAP binding remained unclear. Recently, MalR has been identified as TrmB-like protein, which acts as transcriptional activator of a maltose/maltodextrin transport gene cluster in *S. acidocaldarius* (Wagner et al., 2014). According to the *l-asparaginase* gene the identified repeats, important for MalR mediated transcriptional control of the gene cluster, are located more than 50 bp upstream of the corresponding BRE and TATA-box (61 bp and 94 bp).

The novel identified binding site of TrmBL1 upstream of PF0967 is located within an operon transcribed from PF0971 to PF0965. While PF0971 to PF0968 encode the subunits of the VOR, PF0967 to PF0965 encode subunits of the POR. *In vitro* transcription assays verified the presence of an additional transcript for the genes PF0967 to PF0965 beside the main transcript of the whole operon. The TGM is located downstream of the promoter elements, which causes negative transcriptional control of the additional transcript by TrmBL1. Under gluconeogenic growth conditions this leads to less conversion of pyruvate to acetyl-CoA, which increases the amount of available pyruvate for gluconeogenesis. The function of transcription promoters within archaeal operons and their implications for modulating different responses according to the environmental challenge has been already shown in *H. NRC1* (Koide et al., 2009).

Through regulation of expression of a putative regulator of the PadR-family (PF1476) transcriptional control by TrmBL1 can be extended to promoters of genes, which have no TGM. Transcriptional repression of PF1476 by TrmBL1 is conserved between *P. furiosus* and *T. kodakarensis* (supplemental data microarray analysis, but not mentioned in the manuscript; Kanai et al., 2007). The PF1476 gene is false annotated in the genome, which was indicated by the position of the TrmBL1 binding site identified by ChIP-seq. The presumably correct translation start site is located 153 bp upstream of the annotated translation start site. The corrected PF1476 ORF encodes a ~13 kDa protein, which forms a dimer *in vitro* (BS_3 crosslinking and denaturating SDS-PAGE; data not shown). For preliminary characterization binding of PF1476 to three randomly selected promoter regions was studied by EMSA analysis. The corresponding genes (PF0531 (reverse) / 0532 (forward), cobalt metabolism / acetyl-coA synthetase subunit; PF0972, see below; and PF1911, ferredoxin:NADPH oxidoreductase II) were upregulated during growth of *P. furiosus* under gluconeogenic conditions (Schut et al., 2003). PF1476 was able to bind the intergenic region between the genes PF0971 and PF0972, whereas it was not able to recognize the other promoter regions (Figure 34, not all data shown). PF0972 is co-transcribed with the genes PF0973 and PF0974 within an operon, which is upregulated under gluconeogenic growth conditions *in vivo* (Schut et al., 2003). PF0972 encodes a HMG-CoA synthase and PF0973 encodes an acetyl-CoA acetyltransferase, which function in pyruvate and acetyl-CoA metabolism.

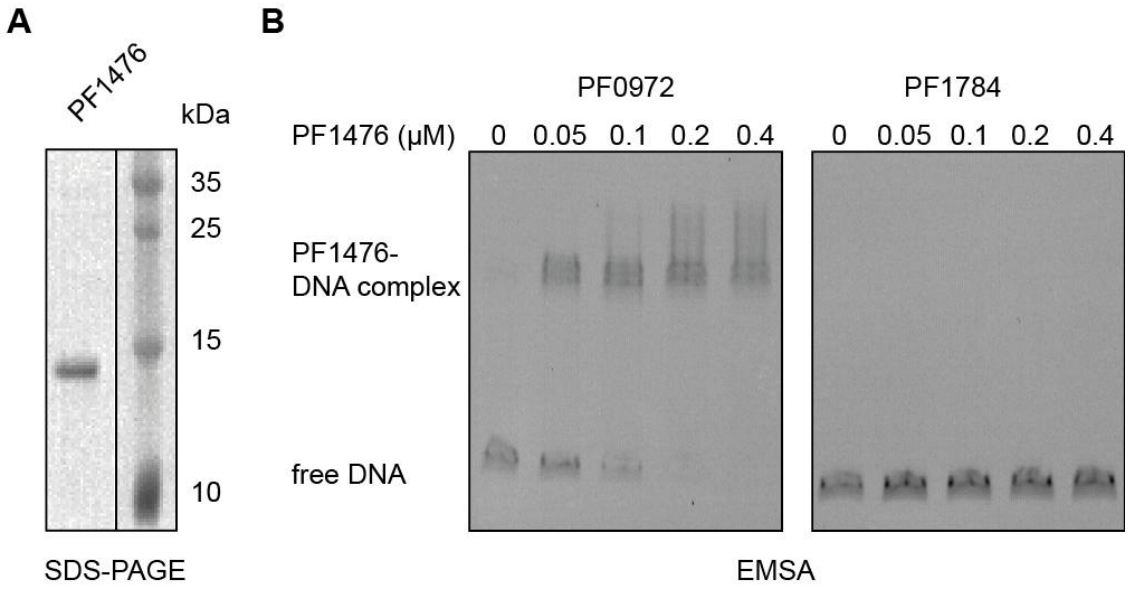


Figure 34. Preliminary analysis revealed binding of PF1476 to the promoter region of the PF0972 gene, which encodes a HMG-CoA synthase.

A, The corrected PF1476 ORF encodes a ~13 kDa protein. PF1476 was recombinantly expressed and purified. 500 ng of the protein were separated on a 15% denaturating SDS polyacrylamid gel. B, EMSA shows that PF1476 binds to the PF0972 promoter, but not to the PF1784 promoter. Cy5 labeled DNA templates were used and gelshifts were separated on a 6% non-denaturating polyacrylamid gel. Gelshifts using the PF0972 or the PF1784 promoter region as DNA template, are presented. PF1476 was used in the given concentrations.

These findings show that TrmBL1 is not only able to regulate expression through direct binding to the promoter regions of genes encoding e.g. gluconeogenic- and glycolytic-specific enzymes, but it also indirectly regulates expression of enzymes from functionally-linked metabolic pathways via transcriptional control of a secondary TF as shown here for the putative PadR-like regulator PF1476. So far, only preliminary characterization of PF1476 was carried out and its function as repressor or activator is still unclear. Thus, elucidation of the function of this PadR like regulator will be a straight forward task for the future to completely understand TrmBL1 mediated global gene regulation. The TrmB protein from *H. NRC1* can control transcription of up to four accessory regulators (Schmid et al., 2009).

Results obtained from ChIP-seq analysis were subsequently validated by *in vivo* (ChIP-qPCR), *in vitro* (EMSA and cell-free transcription) and *in silico* (comparison with a whole genome microarray analysis of *P. furiosus* grown on carbohydrates or peptides; Schut et al., 2003) analysis. This revealed approximately complete consistency of the considered results; however a few discrepancies were observed. The imperfect overlap of TrmBL1 binding sites found by ChIP-seq and differential regulation in the whole genome microarray analysis can be explained by various parameters. The most important one may be the slightly different growth conditions used in both experiments. Contrary to ChIP-seq *P. furiosus* was grown in the presence of S^0 for the microarray analysis. When maltose is used as the carbon source, *P. furiosus* growth is comparable both in the presence and absence of S^0; however significant changes in gene expression are evident from DNA microarray analyses (Adams et al., 2001; Schut et al., 2001; Schut et al., 2007). This may explain some of the found differences. Moreover, in the microarray analysis for induction of gluconeogenic growth peptides (hydrolyzed casein) were used and for induction of glycolytic

growth the disaccharide maltose (break down product of starch). In contrast, for the ChIP seq analysis *P. furiosus* was grown on sodium pyruvate to obtain gluconeogenic conditions and on starch to obtain glycolytic conditions. The differential transcriptional responses of *P. furiosus* cells after growth on various alpha glucans, e.g. maltose, starch, laminarin, glycogen or others, were already shown by two groups using whole genome microarray analysis *in vivo* (Lee et al., 2006a; Comfort et al., 2008). Therefore, effects caused by the differing carbon sources can not be excluded.

Furthermore, *in vitro* analysis revealed that for the *pfk* promoter (PF1784) the three sugars maltose, maltotriose and fructose can act as inducers for TrmBL1-mediated transcriptional repression. In contrast, for the MD-system promoter only maltose can act as inducer, but not maltotriose and fructose. Until now no compounds were found acting as co-repressors for TrmBL1 according to glucose for TrmB (Lee et al., 2007b; Surma, 2011). Taking these results together, it can be assumed that the function of TrmBL1 as global regulator is mainly based on both the presence of the TGM in promoter regions of regulated genes as well as the presence of various inducers e.g. maltose, maltotriose or others and possibly unknown co-repressors under varying growth conditions. This may also explain why not all genes with a predicted TGM in their promoter regions are differentially regulated in the whole microarray analysis, e.g. PF0043 (PEPS) and/or detected by the ChIP-seq analysis, e.g. PF0312 (GLK).

In conclusion, adaptation of an established ChIP protocol for ChIP-seq analysis of TrmBL1 enabled to extend the number of available *in vivo* methods for the hyperthermophilic euryarchaeon *P. furiosus* to study GRNs in more detail. Beside known and predicted TrmBL1 targets, novel TF binding sites were identified, which highlight the versatile functions of this global regulator in transcriptional control (summarized in Figure 35). Validation of the ChIP-seq-identified TrmBL1 binding sites by DNaseI footprinting revealed the high spatial resolution of the method. This is an important feature to study GRNs in Archaea, which have a compact genome regarding their protein-gene coding frequency (Koonin and Wolf, 2008). In the current study the high spatial resolution enabled discrimination between binding of TrmBL1 upstream or downstream of the promoter and prediction of the corresponding transcriptional response, activation or repression respectively. TrmBL1 ChIP-seq analysis showed that according to Schmid et al. (2009) transcriptional control of functionally-linked metabolic pathways in response to nutrient availability appears to be a common feature of TrmB family TFs in Archaea. Like the TrmB protein in *H. NRC1*, TrmBL1 is part of a multi-TF GRN. There it plays an important role for sensing disposability of a carbon source and controlling the appropriate transcriptional response by its own or together with additional regulators (Todor et al., 2013).

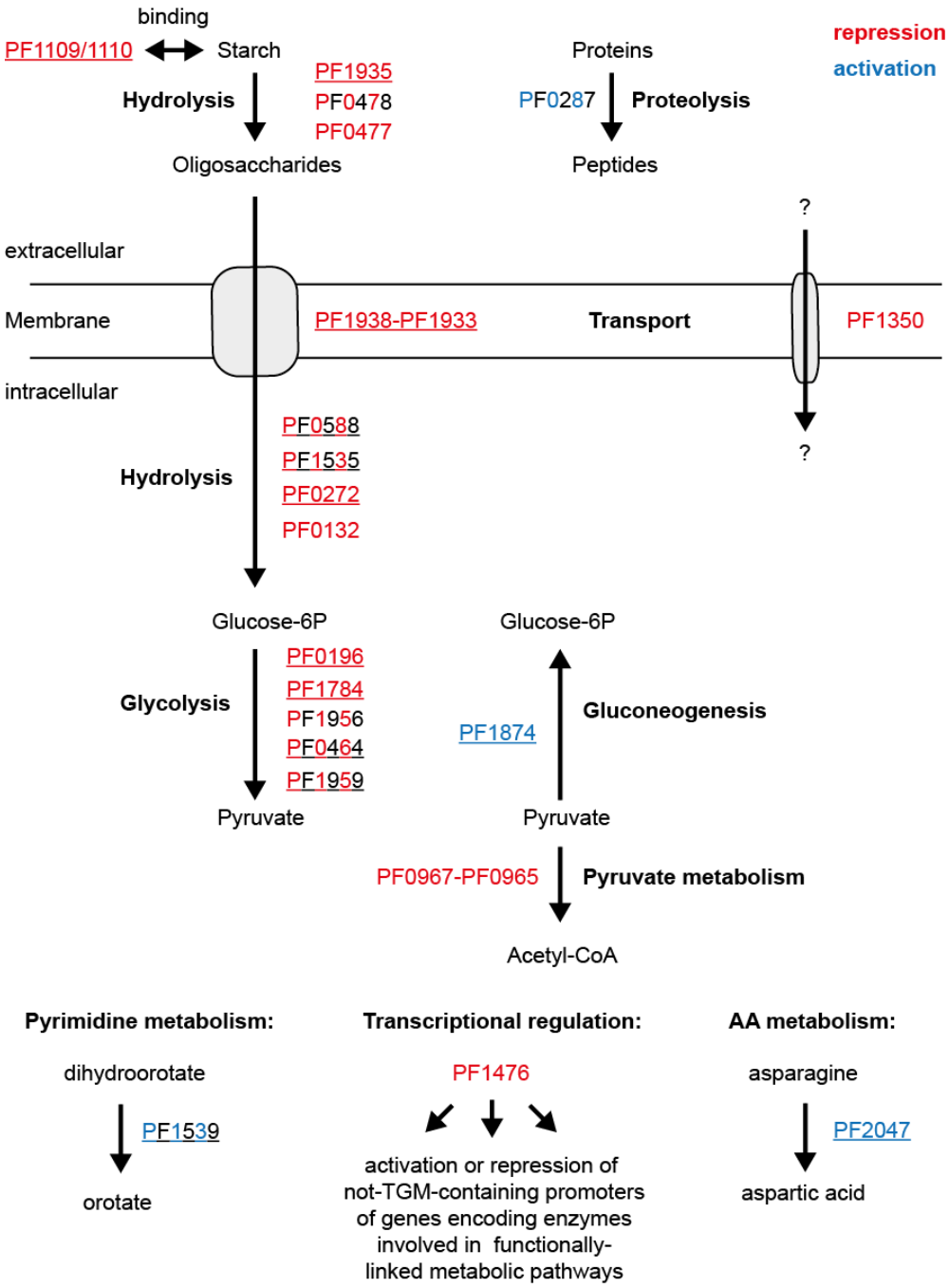


Figure 35. TrmBL1 controls expression of genes involved in functionally-linked metabolic pathways by its own or in concert with secondary regulators.

Genes: PF1109/1110, extracellular starch-binding protein; PF0477, proposed extracellular α-amylase. PF0478, proposed extracellular cyclomaltodextrin glucano-transferase; PF1935, amylo-pullulanase; PF0287, pyrolysin; PF1938-PF1933 maltodextrin-specific ABC transporter (MD-system); PF1350, proposed single-component transport protein; PF0588, phosphoglucose mutase; PF1535, maltodextrin phosphorylase; PF0272, proposed 4-α-glucanotransferase; PF0132, proposed α-glucosidase; PF0196, phosphoglucose isomerase; PF1784, phosphofructokinase; PF1956 phosphoglycerate mutase; PF0464, glycerinaldehyde 3-phosphate:ferredoxin oxidoreductase; PF1959, fructose-1,6-bisphosphate aldolase; PF1874 glycerinaldehyde 3-phosphate dehydroghenase; PF0967-PF0965, pyruvate ferredoxin oxidoreductase; PF1539, dihydroorotate dehydrogenase 1b; PF1476, putative PadR-like regulator; PF2047, l-asparaginase. TrmBL1 mediated repression or activaiton is depicted in red or blue. Complete colouring indicates validation of repression or activation by cell-free transcription. Banded colouring indicates suggested transcriptional effect. Genes differentially expressed in a microarray analysis of *P. furiosus* grown on maltose or peptides (Schut et al., 2003) are underlined.

4.3. TrmBL2 is an abundant chromosomal binding protein

It was speculated that also TrmBL2 functions in the sugar metabolism of *P. furiosus* due to its belonging to the TrmB family of transcriptional regulators. However, the C-terminal domain of TrmBL2 lacks the sugar binding part (Lee et al., 2008). Maruyama et al. (2011) showed that the TrmBL2 orthologue TK0471 from *T. kodakarensis* represents a novel type abundant chromosomal binding protein, which can additionally function as global repressor. ChIP-seq analysis verified abundant distribution of TrmBL2 binding sites over the whole genome of *P. furiosus*. Furthermore, the high consistency of detected TrmBL2 binding sites under glycolytic and gluconeogenic conditions disagrees with the assumption that TrmBL2 functions as transcriptional regulator of sugar metabolism like TrmBL1. This corresponds to the results obtained from a TrmBL2 deletion strain analysis (Waege, 2014). The ΔTrmBL2 strain showed no significant differences compared to the wildtype strain regarding its growth behavior using starch, pyruvate, maltose or cellobiose as solely carbon source. Moreover, only a small overlap of TrmBL1 and TrmBL2 binding events was found. Binding of both proteins in parallel or interchangeably was restricted to three promoter regions. Thus, it is unlikely that both TF regulates in concert transcription of sugar-metabolism-associated genes.

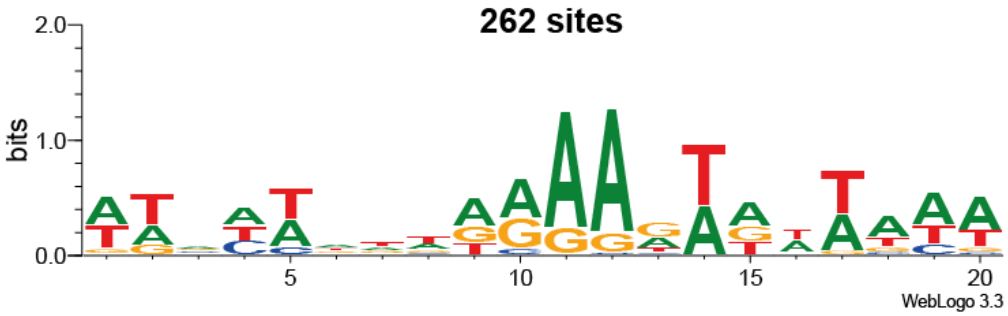


Figure 36. Potential TrmBL2 recognition consensus logo based on the 262 TrmBL2 binding sites identified in this study.

The motif was found by GLAM2 (Score: 2403) in 262 of 263 analyzed regions of *P. furiosus*, which were detected via ChIP-seq analysis of TrmBL2 under glycolytic and gluconeogenic growth conditions. The logo was created using WebLogo 3.3 (http://weblogo.threeplusone.com).

In contrast to the random localization of TK0471 in coding and non-coding regions of the genome, TrmBL2 binding events are enriched in non-coding areas compared with predicted random distribution. This difference likely derives from the higher resolution of the ChIP-seq approach. Furthermore, this indicates specific binding of TrmBL2 to particular areas of the genome. Nevertheless, according to TK0471 extensive search for a TrmBL2 recognition consensus motif based on the ChIP-seq results was not successful using numerous software tools. It was supposed that TK0471/TrmBL2 might bind to relatively nucleosome-free regions independent from any specific DNA recognition motif (Maruyama et al., 2011). Archaeal histone proteins prefer certain sequence motifs (nucleosome positioning signals) when binding to DNA (Bailey et al., 2000). Moreover, it was shown that the occurrence of nucleosomes is depleted in intergenic region, mainly immediately upstream of translation initiation codons (Nalabothula et al., 2013). Figure 36 shows one of the various motifs obtained from *de novo* motif discovery using 100 bp windows enclosing the TrmBL2 binding

sites identified by ChIP-seq. The motif illustrates the common feature of all potential TrmBL2 recognition consensus sequences identified by *de novo* discovery analysis. They all contained A/T-rich regions. This corresponds to the assumption that TrmBL2 especially binds to relatively nucleosome-free region. At the center of archaeal nucleosomes G/C-rich sequences predominate, whereas A/T-rich regions are underrepresented (Ammar et al., 2012; Nalabothula et al., 2013). It has to be proven yet, whether TrmBL2 recognizes A/T-rich regions with a higher affinity or whether this assumed preference is caused by unspecific binding to nucleosome-free regions.

The proposed function of TrmBL2 as global regulator suggests that it binds at least some promoter regions with higher affinity (Maruyama et al., 2011). TrmBL2 gene disruption analysis in *P. furiosus* revealed only rather small differential expression patterns compared to the wildtype strain (Waege, 2014). ChIP-seq demonstrated that TrmBL2 binds to the promoter regions of 77 % of most-high differentially expressed transcription units (single genes or operons). However, these 23 sites represent only 20 % of all TrmBL2 peaks located in intergenic regions. Furthermore, downregulation or upregulation did not correspond with binding of TrmBL2 downstream or upstream of the BRE and TATA-box. Thus, transcriptional control of TrmBL2 may rely on a mechanism differing from the so far described principles in Archaea. It was supposed that TrmBL2 blocks RNAPs access to the promoter regions through building of thick fibrous structures with the DNA (Maruyama et al., 2011). These structures may be composed of TrmBL2 and additional proteins, which are recruited by protein-protein interactions to the proximate chromosomal region according to a mechanism described for bacterial nucleoid-associated proteins like H-NS (Bouffartigues et al., 2007). Only auto repression of its own promoter could be shown *in vitro* by cell free transcription. Direct transcriptional repression of other analyzed promoter regions, which were bound by TrmBL2 in the ChIP-seq experiments, was not observed *in vitro* (Waege, 2014). This may derive from the absence of *in vivo* TrmBL2-interacting proteins in *in vitro* assays. However, no interaction partners could be identified so far.

