Structural and functional analysis of **e**ukaryal-like proteins from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Hao Wu

Promotoren	**Prof. dr. John van der Oost** persoonlijk hoogleraar bij de leerstoelgroep Microbiologie Wageningen Universiteit
	Prof. dr. Willem M. de Vos Hoogleraar in de Microbiologie Wageningen Universiteit
Co-promotor	**Prof. dr. Zihe Rao** Hoogleraar in de Structuur Biologie Tsinghua Universiteit Beijing, China
Leden van de Promotie Commissie	**Prof. dr. Sacco C. de Vries** Wageningen Universiteit
	Dr. Henk A. Schols Wageningen Universiteit
	Dr. Joen Luirink Vrije Universiteit, Amsterdam
	Prof. dr. Bauke W. Dijkstra Rijksuniversiteit Groningen

Dit onderzoek is uitgevoerd binnen de onderzoekschool VLAG

Structural and functional analysis of **e**ukaryal-like proteins from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Hao Wu

Proefschrift
ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit, Prof. dr. M.J. Kropff,
in het openbaar te verdedigen op maandag 10 december 2007
des ochtends te elf uur in de Aula

Hao Wu - Structural and functional analysis of eukaryal-like proteins from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Ph.D. Thesis Wageningen University, Wageningen, Netherlands (2007)
158p. - with summary in Dutch

ISBN 978-90-8504-822-0

Table of contents

Preface

Aim and outline of this thesis

Life on this planet includes organisms belonging to three domains, eukarya, bacteria and archaea. The focus of this thesis is on the latter group, which includes prokaryotic microorganisms that thrive in extreme ecosystems. *Sulfolobus solfataricus* is a thermo-acidophilic archaeon that has become a model organism for several practical reasons, including its easy cultivation (relatively fast growth, aerobically on a range of carbon sources), and the availability of several extra-chromosomal elements (plasmids, viruses). This has led to the fact that *S. solfataricus* was among the earliest archaea that were selected for complete genome sequence analysis, and that genetic tools have been developed that allow gene disruption and overproduction.

The research presented in this thesis is aimed at applying technologies in bioinformatics, biochemistry, structural biology and cell biology to reveal the global regulation network in archaea, gain insights in the mechanism of archaeal signal transduction, and provide details on the evolution of the well-conserved archaeal-eukaryal information processing systems, including transcription, translation, and replication. The global regulation network includes several novel core eukaryal-like proteins that are predicted to operate in the regulation of transcription and/or translation in archaea. In addition, approaches are described to analyze the function of the predicted regulators that involved the development of antibiotic resistance marker for hyperthermophiles since the genetic modification of hyperthermophiles has been hampered, at least in part, by the lack of suitable selection markers.

Chapter 1 – General introduction

The first chapter gives a general overview of archaea and their regulation mechanisms with specific attention for *Sulfolobus solfataricus.* An important enzyme, GTPase and its molecular switch mechanism are introduced as it is an essential regulator in all domains of life.

Chapter 2 – Engineering a selectable marker for hyperthermophiles

The lack of suitable antibiotic selection markers is a major limitation for the development of genetic systems for hyperthermophiles. The goal of this chapter was to engineer an antibiotic resistance marker for hyperthermophiles by applying a directed evolution approach. Stabilized mesophilic resistance markers were successfully selected in the thermophilic bacterium *Thermus thermophilus* .

Chapter 3 – Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*

This chapter describes the characterization of a new enzymes involved in the degradation of α-linked galactose oligosccharides. In plants, α-galactosides such as raffinose and stachyose, serve to store energy in leaves, roots, and tubers. Biochemical and molecular genetics methods were used to gain insight in a novel archaeal α-galactosidase.

Chapter 4 – Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*

The gene for a novel GTP-binding protein from *Sulfolobus solfataricus* has been cloned and overexpressed in *E. coli*. The purified protein was crystallized using the hanging-drop vapor-diffusion technique in the presence of 0.05 *M* cadmium sulfate and 0.8 *M* sodium acetate pH 7.5. A single-wavelength anomalous dispersion data set was collected to a maximum resolution of 2.0 Å using a single cadmium-incorporated crystal and an initial analysis is described.

Chapter 5 – Structural and functional analysis of ubiquitous HflX-like GTPase

The crystal structures are described of SsGBP alone and with cofactor GDP. SsGBP is a novel HflX-like GTPase in the GTPase superfamily. In human, the highly conserved SsGBP-homologs have proposed to play a role in the onset of prostate cancer. In *E. coli*, it is reported to be involved in the lysis/lysogeny decision of phage lambda. So far, no archaeal homologs have been characterized. Structural and biochemical analyses demonstrate that this novel GTPase might be involved in tRNA processing and as such may contribute to translation fidelity.

Chapter 6 – Molecular characterization of the eukaryotic-like Multi-protein bridging factor (MBF) from thermophilic archaeon *Sulfolobus solfataricus*

Multi-protein bridging factor (MBF) is highly conserved in archaea and eukarya, but absent in bacteria. In eukarya, MBF shows a conserved genomic context with several genes including a proteasome-activating nucleotidase, a GTP-binding protein, a basic transcription factor (TFE), and proteins involved in RNA binding/modification. Evidence is provided that at least under certain conditions MBF from *Sulfolobus solfataricus* can interact with DNA or RNA.

Chapter 7 – Summary and conclusions

This chapter is a brief summary of the results in previous chapters and hypothesis based on the results.

Chapter 8 – Nederlandse samenvatting en conclusies

The last chapter is a translation of chapter 7 into dutch. Het laatste hoofdstuk bevat de nederlandse vertaling van hoofdstuk 7 en is een samenvatting van de resultaten uit de vorige hoofdstukken en hypotheses gebaseerd op die resultaten.