In contrast, differential expression patterns observed by gene disruption of TrmBL2 may also rely on indirect transcriptional effects caused by secondary transcription factors. ChIP-seq showed TrmBL2 binding to the promoter region of the *tfb1* gene (PF1377). TFB1 functions as GTF in transcription initiation and it is supposed to be the `housekeeping´ TFB of *P. furiosus*. In the deletion strain analysis TFB1 was not differentially expressed; however, even small differences in expression of this GTF may lead to transcriptional responses in the cells. TFB1 from *P. furiosus* appears to be essential *in vivo* in contrast to TFB1 and TFB2 from *T. kodakarensis* (presence of one is sufficient) (data not shown for TFB1 from *P. furiosus*; Santangelo et al., 2007; Hidese et al., 2014). Additionally, TrmBL2 recognized the promoter regions of numerous predicted bacterial-like TFs in the *P. furiosus* genome (e.g. operon PF0648/PF0649). The functions of these TFs are still unclear and have to be characterized yet. This prohibits so far association of differential expression caused by TrmBL2 deletion and transcriptional regulation mediated by these secondary TFs. It was shown for the nucleoid-associated proteins H-NS and Fis from Bacteria that also indirect and proximal effects exist beside direct effects of protein binding on gene expression (Kahramanoglou et al., 2011; Wang et al., 2013). In conclusion, it can be suggested that the function of TrmBL2 as global regulator relies on both direct as well as indirect effects of protein binding on transcriptional control.

The physiological function of TrmBL2 as global regulator is still unclear. It is not essential for growth under optimal growth conditions using various carbon sources (Ingrid Waege 2014). Moreover, comparison of the TrmBL2/TK0471 gene disruption studies (whole genome

microarray studies; Maruyama et al., 2011; Waege, 2014) and binding sites mapping studies (sucrose gradient centrifugation combined with high-throughput sequencing (Maruyama et al., 2011) or ChIP-seq (this study)) from *T. kodakarensis* and *P. furiosus* revealed no striking overlap regarding differentially expressed genes and occupied genomic regions in contrast to Tgr and TrmBL1 (see above). Database search indicated that particular differentially expressed genes from *P. furiosus* likely encode membrane-associated proteins and transporters. However, these differences in expression between the ΔTrmBL2 strain and the wildtype strain can derive from the selection marker system used for deletion strain construction. The TrmBL2 gene was replaced by the resistance marker cassette for over expression of the HMG-CoA reductase gene (Waege, 2014). This leads to resistance of *P. furiosus* against the antibiotic simvastatin (Matsumi et al., 2007; Waege et al., 2010). The enzyme is essential for archaeal membrane lipid biosynthesis (Cabrera et al., 1986; Bochar et al., 1997). Thus, effects on genes encoding membrane-associated proteins cannot be completely excluded using the simvastatin resistance marker system.

It was an unexpected finding that the number of TrmBL2 binding sites increased in the *P. furiosus* genome and that the binding intensities altered upon heat shock (107°C for 30 min). Assuming that the TrmBL2 occupied regions span approximately 28 bp according to TrmBL1, this means that after heat shock ~12 % of the whole *P. furiosus* genome is occupied by TrmBL2 in contrast to ~4 % under optimal growth conditions (95°C). Moreover, TrmBL2 binding sites are randomly distributed over the whole chromosome after heat shock. This indicates that TrmBL2 likely does not function as key transcriptional regulator at elevated temperatures. However, the TrmBL2 binding site in its own promoter region, which overlaps the TSS, is shifted upstream, which suggests increased accessibility for RNAPs and activation of transcription. This assumption is encouraged by the observation that TrmBL2 protein levels were increased (2 - 4 fold compared to TBP and GDH) after incubation at 105°C for 30 min (Waege, 2014). The heat shock response of *P. furiosus* is at least in parts controlled by the repressor Phr (Vierke et al., 2003; Keese et al., 2010). The Phr regulon comprises several genes involved in the heat shock response and in the promoter region of theses genes a Phr DNA recognition sequence was identified (Vierke et al., 2003; Kanai et al., 2010; Keese et al., 2010). Important genes for the heat shock response encode e.g. an ATPase from the AAA^+ family of ATPases (AAA^+ ATPase; PF1882), the small heat shock protein 20 (HSP20, PF1883) and a myo-inositol-phosphate synthase (PF1616). These genes are upregulated upon heat shock through dissociation of Phr from the promoter regions (Shockley et al., 2003; Liu et al., 2007). HSP20 functions mainly as holding chaperone preventing denaturation and aggregation of proteins at elevated temperatures (Laksanalamai et al., 2001; Techtmann and Robb, 2010). The enzyme myo-inositol-phosphate synthase is involved in the synthesis of the compatible solute di-myo-inositol-1,1′-phosphate, which was shown to accumulate to high molar concentrations in the cytoplasm at suboptimal growth temperatures and acts as thermostabilizer (Scholz et al., 1992). ChIP-seq revealed no striking association between Phr-regulated genes and genome-wide binding of TrmBL2. Moreover, comparison of the TrmBL2 heat shock ChIP-seq results with a whole genome microarray analysis of *P. furiosus* grown under optimal conditions (90°C) and heat shock (105°C) was not feasible because of the unavailability of the complete microarray dataset at the denoted website (Shockley et al., 2003).

Additionally, ChIP-seq demonstrated an altered TrmBL2 binding pattern in the promoter region of the *tfb2* gene (PF0687), whereas the binding site in the intergenic region located upstream of the *tfb1* gene (PF1377) remained consistent. TFB2 is suggested to function as alternative GTF in transcription initiation at elevated temperatures (Shockley et al., 2003;

Micorescu et al., 2008). Comparable regulation mechanisms underlying the heat shock response were described for Bacteria as well as Archaea. In Bacteria numerous heat-shock-associated genes are mainly regulated by alternative σ-factors (Nonaka et al., 2006; Hecker et al., 2007). Furthermore, in halophilic Archaea, which usually contain multiple copies of genes encoding TBP and TFB, it was assumed that the properties of the core promoter elements and adjacent regions determine the heat-responsibility of specific genes. These heat shock inducible promoters are likely recognized by specific TBPs and/or TFBs couples (Thompson and Daniels, 1998; Lu et al., 2008). So far the function of TFB2 in the heat shock response of *P. furiosus* is only proposed and has to be confirmed *in vivo* and *in vitro* yet.

Beside transcriptional control, differential TrmBL2 binding patterns observed upon heat shock may derive from its function as architectural chromosomal binding protein. For the chromatin protein Alba2 from *A. pernix* it was shown that it is able to stabilize especially A/T-rich DNA at elevated temperatures *in vitro* (Črnigoj et al., 2013). A corresponding function can also be suggested for TrmBL2. The TrmBL2 protein itself is heat stabile even at a temperature of 110°C, whereas most other cellular proteins are denaturated at temperatures above 100 °C (Waege, 2014). Thus, TrmBL2 may function as thermo-protective chromosomal protein of A/T-rich genomic regions, which are not bound through the additional nucleoid-associated proteins (histone proteins and Alba) present in *P. furiosus*. This is also encouraged by the finding that upon heat shock the number of clustering TrmBL2 peaks (second binding site located within 1000 bp) increased. The average distance between two TrmBL2 sites is preserved and ranges from ~300 bp to ~800 bp under all conditions. Furthermore, the ΔTrmBL2 strain was more sensitive to elevated temperatures compared to the wildtype strain (Waege, 2014). However, it is not trivial to observe a striking growth phenotype, which appears sensitive to heat shock conditions. Even the Phr-deletion strain from *T. kodakarensis* showed no differences in growth characteristics compared to the wildtype strain at elevated temperatures (Kanai et al., 2010). In conclusion, TrmBL2 ChIP-seq analysis of *P. furiosus* grown at 107°C suggests an unexpected role of the TF during heat shock response; however the exact function (regulatory and/or architectural) has to be characterized in detail yet.

Recapitulating, extending of the ChIP-seq approach to study TrmBL2 has revealed its numerous potential functions as architectural as well as regulatory DNA binding protein. ChIP-seq demonstrated abundant binding of TrmBL2 to the genome of *P. furiosus* according to TK0471 from *T. kodakarensis*. The distribution of TrmBL2 differed from the random distribution detected for TK0471. Moreover, TrmBL2 appears to bind DNA independent from a prominent recognition consensus sequence. However, preferential binding to A/T-rich genomic sequences can be suggested. The function of TrmBL2 as global repressor remained unclear. Detected differential expression pattern obtained from TrmBL2 gene disruption analysis can rely on direct as well as indirect effects of protein binding on gene expression. Furthermore, successful adaptation of the ChIP-seq approach for analysis of cells, which were formaldehyde-treated after incubation at 107°C for 30 min, indicates a novel unexpected role of TrmBL2 in the heat shock response of *P. furiosus*. So far, these conclusions are mainly based *on vivo* experiments. According to TrmBL1 they have to be subsequently validated *in vitro* by e.g. EMSA and cell-free transcription experiments.

4.4. TFB-RF1 exclusively binds to its designated binding site *in vivo*

The transcriptional regulator TFB-RF1 from *P. furiosus* represents the first described archaeal activator, which functions through recruitment of TFB to an imperfect BRE. The mechanism of activation was exclusively elucidated *in vitro*. In contrast, the physiological role of TFB-RF1 *in vivo* remained unclear (Ochs, 2009; Ochs et al., 2012). TFB-RF1 (PF1088) is co-transcribed with PF1089 and *in vivo* expression of the PF1088/PF1089 could not be observed after growth of *P. furiosus* under various growth conditions (Kreuzer, 2009; Yoon et al., 2011; Rost, 2012). To study the role TFB-RF1 *in vivo*, over expression of the TF was achieved by transformation of the vector pYS5a containing an extra copy of the PF1088 coding region under the control of a gluconeogenic promoter (*fbpase* promoter). The resulting *P. furiosus* strain MURPf5 specifically expressed TFB-RF1 after growth under gluconeogenic conditions (Kreuzer, 2009). This strain was used in ChIP experiments for genome wide binding sites mapping of TFB-RF1 in vivo. ChIP-qPCR and ChIP-seq showed specific recognition by TFB-RF1 of its *in vitro* designated binding site in the promoter region of PF1089 *in vivo*. Furthermore, ChIP-seq confirmed binding of TFB-RF1 directly upstream of BRE. Beside the PF1089 promoter no additional binding sites were detected in the genome of *P. furiosus*. This indicates that the mechanism of activation by recruitment of TFB1 is restricted to the operon PF1089/PF1088 representing a unique type of autoregulation.

Nevertheless, additional TFB-RF1 binding sites cannot be excluded through our ChIP-seq analysis. Over expression of TFB-RF1 in *P. furiosus* strain MURPf5 was accomplished by using the promoter of the *fbpase* gene, which led to a moderate over expression of the transcriptional regulator under gluconeogenic conditions (Kreuzer, 2009; Rost, 2012). All attempts to express TFB-RF1 constitutively under the control of the strong *gdh* promoter were not successful (Rost, 2012). Therefore, exclusive binding of TFB-RF1 to its designated binding site *in vivo* may be observed due to the limited TFB-RF1 protein level in cells.

Moreover, using the *fbpase* promoter only allows ChIP-seq analysis under gluconeogenic growth conditions. Database analysis of PF1088 using the conserved-domain-database (CDD, http://www.ncbi.nlm.nih.gov/) indicates that TFB-RF1 belongs to the ArsR-subfamily of HTH transcription regulatory proteins with winged helix topology (cd_00090). This was already mentioned by Okada et al. (2006) and Simon Ochs (2009). Regulators of this subfamily have the following conserved features: A, a putative DNA binding site; B, a putative dimerisation domain and C, a putative ion binding site. DNA binding of TFB-RF1 has been shown (Ochs et al., 2012). Furthermore, there exist various findings indicating dimerisation of TFB-RF1 according to its homologue PH1061 (Okada et al., 2006; Rost, 2012; Winter, 2012). One well studied member of the ArsR-subfamily is the cyanobacterial metallothionein repressor SmtB from *S. PCC7942* (Morby et al., 1993; Cook et al., 1998). This repressor was shown to bind to its operator region in a zinc (metal-ion) dependent manner (Morby et al., 1993; Turner et al., 1996; VanZile et al., 2000). Currently, no such co-factors, acting as co-repressors or inducers, has been described for TFB-RF1. Thus, it cannot be excluded that binding of TFB-RF1 to additional binding sites is prohibited by the presence or absence of such a co-factor under gluconeogenic growth condition.

Additionally, monomeric or dimeric binding of TFB-RF1 to DNA is controversially discussed since its discovery as activator mediating recruitment of TFB1 (Ochs, 2009): On the one hand the dimensions of the shift in the EMSA and results from hydroxyl radical footprints indicate monomeric binding. On the other hand the extent of the DNase I footprint of approximately 20 bp on both strands and the dimeric structure of PH1061 in the crystal and

in solution point to dimeric binding. In the western blot analysis using crude extracts derived from non-treated or formaldehyde-treated *P. furiosus* cells over expressing TFB-RF1, the antibody detected mainly a signal in the range of 25 kDa, which may represent dimeric TFB-RF1 (Rost (2012) and this study). This suggests high stability of the dimeric form even under standard denaturation conditions. Only reversal of the crosslink overnight and extensive denaturation lead to disappearance of the 25 kDa signal indicating break of the dimeric form. It is possible that the dimeric conformation of TFB-RF1 is the closed conformation, which cannot bind DNA (Okada et al., 2006). Presence of a specific co-factor may represent one mechanism to control regulatory activity of TFB-RF1 as monomer or dimer according to the tetrameric-hepta/octameric conformatial change of TrmBL1 in the absence and presence of maltotriose and maltose (Lee et al., 2007b).

In summary, this study showed that combining over expression of a transcriptional regulator, here TFB-RF1, with ChIP-seq for genome wide binding sites mapping represents a powerful tool to analyze GRNs of weakly expressed TFs in *P. furiosus in vivo*. The *in vitro* designated TFB-RF1 binding site in the PF1089 promoter region was confirmed *in vivo*. Moreover, due to the high spatial resolution of the ChIP-seq approach, binding of TFB-RF1 upstream of the BRE could be proved. No additional binding sites for TFB-RF1 in the genome of *P. furiosus* were found, which helps to associate the regulator with a specific metabolic pathway. Thus, the physiological function of this unusual archaeal activator, which is supposed to mediate recruitment of TFB1 to an imperfect BRE, remained unclear.

4.5. ChIP-seq workflow for *P. furiosus* – Conclusions and outlook

While ChIP-seq became one widely used approach in bacterial and eukaryotic model organisms, it has been hitherto rarely applied in archaeal systems. Only two published studies can be found by literature search (Mai 2014). Wojtas et al. (2012) used ChIP-seq to study the general archaeal transcription machinery (RNAP) of *S. acidocaldarius*. Wilbanks et al. (2012) presented a complete epitope-tag-based ChIP-seq workflow for *H. NRC1* suitable for whole-genome TF binding sites mapping of GTFs as well as bacterial-like TFs. In all additional ChIP analyses in archaeal species, ChIP-chip was used for identification of binding sites in a genome-wide manner (e.g. Duc et al., 2012; Sharma et al., 2012; Song et al., 2013)

Studying the genome-wide occupancy of three archaeal TFs enabled the establishment of a complete ChIP-seq workflow for the hyperthermophilic euryarchaeon *P. furiosus*. This extended the number of available *in vivo* approaches to analyze GRNs in *P. furiosus*, which represents beside *T. kodakarensis* one widely used archaeal model organism within the order *Thermococcales* (Waege et al., 2010; Sommer et al., 2014). In concert with the well-established *in vitro* assays of the *Pyrococcus furiosus* transcription system, it represents a powerful tool, which can not only help to understand regulation of gene expression, but also to dissect the underlying mechanisms.

One limitation of the here described approach is, that it relies on the availability of a bio-fermenter (~15 L volume) for generation of formaldehyde-treated cells. Additionally, bio-fermenter-scale cultivations especially restrain the number of performable biological replicates. Downscaling to 2.5 L or even less would be preferable. In contrast, crosslinking of 1 ml cell culture volumes, which were described to be sufficient for ChIP-seq analysis in *H. NRC1*, appears to be absolutely unapproachable (Wilbanks et al., 2012). Furthermore, the ChIP-seq experiments were carried out using cells adapted to steady-state conditions. Todor et al. (2013) showed that the transcriptional response mediated by the TrmB protein from *H. NRC1* starts within 2 min after addition of glucose to the media. Thus, adaptation of the protocol for dynamic studies has to be considered. This can particularly contribute to a better understanding of the novel unexpected role of TrmBL2 in heat shock response of *P. furiosus*, which was shown to be a highly dynamic process based on the combined effects of thermal stress and reduced cell growth (Shockley et al., 2003).

In the presented ChIP-seq experiments, only bacterial-like TFs of *P. furiosus* were examined. The group of Prof. Dr. Michael Thomm also focuses on the general transcription machinery composed of GTFs and the RNAP. The first promising results concerning the occupancy of GTFs (TBP and TFB) as well as the RNAP on specific genes by ChIP-qPCR under varying conditions were obtained by Schramm (2013). However, optimizations are still in progress. In addition, first ChIP-seq results received for TFB and RNAP (data not shown) suggest that genome-wide occupancy analyses will be applicable. Challenges regarding high background signals may be overcome by applying a recently described alternative ChIP-seq protocol termed as ChIP-exo. Applying lambda exonuclease to chromatin immunoprecipitates allowed precise examination of the location of thousands of transcription initiation complexes in the genome of *Saccharomyces sp.* (Rhee and Pugh, 2011, 2012). Furthermore, collaboration with the group of Prof. Dr. Finn Werner from the University College London was started to analyze archaeal elongation complexes (especially RNAP and the Spt4/5 complex) *in vivo* by ChIP-seq. Dr. Katherine L. Smollet spent two months in Regensburg for the generation of formaldehyde-treated *M. jannaschii* cells and was taught the whole ChIP-seq workflow.

The recent development of a genetic system for *P. furiosus* will also expand the possibilities for advanced ChIP-seq studies. One example is presented in this thesis by combining ChIP-seq and *in vivo* over expression of TFB-RF1 in *P. furiosus*. In addition, genetic manipulation of *P. furiosus* enables epitope-tagging of the protein-of-interest (Waege et al., 2010). So far, no successful epitope-tag-based IPs could be conducted and the system has to be improved. However, these experiments will contribute to analyze proteins lacking the disposability of highly-specific polyclonal antibodies like TFB1 and TFB2 from *P. furiosus.* Facciotti et al. (2007) used epitope-tag-based ChIP-chip to study the genome-wide distribution of the numerous GTFs types from *H. NRC1*. Moreover, it will also facilitate to study the proposed cross-regulation network of TrmB and TrmBL1 in transcriptional control of sugar metabolism in *P. furiosus*. In the current study TrmB could not be analyzed by ChIP because of cross-specificity of the antibodies as well as deletion of the TF in the genome. Thus, combination of *in vivo* over expression and epitope-tag-based ChIP-seq analysis has to be accomplished. Combining gene disruption and ChIP analysis enables to study direct as well as proximal effects of protein binding on gene expression (Schmid et al., 2009; Kahramanoglou et al., 2011). For TrmBL2 this combination revealed numerous potential functions as architectural chromosomal binding protein as well as transcriptional regulator. The underlying mechanisms have to be examined in detail yet. Another interesting candidate for combined analysis using both approaches represents the PadR-like regulator PF1476. Expression of PF1476 is controlled by TrmBL1 and preliminary analyses demonstrated that this TF binds at least one promoter region (PF0972) in the genome of *P. furiosus*. Studying the function of PF1476 in a genome-wide manner will not only contribute to a better understanding of the physiological role of the TF itself, but it will also help to dissect the complete GRN of TrmBL1. In contrast to TrmBL2, no architectural function of PF1476 can be expected and evaluation of results should be more feasible.

Additional optimizations of the described ChIP-seq workflow and expanding the studies of genome-wide TF binding site mapping of archaeal TFs in *P. furiosus* will contribute to a better understanding of the diverse underlying mechanisms of transcriptional control mediated by GRNs to achieve optimal growth of the cells in response to differing environmental as well as intracellular conditions. Furthermore, transferring the approach for analyses of the general transcription machinery will help to elucidate basal transcription mechanisms as well as the function of associated proteins, factors and complexes in all steps of the archaeal transcription cycle *in vivo*.

5. Supplementary Data

5.1. Occurrences of TrmB family proteins within 13 *Thermococcales* species

MSA of TrmB_domains from 87 TrmB/TrmBL proteins.

The result from the MSA (T-coffee) is presented in the FASTA format (primary accession number/ AA residues/ ordered locus name).

>F4HJC7/28-86/PNA2_1508
----------GRIYGI---LLLSN-KPLTISELASITGLSRSSVSIALSKLSRDYLV-TYTKEK----K-T-------KY----FSA--VP-A--
>F4HKR5/5-63/PNA2_1760
----------EFVYKVL--AT-KK-KAITLKKLSSELETPMPRLLQTLKKLEDDGLVEII-Y---N-KE-IT--------V---RAK--TM---N
>F4HKY3/14-86/PNA2_0572
--------LRAIGLKKNEIMIYRL---LVEKK-RGMRIKEIQKELGISERSVRTHVLNLYRKGLL-KRE-LI-QRGWLG-------YIY---TAV--SPG---
>F4HIY0/7-60/PNA2_1437
I----------K---IL-EN-GEKSEEEIVNEIKLSRLEVRRILLRLAEQGKV-ES-F-LRN-GK-IY----------WKI--RE---K
>F4HIY4/10-78/PNA2_1441
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-MNQ-PG----K-T-------NKY---RPV--HP-A--
>H3ZM44/1286-1350/OCC_06866
E--------MEKGLTEREAEILQE---I--ST-EWTSREEIQKKVGLSRARLNQLLKNLEEKGYV-ERRM---E-GK-RQF-------V---R-K-------L
>H3ZP63/11-78/OCC_03542
L--------SEIGFTKYEILTYWT---LL-VY-GPSTAKEISTKSGIPYNRVYDTISSLKLRGFV-TE-I---E-GT-PK-------VY---AAY--SP---R
>H3ZKI7/10-77/OCC_09344
L--------QKLGLTKYESLAYIT---LL-KL-GISKATDLTKESGIPHTRIYDVLSSLHRKGFV-DI-M---H-GT-PRM-------Y---KPV--NP---E
>H3ZM74/211-278/OCC_07016
L--------ESFNLNEDEKRALLY---IL-EKGGRASQAEVRTALGIPKTTAWRMFKRLERQGLV-RII-K---GRK-ENWV----------EL--K----P
>H3ZQM3/10-77/OCC_02652
L--------KRLGLKDYEARVYAA---LA-LL-GPSKASEVARESGVPRPKVYEVLKELHRKGFV-DF-S---E-GK-PAF-------F---RAV--EP---E
>H3ZMM5/12-82/OCC_02477
--------LEKFELTSSEIKIYSL---LL--K-EQLTPRQIAKKLDLSERIVREKLKHLLELGLV-ERE-MV-NRGWLG-------YIY---KAK--APK---
>H3ZLE5/10-78/OCC_09913
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-LSQ-PG----K-A-------NKY---RPV--HP-K--
>H3ZKD8/11-84/OCC_09099
-------LLRTLGFKKNELRIYRL---LLEKK-APMRITEIKEELGISERSVREHVLNLYKRGVL-KRE-LI-QRGWLG-------YVY---SAV--SPS---
>H3ZMW3/10-77/OCC_05836
L-------KE-LGLNKYEVNAYLT---LI-KQ-GPLTAGELASLSKVPQPRIYDVVRSLMSKGFV-AIT-SE----R-P-------KKI---VPL--DP-E--
>B7R0N8/14-87/TAM4_639
-------LLKSLGFKKTELRVYRL---LLEKK-RPMRITEIQRELGLSERSVREHVLSLYRKGVL-KRT-LI-EQGWLG-------YVY---TAA--SPK---
>B7R1X6/10-78/TAM4_1705
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-Q--
>B7QZP5/10-77/TAM4_771
L--------QRLGLTKYESVAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DV-M---Q-GT-PRL-------Y---KPV--NP---E
>B7R0B2/18-84/TAM4_1006
----------WEYSHVEGRIYAL---LLLSG-EPMTISELVEETGFSRSAVSTALSRLQRDYLV-EVQKKG----R-T-------KY----FTA--IPAL--
>Q5JFH6/231-302/TK0140
L--------SE----LRELKTYFSILSAM-AS-GRRKPSEIASGAGLDGRKIYPYLETLMRLGFV-ER-ELPVARKEKRG-------LY---RIS--DP---M
>Q5JJ60/10-77/TK1769

L--------QRLGLTKYESLAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DV-M---Q-GT-PRL-------Y---KPV--NP---E
>Q5JD67/10-78/TK0471
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-E--
>Q5JES0/5-85/TK1913
VYERLEALLKSLGVKKTELRIYRL---LLEKK-RPMRVTEIVNELGISERSVREHVLSLYRKGML-RRE-LI-QQGWLG-------YTY---TAV--SPH---
>Q5JI96/8-65/TK0956
----------LE---LI-KA-GVTNEREIAERLGLSSKEVEDIIKILESLGYV-EK-V---EFGSKAC-------ET---CPLKKVC---Y
>C5A748/10-77/TGAM_1558
L-------KE-FGLNEYEVRAYLT---LI-KN-GALTAGELATLSKVPQPRIYDVIRTLMAKGFV-TTS-QS----R-P-------KQV---IPL--SP-D--
>C5A4D8/10-77/TGAM_0598
L--------QRLGLTKYESVAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DV-M---Q-GT-PRL-------Y---KPV--NP---E
>C5A539/20-88/TGAM_0849
M--------VRWGYTATDGKVYAI---LLLSN-RPMTIAELAKETGLSRSSISVALSRLVREYLV-TCRREG----R-T-------KY----FTA--VP-A--
>C5A5I3/14-87/TGAM_0993
-------LLRSLGFKKTELRVYKL---LLEKK-RPMRITEIQRELGLSERSVREHVLNLYRKGVL-KRT-LI-EQGWLG-------YVY---TAA--SPK---
>C5A588/12-82/TGAM_0898
--------LRVFELGESEIKIYSL---LQ--R-EALTPRQIAKTLGLSERIVREKLKHLLELGLV-ERT-LV-NRGWLG-------YVY---YAK--APR---
>C5A6Z8/10-70/TGAM_1508
----------REKIYSV---MF-DR-EGLRIPEIARITGLNYTYTYQVVRSLVREGYV-TVV-GG----R-V------VFL-DRKGF--LF-K--
>C5A7G8/10-78/TGAM_1678
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-Q--
>Q9V2L9/14-86/PAB2311
--------LKAIGLKKNEIRIYRL---LVEKR-RGMRIREIQRELGISERSVRTHVLNLYRRGLL-KRE-LV-QSGWLG-------YIY---TPV--SPL---
>Q9UYS2/265-329/PAB0953
----------QFSEKEIEVLRA---II-ENGGEIKQEELPKIVGYSRPTISRIIQDLEKKGIV-ERE-K---SGKTFIVRV----------I---K---K
>Q9UZ85/15-83/PAB0838
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-MNQ-PG----K-T-------NKY---RPV--HP-A--
>Q9UYX1/234-303/PAB2383
----------E----LRELKTYFSILSAT-AS-GKTRASEIASYAGLEGRKIYPYLDTLVRLGFI-ER-EIPIARNDKRG-------IY---KLK--DP---M
>Q9V2N2/3-60/PAB2321
----------IELLKK---LA-RKK-IFTIEEVAEISGVKKSSLRVLLSRLEKRGLI-ERI-E---RGK-YVVIPLEAEVY---AAY--SP---R
>C6A540/10-77/TSIB_1684
L-------KE-LGLNKYEVNAYFT---LI-KQ-GPLTAGELASLSKVPQPRIYDVVRSLMSKGFV-AVT-SE----R-P-------KKI---VPL--DP-K--
>C6A2R8/12-82/TSIB_0852
--------IKNFELTHSEIKIYSL---LL--K-EELTPRQIAKKVGLSERIVREKLKHLLELGLI-ERN-LI-KQGWLG-------YIY---RAK--NPK---
>C6A583/27-92/TSIB_1727
L--------EELGVTFLEFKALVH---L---KS-AKTQKDLLKEMNVSKSTASKVLSSLERKGIV-KRE-R---RGK-AYIVEL----------T---D---K
>C6A5H5/12-84/TSIB_1819
--------LKNLGFKKNELRIYQL---LLEKN-SAMHITEIKEELGISERSVREHVLNLYRRGVL-KRE-LI-QKGWLG-------YVY---SAV--SPA---
>C6A5D5/10-77/TSIB_1779
L--------QKLGLTKYESLAYIT---LL-KL-GTSKATDLTKESGIPHTRIYDVLSSLHRKGFV-DI-M---H-GT-PRM-------Y---KPV--NP---E
>C6A0Q7/214-278/TSIB_0134
L--------EFFNLNEDERRALLY---IL-EKGGRASQAEVRTALEIPKTTAWRMFKRLEKQGLV-KII-K---GRK-ENWV----------E
>C6A1J5/13-81/TSIB_0424
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-LSQ-PG----K-A-------NKY---RPV--HP-K--
>C6A1D3/11-78/TSIB_0362
L--------SEIGFTKYEILTYWT---LL-VH-GPSTAKEISTKSGIPYNRVYDTISSLKIRGFV-TE-I---D-GT-PR-------VY---AAY--SP---R
>F0LL20/5-65/TERMP_01952

----------EYIYKLL--VV-KK-RAVTLQKLSEELETPMPKLLQTLKSLESDGLV-EV-F---Y-GK-EKAA-----IM---VKA--KT---L
>F0LJS0/12-84/TERMP_01737
--------LRSLGIKKTEIRIYRL---LLEKK-VPMRITEIQKELGISERSVREHVLNLYRKGLL-KRE-LI-QKGWLG-------YVY---TAT--SPA---
>F0LKH6/10-77/TERMP_01835
L-------RE-LGFSEYEVSAYLT---LI-KE-GPLTAGELATLSKVPQPRIYDVIRTLMGKGFV-ITI-GG----R-P-------KKV---VAV--NP-E--
>F0LL88/10-78/TERMP_00762
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-K--
>F0LL63/246-309/TERMP_00737
L--------KEYELTKDEEKALLY---LF-DRGGKAKQAEVREMLGIPKTTAWRMFQRLEKQGLV-RVY-K---KKR-ENW----------V
>F0LL87/6-62/TERMP_00761
V----------L-N---IL-KG-KELTAEEISREVRISHPEVRRILLRLAEQGKV-ES-L-QKE-GK-IL----------WRL--KQ-RTR
>F0LL39/11-78/TERMP_01972
L--------TRLGFTKYEILTYWT---LL-VY-GPSTAKEISERSGIPYNRVYDTIASLKTRRFV-SE-I---D-GS-PR-------VY---VAF--SP---S
>F0LK26/28-91/TERMP_01763
A--------YKVELTELERDVLEF---IL-QNG-EVTQKGLSNKFG--KVRACRVIQNLEKKGLI-ERK-K---KGR-TYVIRV----------I
>F0LII0/12-78/TERMP_00277
L--------QELGLTKREAEVYLT---IW-TK-NGATVKELLDSLDVHQPQLYNIIQSLIRKGFV-KA-S---A-GR-PRI-------Y---TAT--D----I
>F0LJA5/10-77/TERMP_01652
L--------QKLGLTKYESLAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DI-M---H-GT-PRL-------Y---KPV--NP---E
>F0LHI9/12-82/TERMP_01341
--------LEKFKLTRSEIRIYSL---LL--K-EQLTPRQIAKKLGLSERIVREKLQHLLELGLV-ERE-LI-SRGWIG-------YLY---KAK--APN---
>F0LHT5/18-87/TERMP_01360
M-------MIRWGYTHTDGKVYAL---LLLNE-KPLTINELVQLTGLSRSSVSTSLNKLARDYLV-IVRKNG----K-T-------KL----FYP--IP-A--
>F0LL61/209-274/TERMP_00735
M--------EKFDLNDEEKSALLY---IL-EKGGRASQAEVRNALGLPKTTAWRMFKRLEKQGLV-KII-R---GRK-ENWV----------E--------L
>O58555/18-87/PH0825
L-------MERWGYDRTGGKIYAI---LLLSS-KPLTISELTKITGLSRSSVSMALSKLSREYLV-TYTKEK----K-T-------KY----FYA--VP-A--
>O58763/11-78/PH1034
L--------IELGFTKYEVLTYWT---LL-VY-GPSTAKEISIRSGIPYNRVYDTISSLKVRGFV-TE-I---E-GN-PK-------VY---AAY--SP---R
>O58529/13-81/PH0799
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-MNQ-PG----K-T-------NKY---RPV--HP-A--
>O58144/119-183/PH0407
L-------------------ALLA---VL-EN-IPLTSRELYVKSNLPRSTFYHVINNLSKYGLI-GK-R---N-GK-YFLI----EKYELFHIFAKEI---Y
>O58509/11-78/PH0751
L--------QELGLTKREAEVYLA---IL-MK-NGATVKDLLETLDVHQPQLYNIIQSLIRKGFI-RA-S---A-GR-PRV-------Y---TAN--DI---T
>O58571/97-168/PH0840m
L--------SE----LKELKTYFSILSAM-SS-GRRRPSEIASEVGLEGRKIYPYLETLIRLGFV-ER-ELPIARKEKRG-------LY---RIS--DP---M
>O57817/14-86/PH0046
--------LRTMGLKKNEIMIYNL---LVEKK-RGMRIREIQRELGISERSVRAHVLSLYRKGLL-KRE-LV-QKGWLG-------YIY---TAV--SPL---
>B6YUM9/250-313/TON_0578
L--------TEYDLTKDEEKALLY---LF-DRGGKARQSEVRETLGIPKTTAWRMFQRLEKQGLV-RVY-K---KKR-ENW----------V
>B6YV59/10-77/TON_1797
L--------QKLGLTKYESLAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DV-M---H-GA-PRL-------Y---KPV--NP---E
>B6YW45/11-84/TON_1920
-------LLRSLGLKRTELRIYRL---LLEKS-EPMRITEIQRELGISERSVREHVLSLYRKGIL-RRR-LI-EQGWLG-------YVY---TAV--TPS---
>B6YTD0/10-78/TON_0332
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-Q--
>G0HJP5/10-77/GQS_06200

L--------QKLGLTKYESLAYIT---LL-KL-GPSKATDITKESGIPHTRVYDVLSSLHRKGFV-DV-M---Q-GS-PRL-------Y---KPV--NP---E
>G0HNN3/10-77/GQS_09575
L-------KE-FGLNEYEVRAYLT---LI-KN-GALTAGELATLSKVPQPRIYDVIRTLMAKGFV-TTS-QG----R-P-------KQV---IPL--SP-D--
>G0HLU7/212-277/GQS_02185
L--------ENLDLNEEEKRALLY---IF-DKGGKASQAEVREAIGLPKTTAWRMFKRLERKGLV-RIL-K---GKK-ENWV----------E--------L
>G0HJ81/11-84/GQS_05545
-------LLRSLGVKKTELRIYRL---LLDKK-EPMRITEIQRELGISERSVREHVLSLYRKGIL-RRT-LI-EQGWLG-------YTY---SAV--SPS---
>G0HMY8/12-82/GQS_09130
--------LRVFELGESEIKIYSL---LQ--R-ESLTPRQIAKTLGLSERIVREKLKHLLELGLV-ERT-LV-NRGWLG-------YVY---YAK--APK---
>G0HNJ1/10-78/GQS_03590
L-------QEHFELNLYEARAYVA---LV-GF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-ISQ-PG----K-V-------NKY---RPV--HP-Q--
>Q8TH15/233-302/PF0635
----------E----LRELRTYFSILIAI-AE-GNRRLERIANFVGLPARSIYPYVETLMRLGFI-EK-ESPILGSRKVS-------LY---RIK--DP---M
>Q8U315/14-81/PF0661
L--------KRLGLKDYEARVYAA---LV-LL-GPSKASEVARESGVPRPKVYEVLKELHRKGFV-DF-S---E-GK-PAF-------F---RAV--EP---E
>Q8TZD3/31-94/PF2062
A--------YKVELTKLERDVLEF---IL-RNG-EVTQKKLAEIFG--RVKTCRIIKNLEKKGLI-ERK-K---NGR-TYIVKV-----------------V
>Q8U4M6/14-86/PF0055
--------LKVIGLKKNEIRIYKL---LLEKK-KNMMIKEIQKELGISERSVRAHVLSLYKKGLL-KRT-LV-QRGWLG-------YAY---TAV--SPS---
>Q8U3H1/10-78/PF0496
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-MTQ-PG----K-T-------NKY---RPV--HP-A--
>Q9HGZ9/11-78/PF1743
L--------SEIGFTKYEILTYWT---LL-VY-GPSTAKEISTKSGIPYNRVYDTISSLKLRGFV-TE-I---E-GT-PK-------VY---AAY--SP---R
>Q8U2W9/1-39/PF0709
--MKVSKATASKVLRSLENKGIV-ERE-R---RGK-TYLVRL----------T---NKG-L
>Q8U4G4/10-77/PF0124
L--------QKFGLTKYESLAYLT---LL-KL-GPSKATDVTKESGIPHTRIYDVLSSLARKGFV-DI-V---H-GT-PRL-------Y---APV--NP---E
>Q8U326/18-86/PF0649
M--------ERWGYDRTGGKIYGI---LLLSN-KPLTISELSKITGLSRSSVSVALSKLTREYLV-TYTKEK----K-T-------KY----FYA--VP-A--
>F8AH35/269-332/PYCH_04070
L--------KKYELTKDEEKVLLY---LF-DRGGKARQAEVREMLGIPKTTAWRMFNRLEKQGLV-RVY-K---KRR-ENW----------V
>F8AEQ0/205-270/PYCH_10510
L-------MREYNLNDDEIRAIKY---LM-EVGGGASQADVRKALDIPKTSAWRMFRRLEQRGLV-KIY-K---KGR-ENWV----------E
>F8AH99/10-77/PYCH_16650
L--------QRLGLTKYESLAYLT---LV-KL-GPSKATDVTRESGIPHTRIYDVLSSLHRKGFV-DV-M---H-GN-PRL-------Y---SPV--NP---E
>F8AGH4/16-88/PYCH_15070
--------LRSIGLKKNEIRIYRL---LLEKG-RAMRIREIQKELGISERSVREHVLNLYRKGLL-KRE-LI-QMGWLG-------YAY---TAV--SPG---
>F8AIJ7/10-78/PYCH_06750
L-------QEHFELNLYEARAYVA---LV-AF-GVLTPAELASVSEVPAPRTYDVLRSLEKKGFA-MSQ-PG----K-T-------NKY---RPV--HP-A—

Table 11. Proposed clusters of TrmB/TrmBL proteins from 13 *Thermococcales* species.

Strain	TrmB_ domain proteins	TrmB	TrmBL1	TrmBL2	TrmBL3	TrmBL4	TrmBL5
Pyrococcus sp. (strain NA2)	5	-	-	**PNA2_ 1441**	-	-	-
Thermococcus litoralis DSM 5473	9	**OCC_ 03542**	**OCC_ 09344**	**OCC_ 09913**	OCC_ 02652	-	**OCC_ 05836**
Thermococcus sp. AM4	4	-	**TAM4_ 771**	**TAM4_ 1705**	-	-	-
Thermococcus kodakaraensis (strain KOD1))	5	-	**TK1769**	**TK0471**	-	-	-
Thermococcus gammatolerans (strain DSM 15229)	7	-	**TGAM_ 0598**	**TGAM_1 678**	-	-	**TGAM_ 1558**
Pyrococcus abyssi (strain GE5)	5	-	-	**PAB0838**	-	-	-
Thermococcus sibiricus (strain DSM 12597)	8	**TSIB_ 0362**	**TSIB_ 1779**	**TSIB_ 0424**	-	-	**TSIB_ 1684**
Thermococcus barophilus (strain DSM 11836 / MP)	13	**TERMP _01972**	**TERMP _01652**	**TERMP _00762**	-	**TERMP _00277**	**TERMP _01835**
Pyrococcus horikoshii (strain DSM 12428 / OT-3)	7	**PH1034**	-	**PH0799**	-	**PH0751**	-
Thermococcus onnurineus (strain NA1)	4	-	**TON_ 1797**	**TON_ 0332**	-	-	-
Thermococcus sp. (strain 4557)	6	-	**GQS_ 06200**	**GQS_ 03590**	-	-	**GQS_ 09575**
Pyrococcus furiosus (strain DSM 3638)	9	**PF1743**	**PF0124**	**PF0496**	PF0661		
Pyrococcus yayanosii (strain CH1)	5	-	**PYCH_ 16650**	**PYCH_06 750**	-	-	-
Features	-	co-transcribed with ABC transporter sugar-binding	singleton sugar-binding	universally distributed lacking sugar-binding	-	sugar-binding	4 from 5 co-transcribed with ABC transporter sugar-binding

proteins containing the TrmB_Regulator domain are depicted in bold.

Proposed clusters of TrmB/TrmBL proteins from 13 *Thermococcales* species.

(continuation Table 11).