Chapter 1

General Introduction

ARCHAEA, THE THIRD DOMAIN OF LIFE

The archaea constitute one of the three domains of life, next to the bacteria and the eukarya (Woese and Fox, 1977; Woese et al., 1990). In 1859, Charles Darwin originally predicted that the evolutionary drift of genetic information (DNA sequences) is the basis for the diversity of life forms. Classification of life has initially been performed on the basis of morphology, later in combination with biochemistry. Until four decades ago, life was divided into two domains: the prokarya and the eukarya (Stanier and Van Niel, 1962). At first glance, the morphology of organisms that belong to the prokarya domain is very similar, apparently lacking the complex composition of eukaryotic cells. Because of their morphological similarity, the co-existence of two fundamentally different types of prokaryotes was certainly not recognized before the introduction of molecular classification techniques in the 1970s. Comparison of ubiquitous sequences such as the building blocks of the protein synthesis machinery (ribosomal RNA and protein) were used to compose universal phylogenetic trees. In 1977, Carl Woese and George Fox performed a phylogenetic analysis by comparing rRNA sequences and a surprising discovery was obtained that early in the cellular evolution two domains diverged within the prokaryotes: the archaebacteria and eubacteria (Woese and Fox, 1977). This is also so-called “the Woesian revolution” (Doolittle and Brown, 1994). It was concluded that life should be divided into three distinct types of living systems: Bacteria, Archaea and Eukarya (Fig. 1). The comparison of properties among Archaea, Bacteria and Eukarya is listed in Table 1: (i) the RNA polymerase from Archaea is similar to RNA polymerase II from eukaryotes; (ii) the ribosomes of Archaea are similar in size to those of Bacteria (70S), although the sequences of the rRNA molecules and the ribosomal proteins of archaea are more related to the eukaryotic counterparts (Fig 1 is based on 16S/18S rRNA comparison); (iii) lipids in membranes from Archaea are unique, containing ether linkages between the glycerol backbone and the fatty acids, instead of ester linkages; (iv) the cells walls of Archaea are chemically and structurally diverse and do not contain peptidoglycan.

Although initially the majority of the archaea were isolated from extreme environments (high temperature, high salt concentration, extreme pH), it has become clear that archaea also thrive in non-extreme environments; moreover, it has recently been found that archaea have an abundant and global distribution, and play an important role in the bio-geochemical cycles in many ecosystems ranging from soil to oceans (Bintrim et al., 1997; DeLong and Pace, 2001; Pace, 1997; Stetter, 1999). The archaea domain was

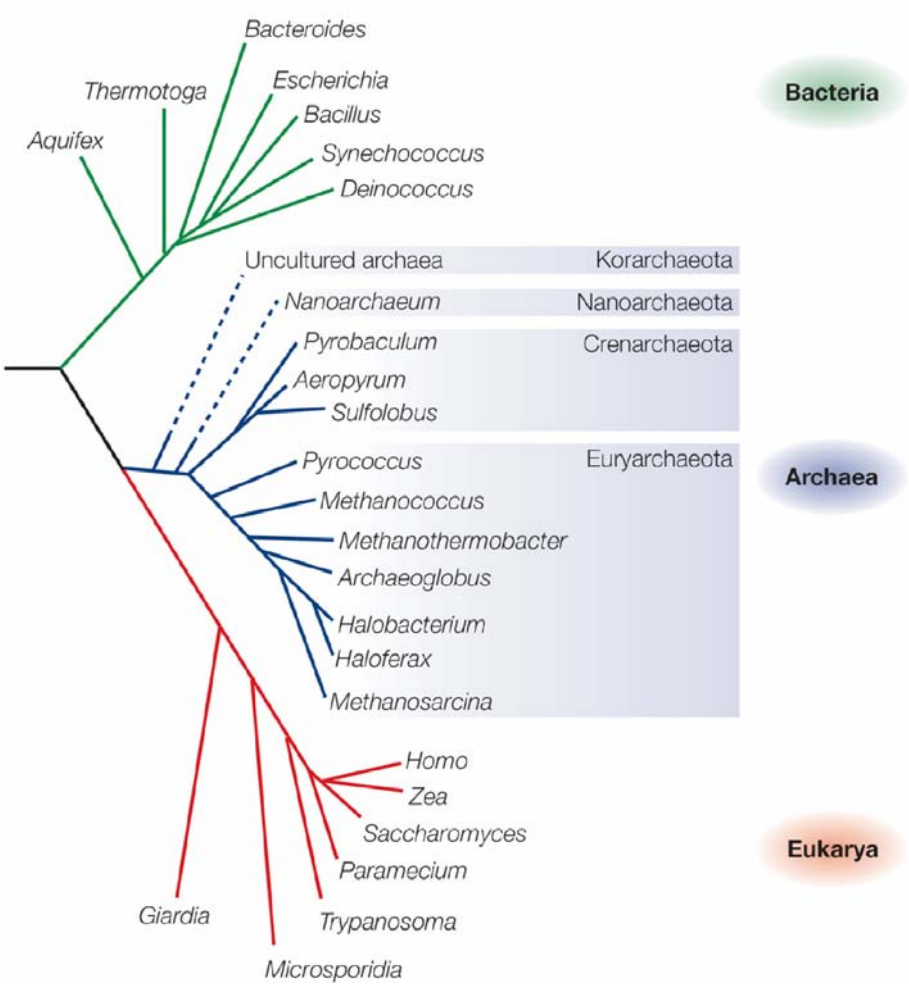


Figure 1. The tree of life. The universal phylogenetic tree of life based on analysis of 16S/18S rRNA and rooted with paralogous protein sequences (Allers and Mevarech, 2005).