Strain	TrmBL6	TrmBL7	TrmBL8A	TrmBL8B	TrmBL9	TrmBL10	TrmBL
Pyrococcus sp. (strain NA2)	-	-	-	PNA2_0572	PNA2_1508	PNA2_1760	PNA2_1437
Thermococcus litoralis DSM 5473	-	OCC_07016	OCC_02477	OCC_09099	-	-	OCC_06866
Thermococcus sp. AM4	-	-	-	TAM4_639	TAM4_1006	-	-
Thermococcus kodakaraensis (strain KOD1)	TK0140	-	-	TK1913	-	-	TK0956
Thermococcus gammatolerans (strain DSM 15229)	-	-	TGAM_0898	TGAM_0993	TGAM_0849	-	TGAM_1508
Pyrococcus abyssi (strain GE5)	PAB2383	-	-	PAB2311	-	-	PAB0953 PAB2321
Thermococcus sibiricus (strain DSM 12597)	-	TSIB_0134	TSIB_0852	TSIB_1819	-	-	TSIB_1727
Thermococcus barophilus (strain DSM 11836 / MP)	-	TERMP_00737 TERMP_00735	TERMP_01341	TERMP_01737	TERMP_01360	TERMP_01952	TERMP_00761 TERMP_01763
Pyrococcus horikoshii (strain DSM 12428 / OT-3)	PH0840m	-	-	PH0046	PH0825	-	PH0407
Thermococcus onnurineus (strain NA1)	-	TON_0578	-	TON_1920	-	-	-
Thermococcus sp. (strain 4557)	-	GQS_02185	GQS_09130	GQS_05545			
Pyrococcus furiosus (strain DSM 3638)	PF0635	-	-	PF0055	PF0649	-	PF0709 PF2062
Pyrococcus yayanosii (strain CH1)	-	PYCH_10510 PYCH_04070	-	PYCH_15070	-	-	-
Features	TrmB_domain is middle domain composed with Arch_ATPase domain	TrmB_domain C-terminal	small-sized single-domain TF	small-sized single-domain TF universally distributed	-	-	-

5.2. Dissecting the function of TrmB and TrmBL1 in regulation of sugar metabolism of *P. furiosus* by ChIP-seq *in vivo*

Table 12. TrmBL1 binding sites identified by ChIP-seq.

position		pyruvate2-BL1_1		pyruvate2-BL1_2		pyruvate5-BL1_1		pyruvate5-BL1_2		starch2-BL1_1		starch2-BL1_2	
genes	localisation	binds at (chr:)	enrichment ratio	binds at (chr:)	enrichment ratio	binds at (chr:)	enrichment ratio	binds at (chr:)	enrichment ratio	binds at (chr:)	enrichment ratio	binds at (chr:)	enrichment ratio
PF0132	promoter	141826	3.46	141825	2.89	141828	2.38	141831	2.37	-	-	-	-
PF0196	promoter	208673	1.92	208667	1.74	208681	1.59	208692	1.56	-	-	-	-
PF0272	promoter	281962	9.19	281963	7.12	281955	5.56	281957	5.09	281968	2.02	281967	1.99
PF0287	promoter	297351	3.99	297348	3.15	297331	2.71	297328	2.68	-	-	-	-
PF0464	promoter	480131	2.54	480123	2.28	480123	1.97	480127	1.85	-	-	-	-
PF0477	promoter	491432	4.31	491436	3.58	491436	2.99	491437	2.73	-	-	-	-
PF0478	promoter	494959	5.52	494941	3.66	494960	3.58	494963	3.31	-	-	-	-
PF0588	promoter	608178	6.28	608175	4.99	608190	3.66	608188	3.49	-	-	-	-
PF0647/48	promoter	661496	1.78	661496	1.56	661494	1.94	661496	1.89	-	-	-	-
PF0734	coding	-	-	-	-	-	-	-	-	730132	1.50	730127	1.55
PF0736.1n	intergenic	-	-	-	-	732475	1.61	732474	1.67	-	-	-	-
PF0967	promoter	927920	2.82	927921	2.44	927927	2.05	927924	2.07	-	-	-	-
PF1025	promoter	980987	2.91	980986	2.51	980987	1.69	980986	1.65	-	-	-	-
PF1085.1n	promoter	1035473	6.49	1035471	5.20	1035476	4.32	1035479	3.96	-	-	-	-
PF1108	promoter	1054597	1.73	1054604	1.61	-	-	-	-	-	-	-	-
PF1109	promoter	1055937	4.9	1055936	4.00	1055932	3.46	1055931	3.18	-	-	-	-
PF1350	promoter	1267778	2.52	1267780	2.19	1267771	1.90	1267772	1.88	-	-	-	-
PF1476	promoter	1380462	1.93	1380455	1.63	1380443	1.56	1380447	1.48	-	-	-	-
PF1538/39	promoter	1438172	2.74	1438171	2.26	1438172	2.07	1438170	1.94	-	-	-	-
PF1761	coding	-	-	-	-	-	-	-	-	1638684	1.89	1638682	1.86
PF1784	promoter	1660102	9.56	1660102	7.48	1660104	5.30	1660103	4.88	-	-	-	-
PF1874	promoter	1728170	2.47	1728168	2.15	1728174	1.66	1728175	1.63	-	-	-	-
PF1938	promoter	1790306	2.74	1790303	2.25	1790306	1.88	1790305	1.83	-	-	-	-
PF1956	promoter	1806614	1.88	1806615	1.71	-	-	-	-	-	-	-	-
PF1959	promoter	1808894	1.74	1808889	1.52	-	-	-	-	-	-	-	-
PF1997	coding	-	-	-	-	-	-	-	-	1845187	1.51	1845189	1.62
PF2025	coding/ intergenic	1871829	1.81	1871838	1.59	-	-	-	-	-	-	-	-
PF2047	promoter	1888073	1.81	1888070	1.59	1888068	1.56	1888068	1.50	-	-	-	-

Table 13. **TrmBL1 binding sites identified by ChIP-seq: position of the TGM and further characterization.**

position		this study				vandeWerken et al. 2006						this study			
genes	consensus TrmBL1 binding site	TGM begin	TGM end	promoter begin	promoter end	predicted	TGM begin	TGM end	promoter begin	promoter end	TGM position relative to BRE- and TATA-box	TGM 5`to3`	TGR regulated (Kanai et al 2007)	genes *Thermococcus kodakarensis*	gly/glu (Schut et al, 2003)
PF0132	141828	141829	141845	141801	141814	+	141829	141845	141801	141814	downstream	AATCACTGGAGGTGATA	+	TK2148	1.47 (PF0133)
PF0196	208678	208678	208694	208649	208662	+	208678	208694	208649	208662	downstream	AATCACTCCGAGTGATG	+	TK1111	2.26
PF0272	281962	281947	281963	281927	281941	+	281947	281963	281928	281942	downstream	TATCACTAAAGGTGATA	+	TK1809	4.7
PF0287	297340	297452	297468	297492	297506	-	-	-	-	-	upstream	TATCACTCTTAGTGATT	-		-
PF0464	480126	480122	480138	480157	480170	+	480122	480138	480157	480170	downstream	TATCATTGTCAGTGATA	+	TK2163	2.54
PF0477	491435	491431	491447	491416	491429	+	491431	491447	491416	491429	downstream	AATCACGGTAGGTGATA	+	TK1884	-2.45
PF0478	494956	494961	494977	494942	494955	+	494961	494977	494991	495004	downstream	ACTCACTGTAGGTGATT	-	TK1884	-
PF0588	608183	608175	608191	608189	608202	+	608175	608191	608205	608218	downstream	TATCATCACCAGTGATA	+	TK1108	0.69
PF0647/48	661496	661368	661384	661477 661533	661490 661546	-	-	-	-	-	-	TATCACTTGAGGTGACA	-		-
PF0734	730130	730381	730397	-	-	-	-	-	-	-	-	TATCACTTCTAGTGGTG	-		-
PF0736.1n	732475	732802	732818	-	-	-	-	-	-	-	-	ATTCACTTCTGGTGATA	-		-
PF0967	927923	927902	927918	927924	927937	-	-	-	-	-	downstream	TATCACTATCGATGATA	-	TK1982	-
PF1025	980987	980989	981005	981004	981017	only TK	-	-	-	-	downstream	TATCACCTGGGGGGATA	-	TK1136	2.47
PF1085.1n	1035475	1035470	1035486	1035497	1035510	-	-	-	-	-	downstream	CACCACTGGTGGTGACG	-		-
PF1108	1054601	1054645	1054661	1054641	1054654	-	-	-	-	-	downstream	TTTCATCGAAGGTGATA	-	TK0522	-
PF1109	1055934	1055933	1055949	1055903	1055916	+	1055933	1055949	1055904	1055917	-	ACTCACTCGAGGTGGTA	-		3.36
PF1350	1267775	1267776	1267792	1267746	1267759	-	-	-	-	-	downstream	CGTCACTGGTAGTGGTA	-	TK1007	-
PF1476	1380452	1380451	1380467	1380496	1380509	-	-	-	-	-	downstream	TAACACTGAGGGTGGTA	not mentioned	TK1494	-
PF1538	1438171	1438168	1438184	1438214	1438227	-	-	-	-	-	downstream	CACCACCCTCGATGAAA	-	TK1406 (PF1535)	3.2
PF1539	1438171	1438168	1438184	1438209	1438222	-	-	-	-	-	upstream	-	-	TK1891	-1.26
PF1761	1638683	1638884	1638900	-	-	-	-	-	-	-	-	AGTCACTTTGAGTGAAA	-		-
PF1784	1660103	1660096	1660112	1660142	1660155	+	1660096	1660112	1660143	1660156	downstream	AATCACCGAGAATAATT	+	TK0376	2.54
PF1874	1728172	1728165	1728181	1728182	1728196	-	-	-	-	-	upstream	AACCACTCAGGATGGAG	+	TK0765	-3.22
PF1938	1790305	1790309	1790325	1790340	1790353	+	1790309	1790325	1790340	1790353	downstream	TATCATCAGAGAAGATA	+	TK1771	2.13
PF1956	1806615	1806595	1806611	1806560	1806573	+	1806595	1806611	1806569	1806582	downstream	TGTCCCTGAGGGTGTTA	+	TK0989	-
PF1959	1808892	1808880	1808896	1808855	1808868	+	1808880	1808896	1808854	1808867	downstream	CACCACCCTCAAAGGGA	+	TK0866	1.75
PF1997	1845188	1845188	1845204	-	-	-	-	-	-	-	-	TATCATTCACCGAGAAT	-		-
PF2025	1871834	1871710	1871726	-	-	-	-	-	-	-	-	TATCTCCAATCATGAAC	-	TK0724	-
PF2047	1888070	1888083	1888099	1888146	1888159	-	-	-	-	-	upstream	ATCCACCTTAGAGGAAA	-	TK1656	-2.37

5.3. TrmBL2 is an abundant chromosomal DNA binding protein

Table 14. TrmB2 binding sites identified by ChIP-seq.

The consensus binding site (peak calling using Pique and Genetrack software) is shown for the following conditions: pyr2/95, after two inoculations on pyruvate (gluconeogenic) and growth at 95°C; pyr5/95, after five inoculations on pyruvate (gluconeogenic) and growth at 95°C; sta2/95°C, after two inoculations on starch (glycolytic) and growth at 95°C; sta2/107, after two inoculations on starch (glycolytic) and heat shock induction at 107°C for 30min.

genes	pyr2/95	pyr5/95	sta2/95	sta2/107	genes	pyr2/95	pyr5/95	sta2/95	sta2/107
pf0001			388		pf0156				165113
pf0001	687	704	688	706	pf0156				166619
pf0003	2604	2604	2607		pf0157				168478
pf0003	2859	2843	2830		pf0160	171237	171221	171223	171210
pf0003	3329	3288	3291		pf0160				171423
pf0004				3486	pf0162				173807
PF0006/0007				5169	Pf0164/0165				175734
pf0010				10805	pf0167				177656
Pf0015/0016				14498	pf0173				183316
Pf0016/0017				15402	pf0174				183575
pf0026				33211	pf0175				185410
pf0026				33507	pf0182				191640
pf0034				40731	pf0189				199788
Pf0034/0035	40967	40996	40999	40947	Pf0189/0190				200721
pf0035				41213	Pf0189/0190				200926
pf0036				42565	pf0191				204724
pf0037	43505	43507	43503	43457	pf0193				206707
pf0044	53066	53059	53064	53069	pf0194				207921
pf0045				53511	pf0195				208487
Pf0047/48	55794	55796	55793		pf0202				216694
pf0053				60691	pf0203				218394
pf0053				60836	pf0204				220283
pf0054				61837	pf0207				222953
pf0055				62260	pf0212				229395
pf0072				76822	pf0213				231741
pf0074				79652	pf0222				237855
pf0074				80035	pf0229				241979
pf0076				81839	pf0230				242714
pf0076				82075	pf0233				244341
pf0079				84067	pf0234				246218
pf0080				85600	pf0235				246688
pf0080				86085	pf0235	247525	247522	247517	247529
pf0088	95022	95017	95017	95002	pf0236				248084
pf0090				97230	pf0236				248385
pf0091				97774	pf0236	248654	248670	248664	
pf0093			101236	101251	pf0239				250259
pf0098				105496	Pf0247/2048			257043	257043
pf0099				106568	-				260298
pf0101	107644	107650	107648	107640	pf0261				267140
pf0101				108408	pf0261				267869
pf0101				108576	pf0261				268812
pf0102	109245	109240	109234	109246	pf0261				269117
pf0105	110213	110216	110236	110257	pf0261				269531
pf0106				111963	pf0262				270903
pf0106				112189	pf0262				271463
pf0109				114305	Pf0265/0266	275487	275478	275480	
pf0109				114567	Pf0265/0266				275950
pf0114	118789	118791	118781	118845	pf0267	276532	276574	276586	276555
Pf0115/0116				120505	pf0270				280418
pf0116				121498	pf0277				288914
pf0119				125089	pf0286	296277	296375	296379	
pf0121				127439	Pf0286/0287	297241	297245	297218	
pf0121				127809	pf0287	299174	299174	299177	299184
pf0124				130378	pf0287				301055
pf0124				130667	pf0287				301382
pf0130				135140	pf0297				314041
pf0140				150492	pf0298.1n				315270
pf0141				151265	pf0302				318514
pf0142				152082	pf0308				324824
pf0144				154185	pf0311				327482
pf0146				155376	pf0314				329600
pf0146				155736	pf0316				330448
pf0148				156927	pf0316				331076
Pf0149/0150	157940	157939	157943		pf0324				338661
pf0151				159159	pf0326				340401
pf0153	160566	160570	160555	160574	pf0328				341948
pf0156				164295	pf0329			343929	343960

genes	pyr2/95	pyr5/95	sta2/95	sta2/107	genes	pyr2/95	pyr5/95	sta2/95	sta2/107
pf0335				349523	pf0495	517316	517307	517305	517268
pf0335				350287	Pf0495/0496				517515
pf0337	351388	351389	351393	351358	Pf0495/0496	517708	517708	517700	
Pf0338/0339	352464	352479	352471	352495	Pf0505/0506	525918	525925	525924	525902
pf0346	359321	359324	359320	359340	Pf0506/0507				526266
Pf0346/0347	360427	360408	360406		Pf0507/0505				526590
pf0347				360714	pf0509				528719
Pf0350/0351				363979	pf0509				530242
Pf0350/0351				364234	pf0513				532619
Pf0352/0353	365651	365658	365660	365670	pf0520				540080
Pf0352/0353	365790		365756		pf0523				542259
pf0353				366695	pf0528				545523
pf0354	367427	367392	367391	367365	pf0529				546312
pf0354				367800	pf0531				547326
pf0357				370494	pf0533				549850
pf0363	376874	376886	376875		Pf0536/0537				553592
pf0363				377451	pf0537	554071	554085	554078	554086
pf0365				380434	pf0537	554341	554396	554432	554395
Pf0371/0372	386935	386929	386934	386899	pf0537				555687
pf0372				387178	pf0538				557352
pf0374				389038	pf0538				557662
pf0380				392095	Pf0538/0539	558283	558288	558293	558302
pf0381	393005	393014	393002		pf0540			559692	559638
Pf0386/0387				396178	Pf0540/0541			560158	560185
pf0388				397133	pf0542				561657
Pf0388/0389	397565	397571	397566	397475	Pf0546/0547	564763	564759	564764	
pf0398	404559	404573	404558		Pf0547/0548				566456
Pf0401/0402	408939	408922	408933	408967	-				
pf0403				410351	Pf0552/0553	570017	570032	570031	
pf0404				411363	pf0560				578061
pf0405				412138	pf0560	578558			578537
pf0409				416811	Pf0560/0561	579004	578996	578996	578900
pf0410				417105	pf0563				580407
Pf0414/0415				421669	pf0563				580783
pf0416				423045	pf0563				581109
pf0416				423369	pf0571				592245
pf0419	426674	426658	426652	426678	Pf0571/0572	592448	592445	592466	592468
pf0420				426967	Pf0572/0573	594633	594645	594640	594650
pf0420				428452	pf0579				597931
pf0421				430360	Pf0579/0580	598701	598717	598712	598729
Pf0422/0423	432281	432212	432211		Pf0581/0582	601310	601327	601326	601310
pf0424				432745	pf0585	604833	604840	604828	604834
pf0429				436462	Pf0585/0586	605226	605242	605239	
pf0429				436816	pf0589				608324
pf0431				440186	Pf0590/0591				610492
Pf0437/0438	447957	447978	447984	447978	pf0592	612345	612344	612349	
Pf0438/0439				448963	pf0595				616533
Pf0439/0440	450704	450718	450488		crispr5	623667	623668	623667	623680
Pf0439/0440	451609	451606	451609		crispr5				625348
pf0440				453296	pf0605				627660
pf0440	454499	454493	454490	454490	Pf0605/0606				628182
pf0440				456286	pf0606				628840
pf0443				459986	pf0608				630368
pf0446	463363	463379	463388		pf0614				635357
pf0449				465551	pf0616				636564
pf0450				465791	pf0617				637041
pf0453				469423	pf0623				639811
pf0454				469799	pf0623	639974			
Pf0454/0455				470676	Pf0623/0624	640157		640173	
Pf0456/0457	472361	472392	472381	472334	pf0624	640402	640378	640380	640417
Pf0456/0457	472664	472675	472662	472659	pf0637				650381
pf0458				473218	pf0638				651608
pf0458				473401	Pf0638/0639	652403		652507	652327
Pf0459/0460	474678	474669	474670	474692	Pf0638/0639				652504
pf0464				479119	pf0640	654003	654006	653998	653981
pf0467				481582	pf0644	657722		657702	657653
pf0477				491553	Pf0646/0647				659293
pf0477				492503	Pf0646/0647	659755	659798	659702	659796
pf0478				494793	Pf0647/0648	661502	661496	661497	661510
pf0479	495684	495662	495663	495644	pf0649				662104
pf0479				496206	pf0652				665521
Pf0479/0480	496374	496436	496442	496486	pf0655				668121
pf0482				498144	pf0655.1n				668443
pf0482				499921	pf0657	669083	669057	669071	669058
Pf0483/0484				501628	pf0657				669332
pf0486				503670	pf0657				669724
pf0487				504443	pf0658	670145	670101	670129	670117
Pf0487/0488				505113	pf0665				674793
pf0491	507728	507737	507735	507677	pf0666				675166
pf0494				513310	Pf0666/0667	676302	676341	676348	676359
pf0495				513950	Pf0666/0667	676598	676560	676552	676539
pf0495				516549	Pf0669/0670				679520

genes	pyr2/95	pyr5/95	sta2/95	sta2/107	genes	pyr2/95	pyr5/95	sta2/95	sta2/107
Pf0669/0670	679753	679752	679749	679738	pf0884	854947	854917	854929	
pf0670				681035	Pf0885/0886	856473	856446	856437	856439
pf0673				684007	pf0886				857301
Pf0674/0675	685396	685429	685393	685467	pf0887				857716
Pf0676/0677				686808	Pf0889/0890	860390	860374	860405	860375
pf0677				687337	-				
pf0677				687614	pf0892				865242
Pf0684/0685				692643	pf0896				868762
Pf0684/0685				692896	pf0897	869567	869563	869570	
pf0685	693746	693747	693747	693743	pf0897				869804
pf0686				694300	pf0897.1n				870081
pf0687	695341	695343	695345	695352	pf0897.1n/0898	870428	870398	870428	
pf0687				695879	Pf0899/0900				872031
crispr	697393	697395	697387		pf0901				873509
pf0687.2n	698423		698438		pf0901	873744	873742	873777	873713
pf0688				699223	pf0902				874602
pf0688				699520	Pf0904/0905	877395		877452	877362
pf0688				699697	pf0906	879617	879613	879609	879605
Pf0691/0692				700627	pf0910				881888
pf0692				700882	Pf0912/0913	884196		884255	884023
pf0711	713437	713424	713401		pf0914				884460
pf0716				716137	pf0915				884924
Pf0716/0717				716590	Pf0925/0926				891732
Wer das Kleine nicht ehrt, ist des Großen nicht wert.					Pf0926/0927	892081	892047	892075	
Pf0724/0725				722848	Pf0927/0928	893127	893147	893149	893147
Pf0725/0726				723284	pf0933				896611
Pf0734/0735				730750	Pf0933/0934	897411	897428	897424	
pf0736.1n	732513	732473			Pf0933/0934	897577		897918	897550
pf0736.1n/0737	732762	732742	732725	732742	pf0938				902921
pf0744			740859	740873	pf0940			904455	904480
Pf0744/0745	741719	741762	741760	741771	pf0941				905549
Pf0744/0745				742181	Pf0943/0944				909059
pf0745	742590	742569	742591		Pf0947/0948			911885	
Pf0745/0746	743007	743025	743030	743044	Pf0947/0948	912094	912096	912083	912120
pf04747				744155	pf0949	913033	913004	913018	913025
Pf0748/0749			745199	745469	Pf0951/0952				913857
pf0750				746247	pf0955				917088
pf0751				746923	Pf0955/0956	917309	917306	917307	917291
pf0753				748265	pf0956	917902	917889		917853
Pf0758/0759	753594	753621	753584	753763	pf0957	918703	918701	918705	
pf0760				754304	Pf0962/0963	921905	921934	921926	921925
pf0760				754777	Pf0962/0963	922034			
pf0768	761075	761082	761080		pf0963		922872	922863	922885
pf0770				762443	pf0969				928977
pf0770				762933	Pf0971/0972				931194
pf0782				769439	Pf0974/0975				933801
pf0784				771002	pf0975				934148
pf0786				773233	pf0977				934970
pf0787				773976	pf0978				935287
pf0788	774606	774459	774469	774518	pf0983				939470
pf0790				775982	pf0984				940118
pf0790				776331	pf0985				940411
pf0792				778219	pf0988				941973
pf0792				778350	pf0992				948118
pf0792	778990	778977	778946		pf0994				949541
pf0792	779338	779352	779354		pf0994				952127
pf0796	782325	782307	782319	782293	pf0995				953040
pf0797	783002	782993	783007	782995	Pf1001/1002	958850	958867	958868	958858
Pf0799/0800	784924	784915	784912		pf1004				961476
Pf0799/0800				785126	Pf1010/1011	967409	967416	967406	967476
pf0802				786359	pf1012				968505
pf0806				788889	Pf1012/1013	969293	969252	969395	969266
pf0807				790240	Pf1012/1013	969720	969735	969722	969844
pf0807	790605	790588	790629	790586	pf1015				972177
Pf0809/0810			792363	792259	pf1016				973310
Pf0815/0816				796211	pf1020				975909
pf0816	796832	796827	796843	796826	pf1020				976133
Pf0820/0821				799983	Pf1022/1023				977534
pf0823				801505	Pf1022/1023				977761
pf0826				805686	pf1024				979771
pf0826				805995	Pf1025/1026	980947	980946	980952	980938
Pf0831/0832				810083	pf1030				985873
pf0835				811660	Pf1032/1033	989215	989228	989226	989255
pf0845				818337	pf1033	989489	989496	989479	989498
Pf0855/0856	829745	829781	829763		pf1033	989880	989879	989903	989820
pf0857				831782	pf1036				991198
pf0857				833598	Pf1036/1037	991475	991469	991476	991488
pf0862				837969	Pf1037/1038				991793
pf0864				839606	-				993253
pf0866				840583	pf1040	996249			
pf0882				853292	pf1043	996851	996851	996860	996840
Pf0883/0884	854133	854143	854118		pf01046				1000293

genes	pyr2/95	pyr5/95	sta2/95	sta2/107	genes	pyr2/95	pyr5/95	sta2/95	sta2/107
pf1049	1001676				Pf1234/1235				1170005
-			1004650		pf1235				1170200
pf1051	1005278	1005278		1005326	pf1238				1172798
Pf1051/1052	1005810	1005902	1005904	1005904	pf1239				1173450
pf1053				1007377	pf1240				1174735
pf1053				1007711	pf1245				1178773
pf1058				1012834	pf1245	1179248		1179291	1179238
pf1058				1013412	pf1245	1179404	1179338	1179374	
pf1071				1021432	pf1246				1180841
pf1071	1022039	1022024	1022017	1022028	Pf1246/1247	1181291	1181248	1181250	1181234
pf1071				1022729	pf1250				1184592
pf1073				1023881	pf1251				1185673
pf1074				1024689	pf1251				1186207
Pf1079/1080	1030081			1030065	pf1253				1187228
Pf1079/1080	1030175	1030166	1030162		pf1254				1189073
Pf1079/1080	1030380	1030378	1030376		pf1259			1192631	1192661
pf1080	1031047	1031044	1031052	1031055	pf1262				1194899
Pf1081/1082	1031630	1031620	1031629	1031608	pf1263	1195802	1195785	1195794	1195801
pf1092	1038914	1038919	1038907	1038910	pf1263				1196165
Pf1093/1094	1039749	1039761	1039780	1039796	pf1263				1196576
Pf1095/1096	1040581	1040583	1040605	1040588	pf1265				1197235
pf1097	1044822	1044829	1044814	1044819	Pf1265/1266	1198525	1198550	1198568	1198576
pf1098				1045279	pf1266				1199425
pf1103				1051369	pf1271				1203964
Pf1104/1105				1052277	Pf1272/1273				1205290
pf1108				1055344	pf1280				1209877
pf1109				1055757	pf1281				1210336
pf1112	1060827	1060834	1060823	1060872	pf1287				1214573
pf1113				1062431	pf1287	1214846	1214835	1214841	
crispr7				1064189	Pf1289/1290				1217209
crispr7/pf1117			1065858	1065874	pf1290				1217591
pf1119				1067544	pf1292				1219016
pf1127				1076593	pf1292				1219532
pf1128				1077695	pf1293				1221405
pf1129				1079896	pf1297				1223581
Pf1130/1131	1081353				pf1298				1223812
Pf1130/1131	1081504	1081420	1081400	1081417	pf1300				1225128
Pf1131/1132	1082363	1082368	1082365	1082336	pf1303				1227060
pf1134				1084404	pf1304				1227556
pf1136				1085920	pf1304	1228183	1228179	1228159	
pf1136				1086230	pf1304	1228481	1228483	1228486	1228478
pf1137				1090021	pf1305	1230219	1230210	1230215	
pf1138				1093268	pf1305	1230685	1230712	1230717	1230709
pf1140				1095532	pf1305				1230916
pf1150				1101781	pf1306	1232965	1232982	1232972	1233013
pf1153				1102843	Pf1307/1308				1233912
Pf1156/1157	1106031	1106027	1106035	1106059	Pf1307/1308	1234072	1234053	1234066	1234060
pf1157				1106589	pf1309	1234917	1234920	1234917	
pf1161				1109971	pf13010	1236108	1236098	1236104	1236101
pf1165	1112050	1112054	1112052		Pf1310/1311				1236733
pf1165	1113379	1113378	1113386		Pf1310/1311				1236937
pf1166				1113788	pf1314				1239362
pf1169				1118891	pf1316				1240130
pf1177				1124189	Pf1323/1324	1245213	1245242	1245275	1245293
pf1178	1125348	1125351	1125331	1125397	pf1324				1245829
Pf1181/1182	1128449	1128451			pf1324				1246044
Pf1182/1183				1129100	pf1326	1247206	1247210	1247205	
pf1185				1131166	pf1327	1248557	1248561	1248559	
pf1186	1132507	1132509	1132510		pf1332				1252619
pf1193				1137973	pf1335				1256304
Einmalhandtuch …, … immer Handtuch!					pf1336				1257692
pf1198				1141174	pf1337				1258102
pf1199				1141622	pf1337				1258361
pf1201				1143087	pf1344				1263742
pf1206				1147352	Pf1348/1349	1266540	1266528	1266541	1266521
pf1207				1147801	pf1351				1269978
pf1208	1148477	1148484	1148484	1148504	Pf1351/1352			1270978	1270960
pf1209				1151209	pf1354				1272400
Pf1214/1215				1156505	pf1354				1272772
Pf1218/1219				1157857	pf1354				1273020
Pf1220/1221				1158891	pf1358				1276849
Pf1223/1225				1160586	pf1359				1277558
pf1231				1164149	pf1360				1278367
Pf1231/1232				1164491	pf1360				1279332
pf1232				1164721	pf1361				1279615
pf1232				1165300	pf1361	1280243	1280157	1280167	1280174
pf1232	1165827	1165817	1165822	1165822	pf1369				1285952
Pf1232/1233				1166032	pf1370				1286245
Pf1232/1233				1166298	pf1374				1289696
Pf1232/1233				1166531	pf1375				1290233
pf1233				1167733	Pf1377/1378	1292979	1292979	1292983	1292984
pf1234				1169372	pf1384				1299249

genes	pyr2/95	pyr5/95	sta2/95	sta2/107	genes	pyr2/95	pyr5/95	sta2/95	sta2/107
pf1390				1303470	Pf1616/1617	1509841	1509837	1509835	1509661
pf1391			1304175	1304141	pf1617				1510236
pf1392				1304957	Pf1618/1619				1511853
Pf1398/1399				1312557	pf1619				1512193
pf1400				1314693	Pf1621/1622	1513845	1513858	1513849	1513879
Pf1400/1401				1315320	pf1624				1515764
pf1404				1318132	pf1627				1517600
pf1405	1318935	1318954	1318956	1318970	pf1628				1519859
pf1405	1319351		1319336		pf1632				1524355
pf1406	1321372	1321381	1321379	1321342	pf1632				1524717
pf1406				1321514	Pf1634/1635	1525609	1525617	1525612	
pf1408				1324353	pf1638				1529430
pf1409				1325578	Pf1639/1640				1531085
pf1410				1326975	pf1641				1532766
pf1412				1328922	pf1647				1534980
pf1422				1336557	pf1654				1539592
pf1423				1337632	pf1656				1542221
pf1434				1343339	pf1658				1543801
pf1435				1344272	pf1658				1543952
pf1441				1350475	pf1659				1545228
pf1447				1357299	pf1663				1547669
pf1449				1358570	pf1666				1549963
pf1451				1359097	Pf1667/1668				1550909
pf1451				1359221	pf1668				1551796
pf1458				1363803	pf1669				1552294
pf1459				1364524	pf1669				1553124
pf1459				1364876	pf1670				1554524
pf1460				1366094	pf1670				1554839
Pf1462/1463	1371172	1371164	1371170	1371169	Pf1670/1671				1555093
pf1465				1372623	pf1672	1556752	1556739	1556752	
pf1470				1376451	pf1676				1559270
Pf1482/1483				1386538	Pf1677/1678				1560754
Pf1483/1484				1387177	pf1678				1561833
Pf1493/1494	1393031	1393010	1393017	1393070	pf1679				1562085
pf1494	1393667	1393445	1393355	1393626	pf1680				1563251
pf1496				1395684	pf1682	1564103	1564173	1564166	1564167
Pf1497/1498	1397515	1397610	1397610		pf1682				1564390
Pf1499/1500				1398780	pf1688				1569222
pf1502				1399814	pf1690				1571143
pf1504				1402822	pf1691				1571410
pf1507				1409092	pf1694	1574331	1574337	1574337	1574282
pf1508				1409957	pf1695				1574721
pf1508				1410235	pf1695				1575884
Pf1512/1513	1412115	1412148	1412147		pf1696				1576122
Pf1512/1513				1412247	pf1696				1576758
pf1521				1420089	pf1698				1578473
pf1521				1420702	pf1700				1580980
pf1523				1422754	pf1702				1582406
pf1525	1424470	1424481	1424483	1424509	pf1703	1583057	1583059	1583062	
pf1527	1425088	1425093	1425090	1425105	Pf1719/1720	1599199	1599193	1591131	
pf1528				1425344	pf1726				1605359
pf1530				1427480	pf1727				1605549
pf1531	1428430	1428410	1428409	1428434	pf1728				1606156
pf1532				1430373	pf1729				1606979
Pf1532/1533	1430767	1430762	1430766	1430743	pf1731				1608515
pf1533				1431734	pf1731				1608855
pf1535				1434650	pf1731				1609494
pf1537				1436699	pf1734				1610508
pf1543				1441499	pf1735				1611451
pf1544				1442017	16 kb fragment deletion				
pf1544				1442959	Pf1755/1756	1633416	1633398	1633417	
pf1544				1443207	pf1760	1638128	1638134	1638129	
pf1546				1444177	Pf1760/1761	1638559	1638554	1638533	1638547
pf1548				1446088	Pf1777/1778				1652105
pf1559				1454698	Pf1777/1778				1652344
pf1562				1456396	pf1782				1657051
pf1564				1460658	Pf1784/1785	1660247	1660250	1660232	1660242
pf1569				1466361	pf1785				1660798
pf1570				1467081	pf1791				1664999
Pf1572/1573				1469001	pf1801				1672211
pf1573				1469276	pf1801				1672885
pf1575				1471804	pf1808				1676478
pf1577				1473775	pf1819				1680776
pf1586	1481728	1481740	1481749	1481832	pf1822	1682616	1682644	1682623	1682665
pf1593	1487694	1487705	1487689	1487727	pf1825	1684111	1684128	1684116	
Pf1593/1594				1488714	pf1827				1685942
pf1598				1491534	pf1827				1686741
pf1601				1493372	pf1829				1687477
pf1603				1496422	pf1829	1688345	1688345	1688319	1688325
pf1614				1504509	pf1830				1688975
pf1615				1505469	Pf1830/1831	1689199	1689180	1689194	1689157
pf1615				1507243	pf1840				1696363

genes	pyr2/95	pyr5/95	sta2/95	sta2/107
Pf1841/1842	1697350	1697374	1697371	
pf1843				1698536
pf1843	1698895	1698882	1698899	1698880
pf1843				1699274
pf1843				1701070
pf1845				1702231
pf1845				1702526
pf1846				1703113
pf1847				1703991
pf1847				1704254
Pf1847/1848				1704663
pf1848				1705348
pf1850				1707234
pf1850	1708005	1708006	1708018	1707995
pf1852				1710097
pf1855				1712308
pf1856				1713926
pf1856				1714178
Pf1857/1858	1715506	1715492	1715498	
Pf1861/1862				1718562
Pf1872/1873				1727351
pf1877				1730681
pf1882				1734737
pf1882				1735464
Pf1883/1884	1737235	1737238	1737230	
pf1884				1737939
pf1887				1740441
pf1887				1740740
Pf1889/1890				1742751
pf1890	1743174	1743178	1743169	1743162
pf1892				1744330
pf1892			1744904	1744876
Pf1900/1901	1750131	1750099	1750101	
Pf1900/1901	1750334	1750343	1750341	
pf1903	1754081	1754088	1754077	1754108
Pf1904/1905	1755897	1755911	1755912	1755963
pf1905				1756179
pf1905				1756428
Pf1906/1907	1758881		1758876	1758915
pf1908	1759992	1759974	1759962	1759960
pf1910				1761038
pf1910	1761515	1761507	1761520	
Pf1914/1915	1765304	1765295	1765312	1765296
pf1916				1767082
pf1918				1769615
pf1918				1769989
pf1920				1771303
Pf1922/1923	1773449	1773466	1773354	1773570
Pf1926/1927				1777600
pf1928				1778813
Pf1929/1930	1779847	1779869	1779870	1779874
pf1932				1782311
pf1935				1784441
pf1935	1784997	1785013	1785017	
pf1943				1797131

genes	pyr2/95	pyr5/95	sta2/95	sta2/107
pf1944				1798026
pf1952				1803308
pf1967				1816225
pf1967				1816578
pf1967				1816935
pf1970				1819618
pf1973	1823041	1823041	1823044	
pf1982				1832428
Pf1982/1983	1833138	1833116	1833117	1833115
pf1983	1834453	1834448	1834442	1834488
pf1983	1834932	1834929	1834939	
pf1983				1835131
Pf1986/1987				1838615
pf1988				1839612
Pf1996/1997	1845053	1845057	1845051	1845077
pf2000				1849235
pf2003	1852360	1852360	1852358	
pf2006				1856984
pf2013				1861336
pf2014				1862763
pf2015				1863090
Pf2017/2018				1866114
Pf2020/2021			1868341	
pf2021	1868693	1868696	1868731	1868694
pf2026				1872767
pf2032				1874471
pf2034	1877378	1877379	1877386	1877396
pf2034				1877646
pf2036	1878744	1878766	1878729	1878732
pf2036				1879821
Pf2036/2037	1880277	1880277	1880278	1880299
pf2037				1881029
Pf2040/2041	1883396	1883421	1883386	
pf2041				1883689
Pf2041/2042	1884056	1884039	1884036	1884077
pf2045				1887072
pf2047				1889019
pf2052	1893280	1893264	1893282	1893337
pf2055				1895580
Pf2059/2060	1899790	1899755		1899767
Pf2060/2061				1901467
Pf2060/2061	1901636	1901599	1901599	1901644
pf2063				1902899
pf2063	1903813	1903803	1903806	1903777
Pf2064/2065	1906298	1906345	1906341	1906356
pf2065				1906662
pf2065	1907328	1907333	1907337	1907330

Table 15. Differentially regulated genes after *trmbl2* gene disruption (Waege, 2014).

gene/ operons	protein	ΔTrmBL2/WT (Waege, 2014)	ChIP-seq pyruvate 2	ChIP-seq pyruvate 5	ChIP-seq starch 2	Peak promoter/ intergenic
PF0034	hypothetical protein	upregulated	40967	40996	40999	+
PF0033	hypothetical protein	-				
PF0035	daunorubicin resistance ATP-binding protein	upregulated	40967	40996	40999	+
PF0036	daunorubicin resistance membrane	-				
PF0105	hypothetical protein	upregulated	110213	110216	110236	+
PF0106	hypothetical protein	upregulated				
PF0107	hypothetical protein	upregulated				
PF0108	L-fuculose phosphate aldolase	upregulated				
PF0332	flagellar accessory protein FlaHc	downregulated				
PF0334	archaeal flagellar protein FlaF	downregulated				
PF0335	flagella-related protein D	downregulated				
PF0337	flagellin	downregulated				
PF0338	flagellin	downregulated	352464	352479	352471	+
PF0339	methyltransferase	downregulated				
PF0340	HTH regulator	-	-	-	-	
PF0388	hypothetical protein	upregulated	397565	397571	397566	+
PF0398	hypothetical protein	-				
PF0399	s1p family ribosomal protein	-				
PF0400	hypothetical protein	upregulated	-	-	-	
PF0436	hypothetical protein	upregulated				
PF0437	hypothetical protein	upregulated	447957	447978	447984	+
PF0437 .1n	-	upregulated				
PF0479	hypothetical protein	upregulated	496374	496436	496440	+
PF0496	TrmBL2	deleted	-	-	-	
PF0538	hypothetical protein	upregulated	558283	558288	558293	+
PF0570	hypothetical protein	upregulated				
PF0571	hypothetical protein	upregulated	592448	592445	592466	+
PF0580	hypothetical protein	upregulated	598701	598717	598712	+
PF0581	hypothetical protein	upregulated	601310	601327	601326	+
PF0648	hypothetical protein	downregulated	661502	661496	661497	+
PF0649	hypothetical protein	-				
PF0848	hypothetical protein	upregulated	-	-	-	
PF0888	hypothetical protein	upregulated	-	-	-	
PF0889	hypothetical protein	upregulated	860390	860374	860405	+
PF0907	hypothetical protein	-	879617	879613	879609	+
PF0908	hypothetical protein	upregulated				

gene/ operons	protein	ΔTrmBL2/WT (Waege, 2014)	ChIP-seq pyruvate 2	ChIP-seq pyruvate 5	ChIP-seq starch 2	Peak promoter/ intergenic
PF0921	ABC transporter	upregulated				
PF0922	hypothetical protein	upregulated				
PF0923	hypothetical protein	upregulated				
PF0924	hypothetical protein	-				
PF0925	heme biosynthesis protein	-				
PF0926	hypothetical protein	-	892081	892047	892079	+
PF0927	hypothetical protein	upregulated	893127	893147	893149	+
PF0956	hypothetical protein	upregulated	917902	917889	917900	-
PF0975	hypothetical protein	upregulated				
PF0976	putative trans-membrane protein	-				
PF0977	hypothetical protein	-				
PF0978	hypothetical protein	-				
PF0992	hypothetical protein	-				
PF0993	hypothetical protein	-				
PF0994	archaeal flagellar protein FlaI	upregulated				
PF0995	hypothetical protein	-				
PF0996	hypothetical protein	-				
PF1052	aspartate kinase	downregulated	1005810	1005902	1005904	+
PF1080	hypothetical protein	upregulated	1030175	1030166	1030162	+
PF1089	hypothetical protein	-				
PF1088	TFB-RF1	upregulated	-	-	-	
PF1100	hypothetical protein	upregulated	-	-	-	
PF1463	hypothetical protein	upregulated	1371172	1371164	1371170	+
PF1464	hypothetical protein	-				
PF1465	hypothetical protein	upregulated				
PF1466	hypothetical protein	upregulated				
PF1467	putative ABC transporter	-				
PF1513	hypothetical protein	upregulated	1412115	1412148	1412147	+
PF1514	hypothetical protein	-				
PF1739	trehalose/maltose binding protein	deleted	-	-	-	
PF1905	protease	upregulated	1755897	1755911	1755913	+
PF1906	DAPA aminotransferase	upregulated				
PF1983	hypothetical protein	upregulated	1833138	1833116	1833119	+
PF1984	hypothetical protein	upregulated				
PF2063	aminopeptidase	upregulated				
PF2064	arylsulfatase	upregulated	1906298	1906345	1906341	+

72 genes (51 showed the highest differential expression) were classified into 35 transcription units according to http://archaea.ucsc.edu and Yoon et al., 2011. The following units were not considered for further analysis: PF0398 – PF0400, only PF0400 differentially expressed; PF0496, deleted; PF0975 – PF0978, only PF0978 differentially expressed; PF0992 – PF0996, only PF0994 differentially expressed; PF1739, deleted; ΔTrmBL2, TrmBL2 deletion strain; WT, wildtype strain DSM3638; ChIP-seq pyruvate 2, pyruvate 5 and starch 2: TrmBL2 binding sites (consensus of replicates) under the corresponding growth condition; Peak promoter/intergenic, position of the TrmBL2 binding site.

6. Appendix

6.1. List of tables

6.2. List of figures

6.3. List of abbreviations

α	alpha
A_{578}	Absorbance at 578 nm
AA	amino acid
ADP	adenosine diphosphate
ArsR	arsenic resistance regulator
ATP	adenosine triphosphate
β	beta
bp	base pair
BSA	bovine serum albumin
CoA	coenzyme A
C_t	cycle threshold
C-terminal	carboxy-terminal
Cy5	cyanine 5
Da	Dalton
°C	degree celsius
Δ	delta
DNA	2'-deoxyribonucleic acid
DNase I	deoxyribonuclease I
dNTP	3'-deoxyribonucleoside-5'-triphosphate
DSMZ	*Deutsche Sammlung von Mikroorganismen und Zellkulturen*
E.	*Escherichia*
EDTA	ethylene diamine tetraacetic acid
e.g.	*exempli gratia*
EGTA	ethylene glycol tetraacetic acid
EMSA	electrophoretic gel mobility shift assay
et al.	et alii
Fis	factor for inversion stimulation
fla	*flagellin*
gDNA	genomic DNA
h	hour
H.	*Halobacterium*
H-NS	histone-like nucleoid structuring protein
IgG	immunglobulin G
IP	immunoprecipitation
IPTG	isopropyl-β-thiogalactopyranoside
kb	kilo base pairs
kDa	kilo Dalton
L	liter
LB	Luria-Broth medium
M	$mol*L^{-1}$
MalR	*mal* regulon activator
ml	milliliter
µ	mikro
min	minute
M. acetivorans	*Methanosarcina acetivorans*
M. jannaschii	*Methanocaldococcus jannaschii*
MreA	*Methanosarcina r*egulator of *e*nergy-converting metabolism
mRNA	messenger RNA
MW	molecular weight
n	number
ng	nanogram
N-terminal	amino-terminal

OD	optical density
ORF	open reading frame
P	p-value
P.	*Pyrococcus*
PadR	transcriptional regulator of phenolic acid decarboxylase
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
Poly-dIC	poly-desoxy-Cytosin-desoxy-Inosin
PMSF	phenylmethyl-sulphonyl fluoride
PVDF	polyvinylidenfluorid
rcf	relative centrifugal force
RNA	ribonucleic acid
RNAP	DNA dependent RNA-polymerase
RNaseA	ribonuclease A
ρ	rho
RT	room temperature
σ	sigma
S.	*Sulfolobus*
S. cerevisiae	*Saccharomyces cerevisiae*
S. PCC7942	*Synechococcus PCC7942*
SD	standard deviation
SDS	sodium dodecylsulphate
sec	second
SurR	sulfur response regulator
T.	*Thermococcus*
TBE	Tris-borate-EDTA buffer
TBS	Tris-buffered-saline
TE	Tris-Cl/EDTA (10:1)
tRNA	transfer RNA
U	units
UTP	uridine triphosphate
v/v	percentage volume to volume
w/v	percentage weight to volume
wt	wild type

in silico: performed on computer or via computer simulation

in vitro: performing a given procedure in a controlled environment outside of a living organism

in vivo: experimentation using a whole, living organism

The IUPAC (International Union of Pure and Applied Chemistry) codes for AAs (one letter and three letter code) and for nucleic acids were used.