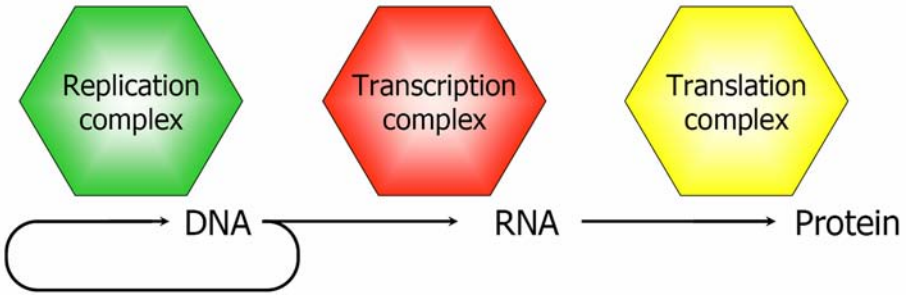


Figure 2. The Central Dogma. The flow of genetic information to protein unctionality proceeds at different stages: replication of DNA by the Replication (DNA Polymerase) complex, transcription of DNA to RNA by the Transcription (RNA Polymerase) complex, and translation of RNA to protein by the Ribosome (Amino Acid Polymerase) complex.

divided into two main phylogenetic groups, the Crenarchaeota and the Euryarchaeota (Woese et al., 1990). With the ongoing genomic sequencing projects, the tree of life has recently been refined on the basis of concatenated ribosomal proteins (Ciccarelli et al., 2006); this has further strengthened the close relation between archaea and eukaryotes, at least of the systems involved in the processing of the genetic information (DNA > RNA > protein). Also, the other two branches of archaea (Korarchaeota and Nanoarchaeota) were annotated recently (Allers and Mevarech, 2005; Makarova and Koonin, 2005). Korarchaea have been isolated from geothermal environments (Barns et al., 1996), as well as from hydrothermal regions (Auchtung et al., 2006; Marteinsson et al., 2001). One species of Nanoarchaetoa, *Nanoarchaeum equitians* has been cultivated by Huber (Huber et al., 2002). It is an extremely small organism, only 400 nm in diameter, that appears to have a parasitic or symbiotic relation with its archaeal host. Other sequences of Nanoarchaeota have recently been identified at many locations (Hohn et al., 2002).

The genomics era made clear that most of the relevant cellular mechanisms in eukaryotes (transcription, translation, and replication) directly evolved from the basic systems that originally developed in archaea, and that these mechanisms in many respects differ substantially from their bacterial counterparts (Bell and Jackson, 1998a; Bell and Jackson, 1998b; Olsen and Woese, 1997; Rivera and Lake, 2004). Recent studies were performed by evaluating completed genomes. It was found that several hypothetical proteins (the core, MBF) are conserved in all archaea, with orthologs in eukarya but not in

Table 1. Comparison of properties between Archaea, Bacteria and Eukarya. Adapted from The Microbial World (http://www.microbiologytext.com/).

Property	Bacteria	Archaea	Eukarya
RNA polymerase	4 proteins Rifampicin sensitive	8-10 proteins Rifampicin resistant	12 proteins (RNA pol II) Rifampicin resistant
Transcription start site	Variable often contains a -35 and -10 region	TATA box	TATA box
Starting amino acid	formylmethionine	methionine	methionine
Ribosome	70S	70S	80S
Lipids	ester-linked	ether-linked	ester-linked
Cell wall composition	G+ peptidoglycan G - peptidoglycan and outer membrane	pseudopeptidoglycan or S-layer of proteins, glycoproteins, or polysaccharides	none or cellulose

bacteria (Koonin et al., 2001). This set of uncharacterized proteins might correspond to unknown structural or regulatory components of the aforementioned cellular mechanisms.

REGULATION MECHANISMS IN ARCHAEA

Proteins are the primary functional molecules of all living cells. The complete set of proteins that is present in a living cell (proteome) corresponds to the genes that are expressed, and are the functional representation of its genome. The flow of genetic information (Fig. 2), i.e. the conversion of the gene's nucleotide sequence to the protein's amino acid sequence, proceeds via different types of RNA: a nucleotide intermediate (mRNA), a nucleotide-amino acid translation system (tRNA), and an amino acid polymerase (rRNA, as part of the ribosome). This universal system for the flow of genetic information is referred to as the Central Dogma of Molecular Biology (Crick, 1970; Crick, 1958; Watson and Crick, 1953a; Watson and Crick, 1953b).

In archaea, regulation at translational level (attenuation, anti-termination) has not been reported yet. A relatively important site of regulation of archaeal metabolism appears to be control at transcriptional level. The archaeal transcription machinery differs significantly from the bacterial four-subunit RNA Polymerase and its Sigma factors. Instead, the archaeal system resembles the core of the eukaryal Pol II system: 10-12 RNA Polymerase subunits, TATA box-Binding Protein (TBP) and Transcription Factor IIB (TFB) (Bell and Jackson, 2001; Thomm, 1996). These general transcription factors constitute the transcription initiation complex (Fig. 3) that interacts with a promoter region upstream a gene, and generally gives rise to a basal level of transcription. In eukarya, transcription is enhanced by transcriptional activator proteins that bind to DNA (enhancer elements) as well as to mediator proteins (co-activators). These co-activators selectively enable the stimulation by subsets of transcriptional activators. Among the best-characterized coactivators are the TAFs (TBP-associated factors, subunits of TFIID). TAFs have been demonstrated to provide the protein interfaces to link the sequence-specific factors to the basal transcriptional machinery, and some exhibit specific enzymatic functions essential for activated gene expression (phosphorylation, acetylation). Moreover, co-activators distinct from those associated with TBP have also been identified (Guarente, 1995; Moreau et al., 1998). An example of the latter class is the Multiprotein Bridging Factor (MBF) that has been characterized in yeast, fly, and man (Jindra et al., 2004; Kabe et al., 1999; Liu et al., 2007; Takemaru et al., 1998; Takemaru et al., 1997).

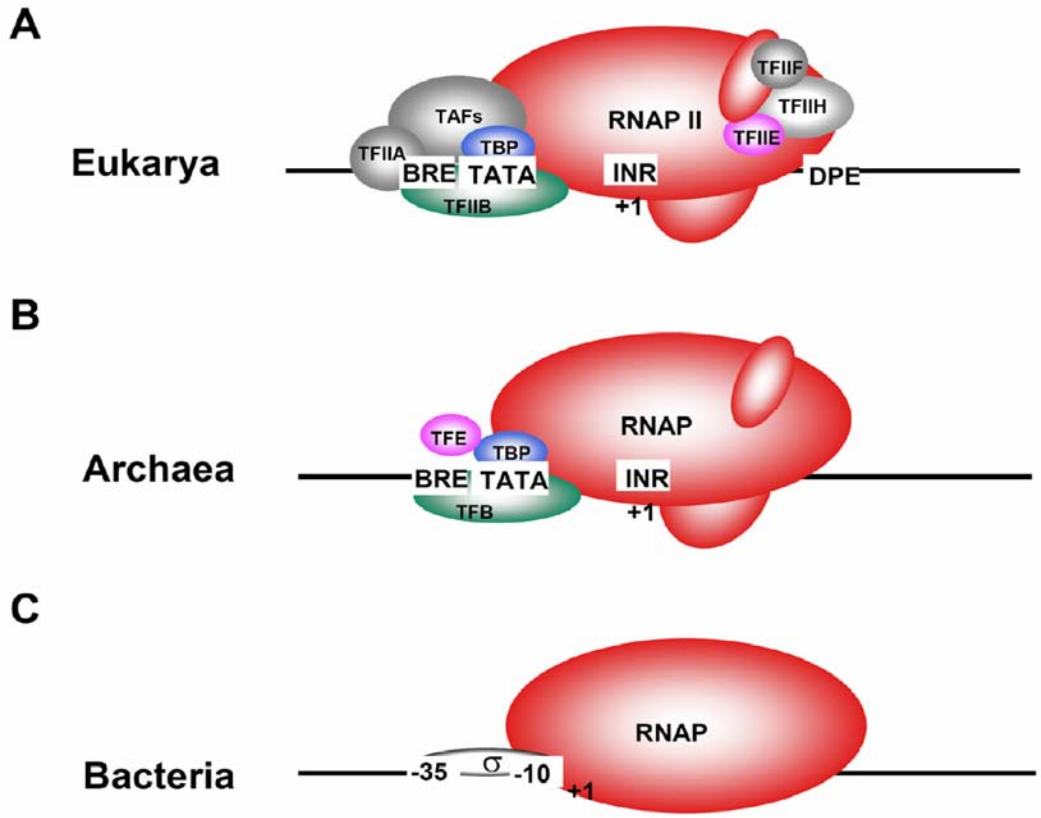


Figure 3. Transcription initiation machinery. Schematic representation of the elements of the eukaryotic (A), archaeal (B), and bacterial (C) basal transcription machinery.The first step in transcription initiation is the assembly of the preinitiation complex at the promoter. The archaeal transcription machinery does not resemble that of bacteria, but is rather eukaryal-like.

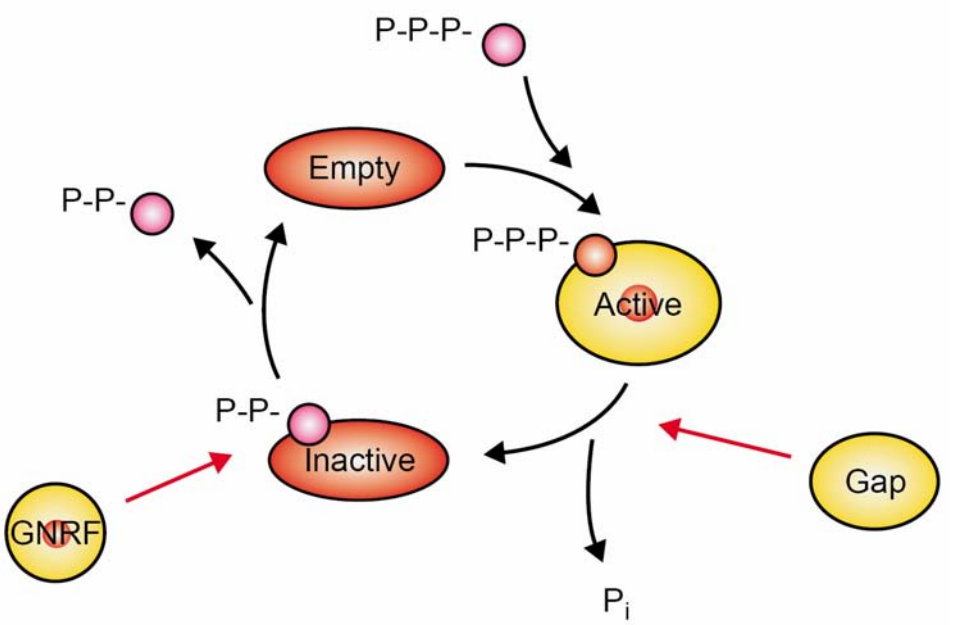


Figure 4. GTPase Cycles. The GTPase Cycles. Three conformational states in GTPase cycle: in nucleotide-free form, GTPase is inactive; GTP-bound form, GTPase becomes active; when GTP is hydrolyzed by GTPase, it enters inactive. During the cycle, GTPase activating proteins (GAPs) can assist the processing of hydrolysis GTP into GDP, and an effect being generated on the target proteins. In the meanwhile, Guanine nucleotide release factors (GNRFs) participate GDP dissociates from the inactive GTPase (Bourne et al., 1990; Caldon and March, 2003).