7. Publication bibliography

References

Adams, M. W., Holden, J. F., Menon, A. L., Schut, G. J., Grunden, A. M., Hou, C., Hutchins, A. M., Jenney, F E Jr, Kim, C., Ma, K., Pan, G., Roy, R., Sapra, R., Story, S. V., and Verhagen, M. F., 'Key role for sulfur in peptide metabolism and in regulation of three hydrogenases in the hyperthermophilic archaeon Pyrococcus furiosus', *Journal of bacteriology*, Vol. 183, No. 2, 2001.

Albert, I., Wachi, S., Jiang, C., and Pugh, B. F., 'GeneTrack--a genomic data processing and visualization framework', *Bioinformatics (Oxford, England)*, Vol. 24, No. 10, 2008.

Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G. D., Giaever, G., and Nislow, C., 'Chromatin is an ancient innovation conserved between Archaea and Eukarya', *eLife*, Vol. 1, 2012.

Aparicio, O., Geisberg, J. V., Sekinger, E., Yang, A., Moqtaderi, Z., and Struhl, K., 'Chromatin immunoprecipitation for determining the association of proteins with specific genomic sequences *in vivo'*, *Current protocols in molecular biology / edited by Frederick M. Ausubel … [et al.]*, Chapter 21, 2005.

Aravind, L., Anantharaman, V., Balaji, S., Babu, M. M., and Iyer, L. M., 'The many faces of the helix-turn-helix domain: transcription regulation and beyond', *FEMS microbiology reviews*, Vol. 29, No. 2, 2005.

Aravind, L., and Koonin, E. V., 'DNA-binding proteins and evolution of transcription regulation in the archaea', *Nucleic acids research*, Vol. 27, No. 23, 1999.

Atomi, H., Imanaka, T., and Fukui, T., 'Overview of the genetic tools in the Archaea', *Frontiers in microbiology*, Vol. 3, 2012.

Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., *Current protocols in molecular biology,* Wiley, regularly updated.

Babu, M. M., Luscombe, N. M., Aravind, L., Gerstein, M., and Teichmann, S. A., 'Structure and evolution of transcriptional regulatory networks', *Current opinion in structural biology*, Vol. 14, No. 3, 2004.

Bailey, K. A., Pereira, S. L., Widom, J., and Reeve, J. N., 'Archaeal histone selection of nucleosome positioning sequences and the procaryotic origin of histone-dependent genome evolution', *Journal of molecular biology*, Vol. 303, No. 1, 2000.

Bailey, T., Krajewski, P., Ladunga, I., Lefebvre, C., Li, Q., Liu, T., Madrigal, P., Taslim, C., and Zhang, J., 'Practical guidelines for the comprehensive analysis of ChIP-seq data', *PLoS computational biology*, Vol. 9, No. 11, 2013.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W., and Noble, W. S., 'MEME Suite: tools for motif discovery and searching', *Nucleic Acids Research*, Vol. 37, suppl 2, 2009.

Baliga, N. S., Goo, Y. A., Ng, W. V., Hood, L., Daniels, C. J., and DasSarma, S., 'Is gene expression in *Halobacterium NRC-1* regulated by multiple TBP and TFB transcription factors?', *Molecular microbiology*, Vol. 36, No. 5, 2000.

Barski, A., Cuddapah, S., Cui, K., Roh, T.-Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K., 'High-resolution profiling of histone methylations in the human genome', *Cell*, Vol. 129, No. 4, 2007.

Bell, S. D., Brinkman, A. B., van Der Oost, J, and Jackson, S. P., 'The archaeal TFIIEalpha homologue facilitates transcription initiation by enhancing TATA-box recognition', *EMBO reports*, Vol. 2, No. 2, 2001a.

Bell, S. D., and Jackson, S. P., 'Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features', *Trends in microbiology*, Vol. 6, No. 6, 1998a.

——, 'Transcription in Archaea', *Cold Spring Harbor symposia on quantitative biology*, Vol. 63, 1998b.

——, 'Mechanism of autoregulation by an archaeal transcriptional repressor', *The Journal of biological chemistry*, Vol. 275, No. 41, 2000.

Bell, S. D., Magill, C. P., and Jackson, S. P., 'Basal and regulated transcription in Archaea', *Biochemical Society transactions*, Vol. 29, Pt 4, 2001b.

Blamey, J. M., and Adams, M. W., 'Purification and characterization of pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Biochimica et biophysica acta*, Vol. 1161, No. 1, 1993.

Blankenberg, D., Kuster, G. von, Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J., 'Galaxy: a web-based genome analysis tool for experimentalists', *Current protocols in molecular biology / edited by Frederick M. Ausubel … [et al.]*, Chapter 19, 2010.

Blat, Y., and Kleckner, N., 'Cohesins bind to preferential sites along yeast chromosome III, with differential regulation along arms versus the centric region', *Cell*, Vol. 98, No. 2, 1999.

Blum, H., Beier, H., and Gross, H. J., 'Improved silver staining of plant proteins, RNA and DNA in polyacrylamide gels', *ELECTROPHORESIS*, Vol. 8, No. 2, 1987.

Bochar, D. A., Brown, J. R., Doolittle, W. F., Klenk, H. P., Lam, W., Schenk, M. E., Stauffacher, C. V., and Rodwell, V. W., '3-hydroxy-3-methylglutaryl coenzyme A reductase of *Sulfolobus solfataricus*: DNA sequence, phylogeny, expression in *Escherichia coli* of the hmgA gene, and purification and kinetic characterization of the gene product', *Journal of bacteriology*, Vol. 179, No. 11, 1997.

Bouffartigues, E., Buckle, M., Badaut, C., Travers, A., and Rimsky, S., 'H-NS cooperative binding to high-affinity sites in a regulatory element results in transcriptional silencing', *Nature structural & molecular biology*, Vol. 14, No. 5, 2007.

Bräsen, C., Esser, D., Rauch, B., and Siebers, B., 'Carbohydrate metabolism in archaea: current insights into unusual enzymes and pathways and their regulation', *Microbiology and molecular biology reviews : MMBR*, Vol. 78, No. 1, 2014.

Bridger, S. L., Lancaster, W. A., Poole, Farris L 2nd, Schut, G. J., and Adams, Michael W W, 'Genome sequencing of a genetically tractable Pyrococcus furiosus strain reveals a highly dynamic genome', *Journal of bacteriology*, Vol. 194, No. 15, 2012.

Brooks, A. N., Turkarslan, S., Beer, K. D., Lo, F. Y., and Baliga, N. S., 'Adaptation of cells to new environments', *Wiley interdisciplinary reviews. Systems biology and medicine*, Vol. 3, No. 5, 2011.

Brown, S. H., Costantino, H. R., and Kelly, R. M., 'Characterization of Amylolytic Enzyme Activities Associated with the Hyperthermophilic Archaebacterium *Pyrococcus furiosus'*, *Applied and environmental microbiology*, Vol. 56, No. 7, 1990.

Cabrera, J. A., Bolds, J., Shields, P. E., Havel, C. M., and Watson, J. A., 'Isoprenoid synthesis in *Halobacterium halobium*. Modulation of 3-hydroxy-3-methylglutaryl coenzyme a concentration in response to mevalonate availability', *The Journal of biological chemistry*, Vol. 261, No. 8, 1986.

Cardenas, M. L., Cornish-Bowden, A., and Ureta, T., 'Evolution and regulatory role of the hexokinases', *Biochimica et biophysica acta*, Vol. 1401, No. 3, 1998.

Chan, P. P., Holmes, A. D., Smith, A. M., Tran, D., and Lowe, T. M., 'The UCSC Archaeal Genome Browser: 2012 update', *Nucleic acids research*, Vol. 40, Database issue, 2012.

Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C., and Doolittle, W. F., 'Transformation methods for halophilic archaebacteria', *Canadian journal of microbiology*, Vol. 35, No. 1, 1989.

Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J.-C., Van Der Oost, John, Weissenbach, J., Zivanovic, Y., and Forterre, P., 'An integrated analysis of the genome of the hyperthermophilic archaeon *Pyrococcus abyssi'*, *Molecular microbiology*, Vol. 47, No. 6, 2003.

Comfort, D. A., Chou, C.-J., Conners, S. B., Vanfossen, A. L., and Kelly, R. M., 'Functional-genomics-based identification and characterization of open reading frames encoding alpha-glucoside-processing enzymes in the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Applied and environmental microbiology*, Vol. 74, No. 4, 2008.

Cook, W. J., Kar, S. R., Taylor, K. B., and Hall, L. M., 'Crystal structure of the cyanobacterial metallothionein repressor SmtB: a model for metalloregulatory proteins', *Journal of molecular biology*, Vol. 275, No. 2, 1998.

Črnigoj, M., Podlesek, Z., Zorko, M., Jerala, R., Anderluh, G., and Ulrih, N. P., 'Interactions of Archaeal Chromatin Proteins Alba1 and Alba2 with Nucleic Acids', *PloS one*, Vol. 8, No. 2, 2013.

Dahlke, I., and Thomm, M., 'A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LrpA, inhibits transcription by abrogating RNA polymerase recruitment', *Nucleic acids research*, Vol. 30, No. 3, 2002.

Darst, S. A., 'Bacterial RNA polymerase', *Current opinion in structural biology*, Vol. 11, No. 2, 2001.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J., 'A framework for variation discovery and genotyping using next-generation DNA sequencing data', *Nature genetics*, Vol. 43, No. 5, 2011.

DiRuggiero, J., Dunn, D., Maeder, D. L., Holley-Shanks, R., Chatard, J., Horlacher, R., Robb, F. T., Boos, W., and Weiss, R. B., 'Evidence of recent lateral gene transfer among hyperthermophilic archaea', *Molecular microbiology*, Vol. 38, No. 4, 2000.

DiTommaso, P., Moretti, S., Xenarios, I., Orobitg, M., Montanyola, A., Chang, J.-M., Taly, J.-F., and Notredame, C., 'T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension', *Nucleic acids research*, Vol. 39, Web Server issue, 2011.

Dorman, C. J., 'Co-operative roles for DNA supercoiling and nucleoid-associated proteins in the regulation of bacterial transcription', *Biochemical Society transactions*, Vol. 41, No. 2, 2013.

Dorr, C., Zaparty, M., Tjaden, B., Brinkmann, H., and Siebers, B., 'The hexokinase of the hyperthermophile *Thermoproteus tenax*. ATP-dependent hexokinases and ADP-dependent glucokinases, teo alternatives for glucose phosphorylation in Archaea', *The Journal of biological chemistry*, Vol. 278, No. 21, 2003.

Du, J., Say, R. F., Lu, W., Fuchs, G., and Einsle, O., 'Active-site remodelling in the bifunctional fructose-1,6-bisphosphate aldolase/phosphatase', *Nature*, Vol. 478, No. 7370, 2011.

Duc, T. N., Hassanzadeh-Ghassabeh, G., Saerens, D., Peeters, E., Charlier, D., and Muyldermans, S., 'Nanobody-based chromatin immunoprecipitation', *Methods in molecular biology (Clifton, N.J.)*, Vol. 911, 2012.

Eisenschink, S., *Analyse von DNA-Protein-Wechselwirkungen verschiedener Transkriptionskomponenten mittels Chromatin Immunpraezipitations-experimenten in Pyrococcus furiosus*, 2010. Diplomarbeit.

Escobar-Paramo, P., Ghosh, S., and DiRuggiero, J., 'Evidence for genetic drift in the diversification of a geographically isolated population of the hyperthermophilic archaeon Pyrococcus', *Molecular biology and evolution*, Vol. 22, No. 11, 2005.

Evan, G. I., Lewis, G. K., Ramsay, G., and Bishop, J. M., 'Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product', *Molecular and cellular biology*, Vol. 5, No. 12, 1985.

Facciotti, M. T., Reiss, D. J., Pan, M., Kaur, A., Vuthoori, M., Bonneau, R., Shannon, P., Srivastava, A., Donohoe, S. M., Hood, L. E., and Baliga, N. S., 'General transcription factor specified global gene regulation in archaea', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 104, No. 11, 2007.

Favorov, A. V., Gelfand, M. S., Gerasimova, A. V., Ravcheev, D. A., Mironov, A. A., and Makeev, V. J., 'A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length', *Bioinformatics (Oxford, England)*, Vol. 21, No. 10, 2005.

Fiala, G., and Stetter, K., '*Pyrococcus furiosus sp. nov.* represents a novel genus of marine heterotrophic archaebacteria growing optimally at 100°C', *Archives of Microbiology*, Vol. 145, No. 1, 1986.

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A., and Wigler, M., 'Purification of a RAS-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method', *Molecular and cellular biology*, Vol. 8, No. 5, 1988.

Finn, R. D., Mistry, J., Schuster-Bockler, B., Griffiths-Jones, S., Hollich, V., Lassmann, T., Moxon, S., Marshall, M., Khanna, A., Durbin, R., Eddy, S. R., Sonnhammer, Erik L L, and Bateman, A., 'Pfam: clans, web tools and services', *Nucleic acids research*, Vol. 34, Database issue, 2006.

Fouqueau, T., Zeller, M. E., Cheung, A. C., Cramer, P., and Thomm, M., 'The RNA polymerase trigger loop functions in all three phases of the transcription cycle', *Nucleic acids research*, Vol. 41, No. 14, 2013.

Fushinobu, S., Nishimasu, H., Hattori, D., Song, H.-J., and Wakagi, T., 'Structural basis for the bifunctionality of fructose-1,6-bisphosphate aldolase/phosphatase', *Nature*, Vol. 478, No. 7370, 2011.

Gao, J., and Wang, J., 'Re-annotation of two hyperthermophilic archaea *Pyrococcus abyssi GE5* and *Pyrococcus furiosus DSM 3638'*, *Current microbiology*, Vol. 64, No. 2, 2012.

Gascuel, O., 'BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data', *Molecular biology and evolution*, Vol. 14, No. 7, 1997.

Gindner, A., *Untersuchungen zur Funktion der Transkriptionsregulator-Familie TrmB und anderen Regulatoren in Pyrococcus furiosus*, 2009. Diplomarbeit.

Goede, B., Naji, S., Kampen, O. von, Ilg, K., and Thomm, M., 'Protein-protein interactions in the archaeal transcriptional machinery: binding studies of isolated RNA polymerase

subunits and transcription factors', *The Journal of biological chemistry*, Vol. 281, No. 41, 2006.

Grabowski, B., and Kelman, Z., 'Archeal DNA replication: eukaryal proteins in a bacterial context', *Annual review of microbiology*, Vol. 57, 2003.

Grohmann, D., Hirtreiter, A., and Werner, F., 'Molecular mechanisms of archaeal RNA polymerase', *Biochemical Society transactions*, Vol. 37, Pt 1, 2009.

Grohmann, D., and Werner, F., 'Recent advances in the understanding of archaeal transcription', *Current opinion in microbiology*, Vol. 14, No. 3, 2011.

Grünberg, S., Bartlett, M. S., Naji, S., and Thomm, M., 'Transcription factor E is a part of transcription elongation complexes', *The Journal of biological chemistry*, Vol. 282, No. 49, 2007.

Hall, T. A., 'BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT', *Nucleic Acids Symposium Series*, Vol. 41, 1999.

Hamilton-Brehm, S. D., Schut, G. J., and Adams, Michael W W, 'Metabolic and evolutionary relationships among *Pyrococcus* Species: genetic exchange within a hydrothermal vent environment', *Journal of bacteriology*, Vol. 187, No. 21, 2005.

Hanzelka, B. L., Darcy, T. J., and Reeve, J. N., 'TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIEalpha', *Journal of bacteriology*, Vol. 183, No. 5, 2001.

Hausner, W., and Thomm, M., 'Events during initiation of archaeal transcription: open complex formation and DNA-protein interactions', *Journal of bacteriology*, Vol. 183, No. 10, 2001.

Hausner, W., Wettach, J., Hethke, C., and Thomm, M., 'Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase', *The Journal of biological chemistry*, Vol. 271, No. 47, 1996.

Hecht, A., Strahl-Bolsinger, S., and Grunstein, M., 'Spreading of transcriptional repressor SIR3 from telomeric heterochromatin', *Nature*, Vol. 383, No. 6595, 1996.

Hecker, M., Pane-Farre, J., and Volker, U., 'SigB-dependent general stress response in *Bacillus subtilis* and related gram-positive bacteria', *Annual review of microbiology*, Vol. 61, 2007.

Heider, J., Mai, X., and Adams, M. W., 'Characterization of 2-ketoisovalerate ferredoxin oxidoreductase, a new and reversible coenzyme A-dependent enzyme involved in peptide fermentation by hyperthermophilic archaea', *Journal of bacteriology*, Vol. 178, No. 3, 1996.

Herzog, B., and Wirth, R., 'Swimming behavior of selected species of Archaea', *Applied and environmental microbiology*, Vol. 78, No. 6, 2012.

Hethke, C., Geerling, A. C., Hausner, W., de Vos, W M, and Thomm, M., 'A cell-free transcription system for the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Nucleic acids research*, Vol. 24, No. 12, 1996.

Hidese, R., Nishikawa, R., Le Gao, Katano, M., Imai, T., Kato, S., Kanai, T., Atomi, H., Imanaka, T., and Fujiwara, S., 'Different roles of two transcription factor B proteins in the hyperthermophilic archaeon *Thermococcus kodakarensis'*, *Extremophiles : life under extreme conditions*, Vol. 18, No. 3, 2014.

Hillis, D. M., and Bull, J. J., 'An Empirical Test of Bootstrapping as a Method for Assessing Confidence in Phylogenetic Analysis', *Systematic Biology*, Vol. 42, No. 2, 1993.

Hirata, A., Kanai, T., Santangelo, T. J., Tajiri, M., Manabe, K., Reeve, J. N., Imanaka, T., and Murakami, K. S., 'Archaeal RNA polymerase subunits E and F are not required for

transcription in vitro, but a *Thermococcus kodakarensis* mutant lacking subunit F is temperature-sensitive', *Molecular microbiology*, Vol. 70, No. 3, 2008a.

Hirata, A., Klein, B. J., and Murakami, K. S., 'The X-ray crystal structure of RNA polymerase from Archaea', *Nature*, Vol. 451, No. 7180, 2008b.

Hochuli, E., Bannwarth, W., Dobeli, H., Gentz, R., and Stuber, D., 'Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent', *Nat Biotech*, Vol. 6, No. 11, 1988.

Hopp, T. P., Prickett, K. S., Price, V. L., Libby, R. T., March, C. J., Pat Cerretti, D., Urdal, D. L., and Conlon, P. J., 'A Short Polypeptide Marker Sequence Useful for Recombinant Protein Identification and Purification', *Nat Biotech*, Vol. 6, No. 10, 1988.

Horlacher, R., Xavier, K. B., Santos, H., DiRuggiero, J., Kossmann, M., and Boos, W., 'Archaeal binding protein-dependent ABC transporter: molecular and biochemical analysis of the trehalose/maltose transport system of the hyperthermophilic archaeon *Thermococcus litoralis'*, *Journal of bacteriology*, Vol. 180, No. 3, 1998.

Huber, M., *Analysen von Protein-DNA-Wechselwirkungen über Crosslinking- und SELEX-Experimente*, 2011. Master thesis.

Huet, J., Schnabel, R., Sentenac, A., and Zillig, W., 'Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type', *The EMBO journal*, Vol. 2, No. 8, 1983.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T. K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., Castro, E. de, Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R. D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A. F., Rivoire, C., Sangrador-Vegas, A., Selengut, J. D., Sigrist, Christian J A, Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P. D., Wu, C. H., Yeats, C., and Yong, S.-Y., 'InterPro in 2011: new developments in the family and domain prediction database', *Nucleic acids research*, Vol. 40, Database issue, 2012.

Huson, D. H., and Bryant, D., 'Application of phylogenetic networks in evolutionary studies', *Molecular biology and evolution*, Vol. 23, No. 2, 2006.

Imanaka, H., Yamatsu, A., Fukui, T., Atomi, H., and Imanaka, T., 'Phosphoenolpyruvate synthase plays an essential role for glycolysis in the modified Embden-Meyerhof pathway in *Thermococcus kodakarensis'*, *Molecular microbiology*, Vol. 61, No. 4, 2006.

Jasiak, A. J., Hartmann, H., Karakasili, E., Kalocsay, M., Flatley, A., Kremmer, E., Strasser, K., Martin, D. E., Soding, J., and Cramer, P., 'Genome-associated RNA polymerase II includes the dissociable Rpb4/7 subcomplex', *The Journal of biological chemistry*, Vol. 283, No. 39, 2008.

Johnson, D. S., Mortazavi, A., Myers, R. M., and Wold, B., 'Genome-wide mapping of *in vivo* protein-DNA interactions', *Science (New York, N.Y.)*, Vol. 316, No. 5830, 2007.

Jones, D. T., Taylor, W. R., and Thornton, J. M., 'The rapid generation of mutation data matrices from protein sequences', *Computer applications in the biosciences : CABIOS*, Vol. 8, No. 3, 1992.

Jorgensen, S., Vorgias, C. E., and Antranikian, G., 'Cloning, sequencing, characterization, and expression of an extracellular alpha-amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis'*, *The Journal of biological chemistry*, Vol. 272, No. 26, 1997.

Kahramanoglou, C., Seshasayee, Aswin S N, Prieto, A. I., Ibberson, D., Schmidt, S., Zimmermann, J., Benes, V., Fraser, G. M., and Luscombe, N. M., 'Direct and indirect effects of H-NS and Fis on global gene expression control in *Escherichia coli'*, *Nucleic acids research*, Vol. 39, No. 6, 2011.

Kanai, T., Akerboom, J., Takedomi, S., van de Werken, Harmen, Blombach, F., Van Der Oost, John, Murakami, T., Atomi, H., and Imanaka, T., 'A global transcriptional regulator in *Thermococcus kodakaraensis* controls the expression levels of both glycolytic and gluconeogenic enzyme-encoding genes', *The Journal of biological chemistry*, Vol. 282, No. 46, 2007.

Kanai, T., Takedomi, S., Fujiwara, S., Atomi, H., and Imanaka, T., 'Identification of the Phr-dependent heat shock regulon in the hyperthermophilic archaeon, *Thermococcus kodakaraensis'*, *Journal of biochemistry*, Vol. 147, No. 3, 2010.

Kandler, O., and Hippe, H., 'Lack of peptidoglycan in the cell walls of *Methanosarcina barkeri'*, *Archives of microbiology*, Vol. 113, 1-2, 1977.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba, S., Kosugi, H., Hosoyama, A., Nagai, Y., Sakai, M., Ogura, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., and Kikuchi, H., 'Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii OT3'*, *DNA research : an international journal for rapid publication of reports on genes and genomes*, Vol. 5, No. 2, 1998.

Keese, A. M., Schut, G. J., Ouhammouch, M., Adams, Michael W W, and Thomm, M., 'Genome-wide identification of targets for the archaeal heat shock regulator phr by cell-free transcription of genomic DNA', *Journal of bacteriology*, Vol. 192, No. 5, 2010.

Kengen, S. W., de Bok, F A, van Loo, N D, Dijkema, C., Stams, A. J., and de Vos, W M, 'Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus'*, *The Journal of biological chemistry*, Vol. 269, No. 26, 1994.

Kengen, S. W., Tuininga, J. E., de Bok, F A, Stams, A. J., and de Vos, W M, 'Purification and characterization of a novel ADP-dependent glucokinase from the hyperthermophilic archaeon *Pyrococcus furiosus'*, *The Journal of biological chemistry*, Vol. 270, No. 51, 1995.

Kibbe, W. A., 'OligoCalc: an online oligonucleotide properties calculator', *Nucleic acids research*, Vol. 35, Web Server issue, 2007.

Koide, T., Reiss, D. J., Bare, J. C., Pang, W. L., Facciotti, M. T., Schmid, A. K., Pan, M., Marzolf, B., Van, P. T., Lo, F.-Y., Pratap, A., Deutsch, E. W., Peterson, A., Martin, D., and Baliga, N. S., 'Prevalence of transcription promoters within archaeal operons and coding sequences', *Molecular systems biology*, Vol. 5, 2009.

Koning, S. M., Konings, W. N., and Driessen, Arnold J M, 'Biochemical evidence for the presence of two alpha-glucoside ABC-transport systems in the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Archaea (Vancouver, B.C.)*, Vol. 1, No. 1, 2002.

Koonin, E. V., Makarova, K. S., and Elkins, J. G., 'Orthologs of the small RPB8 subunit of the eukaryotic RNA polymerases are conserved in hyperthermophilic Crenarchaeota and "Korarchaeota"', *Biology direct*, Vol. 2, 2007.

Koonin, E. V., and Wolf, Y. I., 'Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world', *Nucleic acids research*, Vol. 36, No. 21, 2008.

Korkhin, Y., Unligil, U. M., Littlefield, O., Nelson, P. J., Stuart, D. I., Sigler, P. B., Bell, S. D., and Abrescia, Nicola G A, 'Evolution of complex RNA polymerases: the complete archaeal RNA polymerase structure', *PLoS biology*, Vol. 7, No. 5, 2009.

Kostrewa, D., Zeller, M. E., Armache, K.-J., Seizl, M., Leike, K., Thomm, M., and Cramer, P., 'RNA polymerase II-TFIIB structure and mechanism of transcription initiation', *Nature*, Vol. 462, No. 7271, 2009.

Kreuzer, M., *Untersuchungen zur Transformation von Pyrococcus furiosus*, 2009. Bachelor thesis.

Kreuzer, M., Schmutzler, K., Waege, I., Thomm, M., and Hausner, W., 'Genetic engineering of Pyrococcus furiosus to use chitin as a carbon source', *BMC biotechnology*, Vol. 13, 2013.

Krug, M., Lee, S.-J., Boos, W., Diederichs, K., and Welte, W., 'The three-dimensional structure of TrmB, a transcriptional regulator of dual function in the hyperthermophilic archaeon *Pyrococcus furiosus* in complex with sucrose', *Protein science : a publication of the Protein Society*, Vol. 22, No. 6, 2013.

Krug, M., Lee, S.-J., Diederichs, K., Boos, W., and Welte, W., 'Crystal structure of the sugar binding domain of the archaeal transcriptional regulator TrmB', *The Journal of biological chemistry*, Vol. 281, No. 16, 2006.

Laemmli, U. K., 'Cleavage of structural proteins during the assembly of the head of bacteriophage T4', *Nature*, Vol. 227, No. 5259, 1970.

Laksanalamai, P., Maeder, D. L., and Robb, F. T., 'Regulation and mechanism of action of the small heat shock protein from the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Journal of bacteriology*, Vol. 183, No. 17, 2001.

Lam, W. L., and Doolittle, W. F., 'Shuttle vectors for the archaebacterium *Halobacterium volcanii'*, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 86, No. 14, 1989.

Landt, S. G., Marinov, G. K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B. E., Bickel, P., Brown, J. B., Cayting, P., Chen, Y., DeSalvo, G., Epstein, C., Fisher-Aylor, K. I., Euskirchen, G., Gerstein, M., Gertz, J., Hartemink, A. J., Hoffman, M. M., Iyer, V. R., Jung, Y. L., Karmakar, S., Kellis, M., Kharchenko, P. V., Li, Q., Liu, T., Liu, X. S., Ma, L., Milosavljevic, A., Myers, R. M., Park, P. J., Pazin, M. J., Perry, M. D., Raha, D., Reddy, T. E., Rozowsky, J., Shoresh, N., Sidow, A., Slattery, M., Stamatoyannopoulos, J. A., Tolstorukov, M. Y., White, K. P., Xi, S., Farnham, P. J., Lieb, J. D., Wold, B. J., and Snyder, M., 'ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia', *Genome research*, Vol. 22, No. 9, 2012.

Lane, W. J., and Darst, S. A., 'Molecular evolution of multisubunit RNA polymerases: sequence analysis', *Journal of molecular biology*, Vol. 395, No. 4, 2010.

Langer, D., Hain, J., Thuriaux, P., and Zillig, W., 'Transcription in archaea: similarity to that in eucarya', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 92, No. 13, 1995.

Langmead, B., and Salzberg, S. L., 'Fast gapped-read alignment with Bowtie 2', *Nature methods*, Vol. 9, No. 4, 2012.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L., 'Ultrafast and memory-efficient alignment of short DNA sequences to the human genome', *Genome biology*, Vol. 10, No. 3, 2009.

Langworthy, T. A., Mayberry, W. R., and Smith, P. F., 'Long-chain glycerol diether and polyol dialkyl glycerol triether lipids of *Sulfolobus acidocaldarius'*, *Journal of bacteriology*, Vol. 119, No. 1, 1974.

Lee, H. S., Bae, S. S., Kim, M.-S., Kwon, K. K., Kang, S. G., and Lee, J.-H., 'Complete genome sequence of hyperthermophilic *Pyrococcus sp. strain NA2*, isolated from a deep-sea hydrothermal vent area', *Journal of bacteriology*, Vol. 193, No. 14, 2011.

Lee, H.-S., Shockley, K. R., Schut, G. J., Conners, S. B., Montero, C. I., Johnson, M. R., Chou, C.-J., Bridger, S. L., Wigner, N., Brehm, S. D., Jenney, Francis E Jr, Comfort, D. A., Kelly, R. M., and Adams, Michael W W, 'Transcriptional and biochemical analysis of starch metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Journal of bacteriology*, Vol. 188, No. 6, 2006a.

Lee, S.-J., Bohm, A., Krug, M., and Boos, W., 'The ABC of binding-protein-dependent transport in Archaea', *Trends in microbiology*, Vol. 15, No. 9, 2007a.

Lee, S.-J., Engelmann, A., Horlacher, R., Qu, Q., Vierke, G., Hebbeln, C., Thomm, M., and Boos, W., 'TrmB, a sugar-specific transcriptional regulator of the trehalose/maltose ABC transporter from the hyperthermophilic archaeon *Thermococcus litoralis'*, *The Journal of biological chemistry*, Vol. 278, No. 2, 2003.

Lee, S.-J., Moulakakis, C., Koning, S. M., Hausner, W., Thomm, M., and Boos, W., 'TrmB, a sugar sensing regulator of ABC transporter genes in *Pyrococcus furiosus* exhibits dual promoter specificity and is controlled by different inducers', *Molecular microbiology*, Vol. 57, No. 6, 2005.

Lee, S.-J., Surma, M., Hausner, W., Thomm, M., and Boos, W., 'The role of TrmB and TrmB-like transcriptional regulators for sugar transport and metabolism in the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Archives of microbiology*, Vol. 190, No. 3, 2008.

Lee, S.-J., Surma, M., Seitz, S., Hausner, W., Thomm, M., and Boos, W., 'Characterization of the TrmB-like protein, PF0124, a TGM-recognizing global transcriptional regulator of the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Molecular microbiology*, Vol. 65, No. 2, 2007b.

——, 'Differential signal transduction via TrmB, a sugar sensing transcriptional repressor of *Pyrococcus furiosus'*, *Molecular microbiology*, Vol. 64, No. 6, 2007c.

Lee, T. I., Johnstone, S. E., and Young, R. A., 'Chromatin immunoprecipitation and microarray-based analysis of protein location', *Nature protocols*, Vol. 1, No. 2, 2006b.

Lefrancois, P., Euskirchen, G. M., Auerbach, R. K., Rozowsky, J., Gibson, T., Yellman, C. M., Gerstein, M., and Snyder, M., 'Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing', *BMC genomics*, Vol. 10, 2009.

Leigh, J. A., Albers, S.-V., Atomi, H., and Allers, T., 'Model organisms for genetics in the domain Archaea: methanogens, halophiles, *Thermococcales* and *Sulfolobales'*, *FEMS microbiology reviews*, Vol. 35, No. 4, 2011.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R., 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics (Oxford, England)*, Vol. 25, No. 16, 2009.

Lipscomb, G. L., Keese, A. M., Cowart, D. M., Schut, G. J., Thomm, M., Adams, Michael W W, and Scott, R. A., 'SurR: a transcriptional activator and repressor controlling hydrogen and elemental sulphur metabolism in *Pyrococcus furiosus'*, *Molecular microbiology*, Vol. 71, No. 2, 2009.

Lipscomb, G. L., Stirrett, K., Schut, G. J., Yang, F., Jenney, Francis E Jr, Scott, R. A., Adams, Michael W W, and Westpheling, J., 'Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases', *Applied and environmental microbiology*, Vol. 77, No. 7, 2011.

Liu, K., Linder, C. R., and Warnow, T., 'RAxML and FastTree: comparing two methods for large-scale maximum likelihood phylogeny estimation', *PloS one*, Vol. 6, No. 11, 2011.

Liu, W., Vierke, G., Wenke, A.-K., Thomm, M., and Ladenstein, R., 'Crystal structure of the archaeal heat shock regulator from *Pyrococcus furiosus*: a molecular chimera representing eukaryal and bacterial features', *Journal of molecular biology*, Vol. 369, No. 2, 2007.

Lu, Q., Han, J., Zhou, L., Coker, J. A., DasSarma, P., DasSarma, S., and Xiang, H., 'Dissection of the regulatory mechanism of a heat-shock responsive promoter in Haloarchaea: a new paradigm for general transcription factor directed archaeal gene regulation', *Nucleic acids research*, Vol. 36, No. 9, 2008.

Mahillon, J., and Chandler, M., 'Insertion sequences', *Microbiology and molecular biology reviews : MMBR*, Vol. 62, No. 3, 1998.

Malone, B. M., Tan, F., Bridges, S. M., and Peng, Z., 'Comparison of four ChIP-Seq analytical algorithms using rice endosperm H3K27 trimethylation profiling data', *PloS one*, Vol. 6, No. 9, 2011.

Mankin, A. S., Zyrianova, I. M., Kagramanova, V. K., and Garrett, R. A., 'Introducing mutations into the single-copy chromosomal 23S rRNA gene of the archaeon *Halobacterium halobium* by using an rRNA operon-based transformation system', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 89, No. 14, 1992.

Maruyama, H., Shin, M., Oda, T., Matsumi, R., Ohniwa, R. L., Itoh, T., Shirahige, K., Imanaka, T., Atomi, H., Yoshimura, S. H., and Takeyasu, K., 'Histone and TK0471/TrmBL2 form a novel heterogeneous genome architecture in the hyperthermophilic archaeon *Thermococcus kodakarensis'*, *Molecular biology of the cell*, Vol. 22, No. 3, 2011.

Matsubara, K., Yokooji, Y., Atomi, H., and Imanaka, T., 'Biochemical and genetic characterization of the three metabolic routes in *Thermococcus kodakarensis* linking glyceraldehyde 3-phosphate and 3-phosphoglycerate', *Molecular microbiology*, Vol. 81, No. 5, 2011.

Matsumi, R., Manabe, K., Fukui, T., Atomi, H., and Imanaka, T., 'Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance', *Journal of bacteriology*, Vol. 189, No. 7, 2007.

Matsunaga, F., Forterre, P., Ishino, Y., and Myllykallio, H., '*In vivo* interactions of archaeal Cdc6/Orc1 and minichromosome maintenance proteins with the replication origin', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 98, No. 20, 2001.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A., 'The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data', *Genome research*, Vol. 20, No. 9, 2010.

Meinhart, A., Blobel, J., and Cramer, P., 'An extended winged helix domain in general transcription factor E/IIE alpha', *The Journal of biological chemistry*, Vol. 278, No. 48, 2003.

Meyer, T. S., and Lamberts, B. L., 'Use of coomassie brilliant blue R250 for the electrophoresis of microgram quantities of parotid saliva proteins on acrylamide-gel strips', *Biochimica et biophysica acta*, Vol. 107, No. 1, 1965.

Micorescu, M., Grünberg, S., Franke, A., Cramer, P., Thomm, M., and Bartlett, M., 'Archaeal transcription: function of an alternative transcription factor B from *Pyrococcus furiosus'*, *Journal of bacteriology*, Vol. 190, No. 1, 2008.

Mooney, R. A., Davis, S. E., Peters, J. M., Rowland, J. L., Ansari, A. Z., and Landick, R., 'Regulator trafficking on bacterial transcription units *in vivo'*, *Molecular cell*, Vol. 33, No. 1, 2009.

Morby, A. P., Turner, J. S., Huckle, J. W., and Robinson, N. J., 'SmtB is a metal-dependent repressor of the cyanobacterial metallothionein gene smtA: identification of a Zn inhibited DNA-protein complex', *Nucleic acids research*, Vol. 21, No. 4, 1993.

Mukund, S., and Adams, M. W., 'Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus'*, *The Journal of biological chemistry*, Vol. 270, No. 15, 1995.

Naji, S., Grünberg, S., and Thomm, M., 'The RPB7 orthologue E' is required for transcriptional activity of a reconstituted archaeal core enzyme at low temperatures and stimulates open complex formation', *The Journal of biological chemistry*, Vol. 282, No. 15, 2007.

Nalabothula, N., Xi, L., Bhattacharyya, S., Widom, J., Wang, J.-P., Reeve, J. N., Santangelo, T. J., and Fondufe-Mittendorf, Y. N., 'Archaeal nucleosome positioning *in vivo* and *in vitro* is directed by primary sequence motifs', *BMC genomics*, Vol. 14, 2013.

Näther, D. J., *Untersuchung der Flagellen von Pyrococcus furiosus*, 2008. http://epub.uni-regensburg.de/10717/.

Näther, D. J., Rachel, R., Wanner, G., and Wirth, R., 'Flagella of *Pyrococcus furiosus*: multifunctional organelles, made for swimming, adhesion to various surfaces, and cell-cell contacts', *Journal of bacteriology*, Vol. 188, No. 19, 2006.

Nguyen-Duc, T., Peeters, E., Muyldermans, S., Charlier, D., and Hassanzadeh-Ghassabeh, G., 'Nanobody(R)-based chromatin immunoprecipitation/micro-array analysis for genome-wide identification of transcription factor DNA binding sites', *Nucleic acids research*, Vol. 41, No. 5, 2013.

Nonaka, G., Blankschien, M., Herman, C., Gross, C. A., and Rhodius, V. A., 'Regulon and promoter analysis of the *E. coli* heat-shock factor, sigma32, reveals a multifaceted cellular response to heat stress', *Genes & development*, Vol. 20, No. 13, 2006.

Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., Arkin, A. P., Mironov, A. A., and Dubchak, I., 'RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach', *Nucleic acids research*, Vol. 38, Web Server issue, 2010.

Ochs, S., *Biochemische Analyse von putativen Transkriptionsregulatoren in Pyrococcus furiosus*, 2009. Diplomarbeit.

Ochs, S. M., Thumann, S., Richau, R., Weirauch, M. T., Lowe, T. M., Thomm, M., and Hausner, W., 'Activation of archaeal transcription mediated by recruitment of transcription factor B', *The Journal of biological chemistry*, Vol. 287, No. 22, 2012.

Okada, U., Sakai, N., Yao, M., Watanabe, N., and Tanaka, I., 'Structural analysis of the transcriptional regulator homolog protein from *Pyrococcus horikoshii OT3'*, *Proteins*, Vol. 63, No. 4, 2006.

Orlando, V., and Paro, R., 'Mapping Polycomb-repressed domains in the bithorax complex using *in vivo* formaldehyde cross-linked chromatin', *Cell*, Vol. 75, No. 6, 1993.

Orlando, V., Strutt, H., and Paro, R., 'Analysis of chromatin structure by *in vivo* formaldehyde cross-linking', *Methods (San Diego, Calif.)*, Vol. 11, No. 2, 1997.

Ouhammouch, M., Dewhurst, R. E., Hausner, W., Thomm, M., and Geiduschek, E. P., 'Activation of archaeal transcription by recruitment of the TATA-binding protein', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No. 9, 2003.

Ouhammouch, M., and Geiduschek, E. P., 'A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon *Methanococcus jannaschii'*, *The EMBO journal*, Vol. 20, 1-2, 2001.

Ouhammouch, M., Langham, G. E., Hausner, W., Simpson, A. J., El-Sayed, Najib M A, and Geiduschek, E. P., 'Promoter architecture and response to a positive regulator of archaeal transcription', *Molecular microbiology*, Vol. 56, No. 3, 2005.

Page, R. D., 'TreeView: an application to display phylogenetic trees on personal computers', *Computer applications in the biosciences : CABIOS*, Vol. 12, No. 4, 1996.

Park, P. J., 'ChIP-seq: advantages and challenges of a maturing technology', *Nature reviews. Genetics*, Vol. 10, No. 10, 2009.

Peeters, E., Peixeiro, N., and Sezonov, G., 'Cis-regulatory logic in archaeal transcription', *Biochemical Society transactions*, Vol. 41, No. 1, 2013.

Pepke, S., Wold, B., and Mortazavi, A., 'Computation for ChIP-seq and RNA-seq studies', *Nature methods*, Vol. 6, 11 Suppl, 2009.

Perez-Rueda, E., and Janga, S. C., 'Identification and genomic analysis of transcription factors in archaeal genomes exemplifies their functional architecture and evolutionary origin', *Molecular biology and evolution*, Vol. 27, No. 6, 2010.

Pfaffl, M. W., 'A new mathematical model for relative quantification in real-time RT-PCR', *Nucleic acids research*, Vol. 29, No. 9, 2001.

Poole, F. L., Gerwe, B. A., Hopkins, R. C., Schut, G. J., Weinberg, M. V., Jenney, Francis E Jr, and Adams, Michael W W, 'Defining genes in the genome of the hyperthermophilic archaeon *Pyrococcus furiosus*: implications for all microbial genomes', *Journal of bacteriology*, Vol. 187, No. 21, 2005.

Price, M. N., Dehal, P. S., and Arkin, A. P., 'FastTree 2--approximately maximum-likelihood trees for large alignments', *PloS one*, Vol. 5, No. 3, 2010.