Different types of transcriptional regulators have been identified in archaea. (i) Unique archaeal regulators, such as the transcriptional activator of gas vesicle formation in Haloarchaea (GvpE), a proposed leucine zipper that has no obvious homolog in bacteria and eukarya (Zimmermann and Pfeifer, 2003). (ii) Bacterial-like regulators that often block transcription initiation by competing with the transcription factors (TBP, TFB) or the RNA Polymerase (RNAP) for their respective binding site (e.g. Lrp-like regulators) (Brinkman et al., 2003). Recently bacterial-like activators have also been described, that stimulate the formation of the TBP-TFB-RNAP-promoter complex by affecting the DNA structure, by a direct interaction (Ouhammouch et al., 2003), or via an additional bridging protein (Brinkman et al., 2002). (iii) Eukaryal-like regulators are not abundant, but interestingly one is present in all available archaeal genomes, even in the 0.5 Mbp minimal genome of *Nanoarchaeum equitans* (Waters et al., 2003): the aforementioned multi-protein bridging factor (MBF). Whereas the function of eukaryal MBF is to connect transcriptional regulators (generally containing bZIP-type leucine zipper domains, such as Jun and Fos; (Jindra et al., 2004) with the pre-initiation complex (TBP-TFB-RNAP), the role of the archaeal MBF has not yet been established.

MBF is highly conserved both in archaea and eukarya, but not in bacteria. An *in silico* analysis was performed of the conserved genomic neighborhood of archaeal MBF genes. In a total of 11 crenarchaeal genomes, the *mbf* is always clustered with *gbp,* encoding a HflX-like GTP-binding protein (except for the Sponge symbiont *Crenarchaeum symbiosum*). It implies that these two genes might functionally interact with each other during the transcription and translation regulation in archaea.

GTPase

Members of the GTPase superfamily share conserved structural and molecular mechanisms for their functions as molecular switches in diverse biological processes (Bourne et al., 1991). It suggests that all GTPases derive from a single primordial protein. During the course of evolution, these GTPases were repeatedly modified to perform various functions (Bourne et al., 1990). They have been harnessed to direct ribosomal protein synthesis (Hwang and Inouye, 2006; Karbstein, 2007), to mediate transmembrane signalling (Conner et al., 2005; Reiter and Lefkowitz, 2006), to translocate nascent proteins into endoplasmic reticulum (Lee et al., 2006), to control differentiation and cell proliferation (McCormick, 1989), and to guide intracellular vesicular traffic (Salminen and Novick, 1987).

GTPases can hydrolyzes the γ phosphate of GTP into GDP, and this hydrolysis is Mg^{2+}-dependence. By hydrolysis of the GTP through intrinsic GTPase-activity, GTPases can perform throwing the switch from the active GTP-bound form to the inactive GDP-bound form and effectively switch the GTPase off (Fig. 4). In performing different physiological functions, GTPases can interact with a variety of cellular components by switching from an inactive GDP-bound to an active GTP-bound form (Leipe et al., 2003; Vetter and Wittinghofer, 2001). There are two crucial protein and factor are involved in GTPase cycles (Geyer and Wittinghofer, 1997) (Fig. 4). They are Guanine nucleotide exchange factors (GEFs) (Thomas et al., 2007) (Berken et al., 2005) and GTPase-activating proteins (GAPs) (Scheffzek et al., 1997). Both GEFs and GAPs are multidomain proteins (Bos et al., 2007). Usually, GEFs can turn on signaling by catalyzing the exchange from G-protein-bound GDP to GTP; whereas GAPs terminate signaling by inducing GTP hydrolysis and regulate the activity of small GTPases to control cellular functions (Bos et al., 2007; Wittinghofer et al., 1997). GTPase can transduce a signal to a reaction chain via this GTPase cycles (Fig. 4). It suggests that these GAPs and GEFs may be potential therapeutic targets for developing drugs to treat various diseases, including cancer (Bos et al., 2007). New functions of the highly adaptable GTPase molecular switch have being continually uncovered. Comparison of the features of the conserved G-proteins from different subfamilies and (super)families are listed in Table 2.

THERMOSTABILITY

In general, living organisms can be grouped into four main categories as defined by the temperature range where they grow in: psychrophiles, mesophiles, thermophiles and hyperthermophiles (Unsworth et al., 2007). To date, no hyperthermophilic eukaryote has yet been discovered. However, hyperthermophilic microbes are found in the most basal positions in the universal tree of life in both bacteria and Archaea (Luke et al., 2007). A great phylogenetic diversity of hyperthermophilic Archaea has been found, for example *Pyrococcus, Thermococcus, Methanothermus and Sulfolobus*. They thrive in a variety of hyperthermophilic environments, such as a Yellowstone National Park hot spring (Barns et al., 1996); a shallow marine hydrothermal vent (Takai and Horikoshi, 1999b), and various deep-sea hydrothermal vents (Takai and Horikoshi, 1999a). So far, only two genera of hyperthermophilic organisms (*Aquifex and Thermotoga*) were found in bacteria (Bocchetta et al., 2000; Luke et al., 2007). These hyperthermophilic organisms may thus bear

similarities to ancient life forms (DeLong, 2001). Evidences show that proteins from thermophilic (optimal growth temperature 50-80 ^{0}C) and hyperthermophilic (optimal growth temperature ≥ 80 ^{0}C) organisms exhibit remarkable thermal stability and resisitance to chemical denaturants (Luke et al., 2007). It appears no unique feature responsible for the remarkable heat stability properties of hyperthermostable proteins (Unsworth et al., 2007). The identification of the determinants of protein thermal stabilization is often pursued by comparing enzymes from hyperthermophiles with their mesophilic counterparts (Danson and Hough, 1998). By using a variety of strategies combined of virtually all known structural parameters, thermostability of enzymes appears to be implemented in laboratory. Basically, there are nine strategies applied (Luke et al., 2007). (i) Increasing number of ionic interactions. (ii) Increasing extent of hydrophobic-surface burial. (iii) Increasing number of prolines. (iv) Decreasing number of glutamines. (v) Improving core packing. (vi). Creating greater rigidity. (vii) Extending secondary structure. (viii) Shortening of surface loops. (ix) Making higher states of oligomerization.

The classical genetic methods to select and improve the thermostability of an enzyme is relatively slow (Valetti and Gilardi, 2004). During the last decade many methods have been developed for more efficient protein engineering, i.e. by directed-evolution (Arnold and Volkov, 1999; Giver et al., 1998; Kuchner and Arnold, 1997; Moore and Arnold, 1996) and computationally design (Dwyer et al., 2004; Looger et al., 2003). Directed evolution has emerged in a decade, and became a full-grown tool in molecular biology nowadays (Otten and Quax, 2005). It is widely applied in protein engineering, for example, for improving the activity and stability of the interesting enzymes (Giver et al., 1998; Valetti and Gilardi, 2004), as well as changing the substrate specificity of an enzyme (Zhang et al., 1997). DNA shuffling (Stemmer, 1994) is a powerful process for directed evolution (Crameri et al., 1998). It generates diversity by recombination, combining useful mutations from individual genes. In 1994, Pim Stemmer first introduced DNA shuffling in protein engineering (Stemmer, 1994). Since then, this revolution in customizing protein properties has been largely applied (Crameri et al., 1998; Kaper et al., 2002; Zhang et al., 1997). During the past years, computational design offers enormous generality for engineering protein structure and function (Dwyer et al., 2004). Many enzymes can be designed by the computational design. It remarkably improve the substrate selectivity and specificity, and the de novo design of enzyme activities within scaffolds of known structure (Chica et al., 2005; Looger et al., 2003).

Table 2 Comparison of the features of the conserved G domain of SsGBP with G-proteins from different subfamilies and (super)families

Name	E-value	PDB ID	R.M.S.D.	Z-score	GTP/Mg2+ binding site	Switch I region	Switch II region	G1 box	G2 box	G3 box	G4 box	G5 box	References
SsGBP (HlfX-like)	**-**	**2QTF 2QTH**	**-**	**-**	**H Y-NSKTS NK-DK SAL**	**GLTQKV DTKLFT TMS**	**DTVGFIRG IPPQIVDA FFVTLSEA K**	IV**GYT NSGK T**SL	LFT**T** MS	**D**TV**G**	**NKID**	**SAL**	This research
Ras_like_ GTPase	1e-09	1CEE, 1MR3, 1HE8	2.6 3.0 2.9	15.0 15.1 15.8	*G*ACGKTC G NK-D S*A*K	E*NY*	*A*G D*V*	G*DG*A CGKT	T	DT*A*G	NKKD	S*A*K	(1–4)
eIF2_ gamma	-	1KK1[a] 1SOU[b]	2.4 3.5	16.0 2.4	D-GKTT NK-E SAL	EELRRG ITIKIG	PGHEALM TTMLAGA SLMDG	GHVD HGKT	T	DAPG	NKIE	SAL	(1–7)
Era_like	2e-16	1LNX	-	-	SVGKST NK-D S*A*	TT*LV*P	*AD*L*P*GL*R*	G*F*PSV GKS	T	D*L*PG	NKXD	S*AV*	(1–4)
Era	2e-12	1WF3	2.5	16.9	PNVGKST S QTT G NK-D SA	GKISITS RKAQTT RHR	VDTPGL *G*D	GRPN VGKS	T	DTPG	NKVD	SAE	(1–4),8
Obg_like	1e-14	1LNX	-	-	SVGKST NK-D SA	FTTL*V*P	AD*L*PGLIR	GFPSV GKS	T	D*L*PG	NKXD	SA*V*	(1–4)
Obg	4e-11	1LNZ[c] 1UDX[d]	11.1 8.3	15.5 16.9	SVGKST NK-D SA	SA*K*PKIA DYHFTT L*V*P	GLIEGAH QGLGHQF LRHIERT	GFPSV GKS	T	DLPG	NK*X*D	SA*V*	(1–4)
DRG_Obg family	2e-08	*1EGA*	2.9	16.9	N-GKSA NQ-P D*PN*	SYPYTT *K*EP	VE*I*PPIY*E* GFTRGDG *S*KFVG *V*IRN	GPPN AGKS	T	E*L*PP	NKGD	SA*E*	(1–4)
EngA1	9e-12	1MKY	3.6	16.6	N-VGKST NK-E SAE	KAIV*E*D E TRDP	VDTCGV EA	GRPN VGKS	T	DTCG	NKAE	SAE	(1–4)
EngA2	9e-10	1MKY	3.6	16.6	N-GKST NK-D SAD	NRALVS PIPGTT RDP	GL KA	GRPN VGKS	T	DTAG	NKWD	SAD	(1–4), 9
YihA_EngB	9e-10	1SVW	2.8	16.3	SNAGKSS *TSK* GRT D K-D	PGRTQ LI	GYGYMK*R* *KWQR*AL*G* EYLE*KS*L	GRSN AGKS	T	D*L*PG	TK*A*D	SS*L*	(1–4)
trmE	4e-07	1XZP	3.4	10.8	N-GKST NK-D SA*L*	*N*EDRAI VTDIPG TTRD*V*	VDTAGVR *SE*TNVE*R* *L*GIERTLQ EIEK	GKPN VGKS	T	DTAG	NK*V*D	SA*L*	(1–4) ,10