Qureshi, S. A., Baumann, P., Rowlands, T., Khoo, B., and Jackson, S. P., 'Cloning and functional analysis of the TATA binding protein from *Sulfolobus shibatae'*, *Nucleic acids research*, Vol. 23, No. 10, 1995a.

Qureshi, S. A., Khoo, B., Baumann, P., and Jackson, S. P., 'Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae'*, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 92, No. 13, 1995b.

Rashid, N. U., Giresi, P. G., Ibrahim, J. G., Sun, W., and Lieb, J. D., 'ZINBA integrates local covariates with DNA-seq data to identify broad and narrow regions of enrichment, even within amplified genomic regions', *Genome biology*, Vol. 12, No. 7, 2011.

Reichelt, R., *Untersuchung zur Funktion der TrmB-Familie in Pyrococcus furiosus anhand von Chromatin Immunpraezipitationsexperimenten*, 2009. Diplomarbeit.

Reichlen, M. J., Vepachedu, V. R., Murakami, K. S., and Ferry, J. G., 'MreA functions in the global regulation of methanogenic pathways in *Methanosarcina acetivorans'*, *mBio*, Vol. 3, No. 4, 2012.

Ren, B., Robert, F., Wyrick, J. J., Aparicio, O., Jennings, E. G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T. L., Wilson, C. J., Bell, S. P., and Young, R. A., 'Genome-wide location and function of DNA binding proteins', *Science (New York, N.Y.)*, Vol. 290, No. 5500, 2000.

Rhee, H. S., and Pugh, B. F., 'Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution', *Cell*, Vol. 147, No. 6, 2011.

——, 'Genome-wide structure and organization of eukaryotic pre-initiation complexes', *Nature*, Vol. 483, No. 7389, 2012.

Robb, F. T., Maeder, D. L., Brown, J. R., DiRuggiero, J., Stump, M. D., Yeh, R. K., Weiss, R. B., and Dunn, D. M., 'Genomic sequence of hyperthermophile, *Pyrococcus furiosus*: implications for physiology and enzymology', *Methods in enzymology*, Vol. 330, 2001.

Robinson, J. T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P., 'Integrative genomics viewer', *Nature biotechnology*, Vol. 29, No. 1, 2011.

Rost, M., *Expressionsanalysen des hypothetischen Proteins PF1088 in Pyrococcus furiosus*, 2012. Zulassungsarbeit.

Rozowsky, J., Euskirchen, G., Auerbach, R. K., Zhang, Z. D., Gibson, T., Bjornson, R., Carriero, N., Snyder, M., and Gerstein, M. B., 'PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls', *Nature biotechnology*, Vol. 27, No. 1, 2009.

Santangelo, T. J., Cubonova, L., James, C. L., and Reeve, J. N., 'TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro', *Journal of molecular biology*, Vol. 367, No. 2, 2007.

Sapra, R., Bagramyan, K., and Adams, Michael W W, 'A simple energy-conserving system: proton reduction coupled to proton translocation', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 100, No. 13, 2003.

Sato, T., Fukui, T., Atomi, H., and Imanaka, T., 'Targeted gene disruption by homologous recombination in the hyperthermophilic archaeon *Thermococcus kodakaraensis KOD1'*, *Journal of bacteriology*, Vol. 185, No. 1, 2003.

Sato, T., Imanaka, H., Rashid, N., Fukui, T., Atomi, H., and Imanaka, T., 'Genetic evidence identifying the true gluconeogenic fructose-1,6-bisphosphatase in *Thermococcus kodakaraensis* and other hyperthermophiles', *Journal of bacteriology*, Vol. 186, No. 17, 2004.

Say, R. F., and Fuchs, G., 'Fructose 1,6-bisphosphate aldolase/phosphatase may be an ancestral gluconeogenic enzyme', *Nature*, Vol. 464, No. 7291, 2010.

Schäfer, T., and Schönheit, P., 'Pyruvate metabolism of the hyperthermophilic archaebacterium *Pyrococcus furiosus'*, *Archives of Microbiology*, Vol. 155, No. 4, 1991.

——, 'Maltose fermentation to acetate, CO2 and H2 in the anaerobic hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for the operation of a novel sugar fermentation pathway', *Archives of Microbiology*, Vol. 158, No. 3, 1992.

——, 'Gluconeogenesis from pyruvate in the hyperthermophilic archaeon *Pyrococcus furiosus*: involvement of reactions of the Embden-Meyerhof pathway', *Archives of Microbiology*, Vol. 159, No. 4, 1993.

Schäfer, T., Selig, M., and Schönheit, P., 'Acetyl-CoA synthetase (ADP forming) in archaea, a novel enzyme involved in acetate formation and ATP synthesis', *Archives of Microbiology*, Vol. 159, No. 1, 1993.

Schmid, A. K., Pan, M., Sharma, K., and Baliga, N. S., 'Two transcription factors are necessary for iron homeostasis in a salt-dwelling archaeon', *Nucleic acids research*, Vol. 39, No. 7, 2011.

Schmid, A. K., Reiss, D. J., Pan, M., Koide, T., and Baliga, N. S., 'A single transcription factor regulates evolutionarily diverse but functionally linked metabolic pathways in response to nutrient availability', *Molecular systems biology*, Vol. 5, 2009.

Schmidt, Thomas G M, and Skerra, A., 'The Strep-tag system for one-step purification and high-affinity detection or capturing of proteins', *Nature protocols*, Vol. 2, No. 6, 2007.

Scholz, S., Sonnenbichler, J., Schafer, W., and Hensel, R., 'Di-myo-inositol-1,1'-phosphate: a new inositol phosphate isolated from *Pyrococcus woesei'*, *FEBS letters*, Vol. 306, 2-3, 1992.

Schönheit, P., and Schäfer, T., 'Metabolism of hyperthermophiles', *World Journal of Microbiology and Biotechnology*, Vol. 11, No. 1, 1995.

Schopf, S., Wanner, G., Rachel, R., and Wirth, R., 'An archaeal bi-species biofilm formed by *Pyrococcus furiosus* and *Methanopyrus kandleri'*, *Archives of microbiology*, Vol. 190, No. 3, 2008.

Schramm, M., *ChIP-Analyse des Transkriptionsapparates nach Hitzeschock in Pyrococcus furiosus*, 2013. Bachelor thesis.

Schut, G. J., Boyd, E. S., Peters, J. W., and Adams, Michael W W, 'The modular respiratory complexes involved in hydrogen and sulfur metabolism by heterotrophic hyperthermophilic archaea and their evolutionary implications', *FEMS microbiology reviews*, Vol. 37, No. 2, 2013.

Schut, G. J., Brehm, S. D., Datta, S., and Adams, Michael W W, 'Whole-genome DNA microarray analysis of a hyperthermophile and an archaeon: *Pyrococcus furiosus* grown on carbohydrates or peptides', *Journal of bacteriology*, Vol. 185, No. 13, 2003.

Schut, G. J., Bridger, S. L., and Adams, Michael W W, 'Insights into the metabolism of elemental sulfur by the hyperthermophilic archaeon *Pyrococcus furiosus*: characterization of a coenzyme A- dependent NAD(P)H sulfur oxidoreductase', *Journal of bacteriology*, Vol. 189, No. 12, 2007.

Schut, G. J., Zhou, J., and Adams, M. W., 'DNA microarray analysis of the hyperthermophilic archaeon *Pyrococcus furiosus*: evidence for anNew type of sulfur-reducing enzyme complex', *Journal of bacteriology*, Vol. 183, No. 24, 2001.

Sharma, K., Gillum, N., Boyd, J. L., and Schmid, A., 'The RosR transcription factor is required for gene expression dynamics in response to extreme oxidative stress in a hypersaline-adapted archaeon', *BMC genomics*, Vol. 13, 2012.

Shockley, K. R., Ward, D. E., Chhabra, S. R., Conners, S. B., Montero, C. I., and Kelly, R. M., 'Heat shock response by the hyperthermophilic archaeon *Pyrococcus furiosus'*, *Applied and environmental microbiology*, Vol. 69, No. 4, 2003.

Siebers, B., and Schönheit, P., 'Unusual pathways and enzymes of central carbohydrate metabolism in Archaea', *Current opinion in microbiology*, Vol. 8, No. 6, 2005.

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G., 'Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega', *Molecular systems biology*, Vol. 7, 2011.

Soding, J., Biegert, A., and Lupas, A. N., 'The HHpred interactive server for protein homology detection and structure prediction', *Nucleic acids research*, Vol. 33, Web Server issue, 2005.

Solomon, M. J., Larsen, P. L., and Varshavsky, A., 'Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene', *Cell*, Vol. 53, No. 6, 1988.

Solomon, M. J., and Varshavsky, A., 'Formaldehyde-mediated DNA-protein crosslinking: a probe for *in vivo* chromatin structures', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 82, No. 19, 1985.

Sommer, B., Waege, I., Pollmann, D., Seitz, T., Thomm, M., Sterner, R., and Hausner, W., 'Activation of a chimeric Rpb5/RpoH subunit using library selection', *PloS one*, Vol. 9, No. 1, 2014.

Song, N., Nguyen Duc, T., van Oeffelen, L., Muyldermans, S., Peeters, E., and Charlier, D., 'Expanded target and cofactor repertoire for the transcriptional activator LysM from *Sulfolobus'*, *Nucleic acids research*, Vol. 41, No. 5, 2013.

Soppa, J., 'Normalized nucleotide frequencies allow the definition of archaeal promoter elements for different archaeal groups and reveal base-specific TFB contacts upstream of the TATA box', *Molecular microbiology*, Vol. 31, No. 5, 1999a.

——, 'Transcription initiation in Archaea: facts, factors and future aspects', *Molecular microbiology*, Vol. 31, No. 5, 1999b.

Soppa, J., 'Ploidy and gene conversion in Archaea', *Biochemical Society transactions*, Vol. 39, No. 1, 2011.

Surma, M., *Charakterisierung von TrmB und TrmBL1, zwei Transkriptionsregulatoren im Zentrum des Zuckerstoffwechsels von Pyrococcus furiosus*, 2011. http://epub.uni-regensburg.de/14529/.

Techtmann, S. M., and Robb, F. T., 'Archaeal-like chaperonins in bacteria', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 107, No. 47, 2010.

Thompson, D. K., and Daniels, C. J., 'Heat shock inducibility of an archaeal TATA-like promoter is controlled by adjacent sequence elements', *Molecular microbiology*, Vol. 27, No. 3, 1998.

Thorgersen, M. P., Stirrett, K., Scott, R. A., and Adams, Michael W W, 'Mechanism of oxygen detoxification by the surprisingly oxygen-tolerant hyperthermophilic archaeon, *Pyrococcus furiosus'*, *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 109, No. 45, 2012.

Todor, H., Sharma, K., Pittman, Adrianne M C, and Schmid, A. K., 'Protein-DNA binding dynamics predict transcriptional response to nutrients in archaea', *Nucleic acids research*, Vol. 41, No. 18, 2013.

Towbin, H., Staehelin, T., and Gordon, J., 'Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 76, No. 9, 1979.

Tuininga, J. E., Verhees, C. H., van Der Oost, J, Kengen, S. W., Stams, A. J., and de Vos, W M, 'Molecular and biochemical characterization of the ADP-dependent phosphofructokinase from the hyperthermophilic archaeon Pyrococcus furiosus', *The Journal of biological chemistry*, Vol. 274, No. 30, 1999.

Turner, J. S., Glands, P. D., Samson, A. C., and Robinson, N. J., 'Zn2+-sensing by the cyanobacterial metallothionein repressor SmtB: different motifs mediate metal-induced protein-DNA dissociation', *Nucleic acids research*, Vol. 24, No. 19, 1996.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., and Rozen, S. G., 'Primer3--new capabilities and interfaces', *Nucleic acids research*, Vol. 40, No. 15, 2012.

VanDerAuwera, Geraldine A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., and DePristo, M. A., 'From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline', in , *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc, 2002.

VanDerOost, J., Schut, G., Kengen, S. W., Hagen, W. R., Thomm, M., and de Vos, W M, 'The ferredoxin-dependent conversion of glyceraldehyde-3-phosphate in the

hyperthermophilic archaeon *Pyrococcus furiosus* represents a novel site of glycolytic regulation', *The Journal of biological chemistry*, Vol. 273, No. 43, 1998.

VanDeWerken, H., Verhees, C. H., Akerboom, J., De Vos, W. M., and Van Der Oost, J., 'Identification of a glycolytic regulon in the archaea *Pyrococcus* and *Thermococcus'*, *FEMS microbiology letters*, Vol. 260, No. 1, 2006.

Vanfossen, A. L., Lewis, D. L., Nichols, J. D., and Kelly, R. M., 'Polysaccharide degradation and synthesis by extremely thermophilic anaerobes', *Annals of the New York Academy of Sciences*, Vol. 1125, 2008.

Vannini, A., and Cramer, P., 'Conservation between the RNA polymerase I, II, and III transcription initiation machineries', *Molecular cell*, Vol. 45, No. 4, 2012.

VanZile, M. L., Cosper, N. J., Scott, R. A., and Giedroc, D. P., 'The zinc metalloregulatory protein Synechococcus PCC7942 SmtB binds a single zinc ion per monomer with high affinity in a tetrahedral coordination geometry', *Biochemistry*, Vol. 39, No. 38, 2000.

Verhees, C. H., Akerboom, J., Schiltz, E., De Vos, Willem M, and Van Der Oost, John, 'Molecular and biochemical characterization of a distinct type of fructose-1,6-bisphosphatase from *Pyrococcus furiosus'*, *Journal of bacteriology*, Vol. 184, No. 12, 2002.

Verhees, C. H., Kengen, Serve W M, Tuininga, J. E., Schut, G. J., Adams, Michael W W, De Vos, Willem M, and Van Der Oost, John, 'The unique features of glycolytic pathways in Archaea', *The Biochemical journal*, Vol. 375, Pt 2, 2003.

Vierke, G., *Die Hitzeschockantwort oberhalb des Siedepunktes von Wasser: Funktion und Struktur des neuartigen Transkriptionsregulators Phr aus* Pyrococcus furiosus, 2007. http://epub.uni-regensburg.de/10504/.

Vierke, G., Engelmann, A., Hebbeln, C., and Thomm, M., 'A novel archaeal transcriptional regulator of heat shock response', *The Journal of biological chemistry*, Vol. 278, No. 1, 2003.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M., and Pennacchio, L. A., 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature*, Vol. 457, No. 7231, 2009.

Waege, I., *Charakterisierung von RNA-Polymerase-Mutanten und von Transkriptionsfaktoren mit Hilfe des genetischen Systems von Pyrococcus furiosus*, 2014. PhD thesis unpublished.

Waege, I., Schmid, G., Thumann, S., Thomm, M., and Hausner, W., 'Shuttle vector-based transformation system for *Pyrococcus furiosus'*, *Applied and environmental microbiology*, Vol. 76, No. 10, 2010.

Wagner, M., Wagner, A., Ma, X., Kort, J. C., Ghosh, A., Rauch, B., Siebers, B., and Albers, S.-V., 'Investigation of the malE promoter and MalR, a positive regulator of the maltose regulon, for an improved expression system in *Sulfolobus acidocaldarius'*, *Applied and environmental microbiology*, Vol. 80, No. 3, 2014.

Wang, H., Liu, B., Wang, Q., and Wang, L., 'Genome-Wide Analysis of the Salmonella Fis Regulon and Its Regulatory Mechanism on Pathogenicity Islands', *PloS one*, Vol. 8, No. 5, 2013.

Waterhouse, A. M., Procter, J. B., Martin, David M A, Clamp, M., and Barton, G. J., 'Jalview Version 2--a multiple sequence alignment editor and analysis workbench', *Bioinformatics (Oxford, England)*, Vol. 25, No. 9, 2009.

Werner, F., 'Molecular mechanisms of transcription elongation in archaea', *Chemical reviews*, Vol. 113, No. 11, 2013.

Werner, F., and Grohmann, D., 'Evolution of multisubunit RNA polymerases in the three domains of life', *Nature reviews. Microbiology*, Vol. 9, No. 2, 2011.

White, J. R., Escobar-Paramo, P., Mongodin, E. F., Nelson, K. E., and DiRuggiero, J., 'Extensive genome rearrangements and multiple horizontal gene transfers in a population of pyrococcus isolates from Vulcano Island, Italy', *Applied and environmental microbiology*, Vol. 74, No. 20, 2008.

Wierzbicki, A. T., Haag, J. R., and Pikaard, C. S., 'Noncoding transcription by RNA polymerase Pol IVb/Pol V mediates transcriptional silencing of overlapping and adjacent genes', *Cell*, Vol. 135, No. 4, 2008.

Wilbanks, E. G., Larsen, D. J., Neches, R. Y., Yao, A. I., Wu, C.-Y., Kjolby, Rachel A S, and Facciotti, M. T., 'A workflow for genome-wide mapping of archaeal transcription factors with ChIP-seq', *Nucleic acids research*, Vol. 40, No. 10, 2012.

Winter, J., *Untersuchung zur Funktion von TFB-RF1*, 2012. Bachelor thesis.

Woese, C. R., Kandler, O., and Wheelis, M. L., 'Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya', *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 87, No. 12, 1990.

Wojtas, M. N., Mogni, M., Millet, O., Bell, S. D., and Abrescia, Nicola G A, 'Structural and functional analyses of the interaction of archaeal RNA polymerase with DNA', *Nucleic acids research*, Vol. 40, No. 19, 2012.

Yoon, S. H., Reiss, D. J., Bare, J. C., Tenenbaum, D., Pan, M., Slagel, J., Moritz, R. L., Lim, S., Hackett, M., Menon, A. L., Adams, Michael W W, Barnebey, A., Yannone, S. M., Leigh, J. A., and Baliga, N. S., 'Parallel evolution of transcriptome architecture during genome reorganization', *Genome research*, Vol. 21, No. 11, 2011.

Zhang, J., Li, E., and Olsen, G. J., 'Protein-coding gene promoters in *Methanocaldococcus* (*Methanococcus*) *jannaschii'*, *Nucleic acids research*, Vol. 37, No. 11, 2009.

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nusbaum, C., Myers, R. M., Brown, M., Li, W., and Liu, X. S., 'Model-based analysis of ChIP-Seq (MACS)', *Genome biology*, Vol. 9, No. 9, 2008.

Zhu, Y., Kumar, S., Menon, A. L., Scott, R. A., and Adams, Michael W W, 'Regulation of iron metabolism by *Pyrococcus furiosus'*, *Journal of bacteriology*, Vol. 195, No. 10, 2013.

Zillig, W., Stetter, K. O., and Janekovic, D., 'DNA-dependent RNA polymerase from the archaebacterium *Sulfolobus acidocaldarius'*, *European journal of biochemistry / FEBS*, Vol. 96, No. 3, 1979.

Erklärung

Ich versichere hiermit, daß ich die Arbeit selbständig angefertigt habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt sowie die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht habe. Diese Arbeit war bisher noch nicht Bestandteil eines Prüfungsverfahren, andere Promotionsversuche wurden nicht unternommen.

Regensburg, den 26.05.2014

.....

Robert Reichelt

Danksagung

Zuallererst möchte ich mich sehr herzlich bei Herrn Prof. Dr. Michael Thomm für die Vergabe dieses spannenden and vielfältigen Themas und die Betreuung während meiner Doktorarbeit bedanken. Insbesonders danke ich ihm für seine Geduld und die Zeit, die er mir gegeben hat, um diese komplexe Fragestellung zu einem Abschluss zu bringen.

Desweiteren bedanke ich mich sehr herzlich bei PD Dr. Winfried Hausner, der sich als mein 1. Mentor immer Zeit für anregende wissenschaftliche Diskussionen genommen hat. Seine experimentelle, organisatorische Erfahrung und Hilfe hat mit zum Gelingen dieser Arbeit beigetragen.

Thanks a lot to my second mentor Assoc. Prof. Dr. Michael Bartlett for the wonderful stay in Portland (OR, USA). I really enjoyed working in your laboratory. In addition, our scientific discussions via Skype helped me a lot.

Zudem bedanke ich mich bei Herrn Prof. Dr. Herbert Tschochner dafür, dass er sich dazu bereit erklärt hat diese Arbeit als Zweitkorrektor zu bewerten. Auch bei meinem 3. Prüfer Herrn Prof. Dr. Reinhard Sterner möchte ich mich bedanken.

Vor allem danke ich meiner Freundin Simone Wurm dafür, dass sie die letzten 10 Jahre immer für mich da war und wir gemeinsam sowohl die Strapazen als auch schönen Dinge einer Doktorarbeit durchstehen konnten.

Ein herzlicher Dank geht an meine Freunde und Laborkollegen mit denen ich die letzten Jahre zusammenarbeiten durfte: Thomas Fouqueau und Mirijam Zeller, Ingrid Waege, Alexander Probst, Stefan Dexl, Antonia Gindner, und Lydia Kreuter.

Zudem möchte ich mich bei allen früheren und aktuellen Mitarbeitern des Lst. für Mikrobiologie bedanken. Insbesondere bei Frau Elisabeth Nagelfeld für die liebenswerte Hilfe bei jeglicher Art von administrativen Problemen, bei Herrn Dr. Sebastian Grünberg für Einführung in die archaeelle Transkription und dazugehörigen Techniken und bei Wolfgang, Renate, Gabi, Thomas und Konni für die experimentelle Unterstützung.

Zudem bedanke ich mich bei allen weiteren Doktoranden, Studenten, usw. des Lehrstuhls für Mikrobiologie und Archaeenzentrum.

Und zuletzt, ein besonderer Dank gilt auch meiner Familie, die mich immer liebevoll unterstützt hat.