Name	E-value	PDB ID	R.M.S.D.	Z-score	GTP/Mg2+ binding site	Switch I region	Switch II region	G1 box	G2 box	G3 box	G4 box	G5 box	
NOG	2e-08	*1EGA*	2.9	13.2	N-GKSS S-CD SVK	GKPEIC NYPFTT RGI	GLLRRC*D* *E*DRNNLE KLTL*A*VL*T* HL	GAPN VGKS	T	DTPG	SKCD	SV*K*	(1–4)
IF2_eIF5B	8e-09	1G7R 1G7T	2.9 2.7	12.2 13.9	VDHGKTT E NK-DR SAI	GITQHIG	PGHEAFT TLRKRGG ALADL	GHVD HGKT	T	DTPG	NKID	SAI	(1–4) (11–15)
Arl2l1_ Arl13 _like	1e-07	1KSG	2.7	17.0	DNAGKT*A* PT G NK-D SA*V*	GIQGHP EDVAPT VGF	DLGGG*K*R IRGIWKNY YAE	GLDN AGKT	T	DLGG	NKQD	SA*V*	(1–4), 6. 9, 13
GTP_ translation_ factor	5e-07	1TUI[e] 1EFU[f]	- 2.3	- 15.6	DHGKST NK-D SGL	ITI*T*	PGH*V*F*SS* EV*TAALT*D *G*	AHVD HGKS	T	DSPG	NKVD	SGL	(1–4), (16–18),
EF-G_bact	4e-05	1WDT	2.9	9.6	*A*GSGKTT TT G K-DK	*R*TTV	PGYG DF*V*GEI*R* *GA*LEAAD A	GH*A*G SGKT	T	DAPG	TKLD	-	(1–4), 19
Ar10_like	4e-06	2AL7	2.5	18.1	YSGKTT NK-DL SCK	VIA*S* FSEDM IPTVGF	DIGGQPR FRSMWER Y	GLQY SGKT	T	DIGG	NKRD	SCK	(1–4),20, (21–23)
Arl3	8e-05	1FZQ	2.7	16.6	NAGKTT NK-DL SAL	QLASED ISHI*T*PT QGF	DIGGQRKI R*P*YW*R*SY FEN	GLDN AGKT	T	DIGG	NKQD	SAL	(1–4),20, (21–23) , 24
Rab	0.001	3RAB	2.8	16.4	SSVGKTS FT G NK-D SAK	FTVGIK	GERYRITT AYY	GNSS VGKT	T	DTAG	NKCD	SAK	(1–4), 6, (25–28)

Note:

a. 1KK1 is the heterotrimeric factor e/aIF2 plays a central role in eukaryotic/archaeal initiation of translation.
b. 1SOU is eIF2-gamma from *Aquifex aeolicus* binds zinc, defined using 3.5 Å contacts.
c. 1LNZ shows Obg from *Bacillus subtilis* binds guanosine-5', 3'-tetraphosphate, a GTP analog and 2 Mg^{2+}, defined using 3.5 Å contacts.
d. 1UDX is the GTP-binding Protein Obg from *Thermus thermophilus* HB8.
e. 1TUI is the structure that EF-Tu from *Thermus aquaticus* binds GDP and Mg^{2+}, defined using 3.5 Å contacts.
f. 1EFU shows the EF-Tu interacts with EF-Ts, the guanine nucleotide exchange factor(GEF) for EF-Tu, in *Escherichia coli*, defined using 3.5 Å contacts.
g. Abbreviations in the table: HflX, High Frequency of Lysogenization X; Era-like, *E. coli* Ras-like protein; FeoB, Ferrous iron transport protein B; Obg, Spo0B-associated GTP-binding protein; IF2/eIF5B, initiation factors 2/ eukaryotic initiation factor 5B; NOG1, a nucleolar GTP-binding protein; DRG, developmentally regulated GTP-binding protein; Arf, ADP-ribosylation factor small GTPase.

Others: Eng A 1 and Eng A 2 subfamilies are composed of two adjacent GTPase domains. YihA is EngB subfamily. Arl2l1 (Arl2-like protein 1) and Arl13 form a subfamily of the Arf family of small GTPases; TrmE (MnmE, ThdF, MSS1) is a 3-domain protein found in bacteria and eukaryotes. Arl9/Arl10 was identified from a human cancer-derived EST dataset. Rab GTPases form the largest family within the Ras subfamily

h. The citation numbers in this table are correspondence to these literatures: 1– 5: (Buglino et al., 2002; Colicelli, 2004; Pandit and Srinivasan, 2003; Ruzheinikov et al., 2004; Wennerberg et al., 2005) ; 6– 7: (Roll-Mecak et al., 2004; Schmitt et al., 2002); 8. (Chen et al., 1999); 9. (Robinson et al., 2002); 10. (Scrima et al., 2005); 11–15: (Dahl et al., 2006; Hanzal-Bayer et al., 2002; Kawashima et al., 1996; Roll-Mecak et al., 2000; Sprang, 1997); 16–18: (Hanzal-Bayer and Hancock, 2007; Kawashima et al., 1996; Kremer et al., 2004); 19. (Polekhina et al., 1996); 20. (Scrima, 2005); 21–23: (Aridor et al., 2001; Bi et al., 2002; Menetrey et al., 2000; Pasqualato et al., 2002); 24. (Dumas et al., 1999); 25–28: (Itzen et al., 2006; Ostermeier and Brunger, 1999; Pereira-Leal and Seabra, 2001; Rak et al., 2003; Roll-Mecak et al., 2004).

REFERENCES

Allers, T. and Mevarech, M. (2005) Archaeal genetics - the third way. Nat Rev Genet, 6, 58-73.
Aridor, M., Fish, K.N., Bannykh, S., Weissman, J., Roberts, T.H., Lippincott-Schwartz, J. and Balch, W.E. (2001) The Sar1 GTPase coordinates biosynthetic cargo selection with endoplasmic reticulum export site assembly. J Cell Biol, 152, 213-229.
Arnold, F.H. and Volkov, A.A. (1999) Directed evolution of biocatalysts. Curr Opin Chem Biol, 3, 54-59.
Auchtung, T.A., Takacs-Vesbach, C.D. and Cavanaugh, C.M. (2006) 16S rRNA phylogenetic investigation of the candidate division "Korarchaeota". Appl Environ Microbiol, 72, 5077-5082.
Barns, S.M., Delwiche, C.F., Palmer, J.D. and Pace, N.R. (1996) Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. Proc Natl Acad Sci U S A, 93, 9188-9193.
Bell, S.D. and Jackson, S.P. (1998a) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. Trends Microbiol, 6, 222-228.
Bell, S.D. and Jackson, S.P. (1998b) Transcription in Archaea. Cold Spring Harb Symp Quant Biol, 63, 41-51.
Bell, S.D. and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. Curr Opin Microbiol, 4, 208-213.
Berken, A., Thomas, C. and Wittinghofer, A. (2005) A new family of RhoGEFs activates the Rop molecular switch in plants. Nature, 436, 1176-1180.
Bi, X., Corpina, R.A. and Goldberg, J. (2002) Structure of the Sec23/24-Sar1 pre-budding complex of the COPII vesicle coat. Nature, 419, 271-277.
Bintrim, S.B., Donohue, T.J., Handelsman, J., Roberts, G.P. and Goodman, R.M. (1997) Molecular phylogeny of Archaea from soil. Proc Natl Acad Sci U S A, 94, 277-282.
Bocchetta, M., Gribaldo, S., Sanangelantoni, A. and Cammarano, P. (2000) Phylogenetic depth of the bacterial genera Aquifex and Thermotoga inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. J Mol Evol, 50, 366-380.
Bos, J.L., Rehmann, H. and Wittinghofer, A. (2007) GEFs and GAPs: critical elements in the control of small G proteins. Cell, 129, 865-877.
Bourne, H.R., Sanders, D.A. and McCormick, F. (1990) The GTPase superfamily: a conserved switch for diverse cell functions. Nature, 348, 125-132.
Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. Nature, 349, 117-127.
Brinkman, A.B., Bell, S.D., Lebbink, R.J., de Vos, W.M. and van der Oost, J. (2002) The Sulfolobus solfataricus Lrp-like protein LysM regulates lysine biosynthesis in response to lysine availability. J Biol Chem, 277, 29537-29549.
Brinkman, A.B., Ettema, T.J., de Vos, W.M. and van der Oost, J. (2003) The Lrp family of transcriptional regulators. Mol Microbiol, 48, 287-294.
Buglino, J., Shen, V., Hakimian, P. and Lima, C.D. (2002) Structural and biochemical analysis of the Obg GTP binding protein. Structure, 10, 1581-1592.
Caldon, C.E. and March, P.E. (2003) Function of the universally conserved bacterial GTPases. Curr Opin Microbiol, 6, 135-139.
Chen, X., Court, D.L. and Ji, X. (1999) Crystal structure of ERA: a GTPase-dependent cell cycle regulator containing an RNA binding motif. Proc Natl Acad Sci U S A, 96, 8396-8401.
Chica, R.A., Doucet, N. and Pelletier, J.N. (2005) Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design. Current Opinion in Biotechnology, 16, 378-384.
Ciccarelli, F.D., Doerks, T., von Mering, C., Creevey, C.J., Snel, B. and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. Science, 311, 1283-1287.
Colicelli, J. (2004) Human RAS superfamily proteins and related GTPases. Sci STKE, 2004, RE13.
Conner, A.C., Hay, D.L., Simms, J., Howitt, S.G., Schindler, M., Smith, D.M., Wheatley, M. and Poyner, D.R. (2005) A key role for transmembrane prolines in calcitonin receptor-like receptor agonist binding and signalling: implications for family B G-protein-coupled receptors. Mol Pharmacol, 67, 20-31.
Crameri, A., Raillard, S.A., Bermudez, E. and Stemmer, W.P. (1998) DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature, 391, 288-291.
Crick, F. (1970) Central dogma of molecular biology. Nature, 227, 561-563.
Crick, F.H. (1958) On protein synthesis. Symp Soc Exp Biol, 12, 138-163.
Dahl, L.D., Wieden, H.J., Rodnina, M.V. and Knudsen, C.R. (2006) The importance of P-loop and domain movements in EF-Tu for guanine nucleotide exchange. J Biol Chem, 281, 21139-21146.

Danson, M.J. and Hough, D.W. (1998) Structure, function and stability of enzymes from the Archaea. Trends Microbiol, 6, 307-314.

DeLong, E.F. (2001) A phylogenetic perspective on hyperthermophilic microorganisms. Methods Enzymol, 330, 3-11.

DeLong, E.F. and Pace, N.R. (2001) Environmental diversity of bacteria and archaea. Syst Biol, 50, 470-478.

Doolittle, W.F. and Brown, J.R. (1994) Tempo, mode, the progenote, and the universal root. Proc Natl Acad Sci U S A, 91, 6721-6728.

Dumas, J.J., Zhu, Z., Connolly, J.L. and Lambright, D.G. (1999) Structural basis of activation and GTP hydrolysis in Rab proteins. Structure, 7, 413-423.

Dwyer, M.A., Looger, L.L. and Hellinga, H.W. (2004) Computational design of a biologically active enzyme. Science, 304, 1967-1971.

Geyer, M. and Wittinghofer, A. (1997) GEFs, GAPs, GDIs and effectors: taking a closer (3D) look at the regulation of Ras-related GTP-binding proteins. Curr Opin Struct Biol, 7, 786-792.

Giver, L., Gershenson, A., Freskgard, P.O. and Arnold, F.H. (1998) Directed evolution of a thermostable esterase. Proc Natl Acad Sci U S A, 95, 12809-12813.

Guarente, L. (1995) Transcriptional coactivators in yeast and beyond. Trends Biochem Sci, 20, 517-521.

Hanzal-Bayer, M., Renault, L., Roversi, P., Wittinghofer, A. and Hillig, R.C. (2002) The complex of Arl2-GTP and PDE delta: from structure to function. Embo J, 21, 2095-2106.

Hohn, M.J., Hedlund, B.P. and Huber, H. (2002) Detection of 16S rDNA sequences representing the novel phylum "Nanoarchaeota": indication for a wide distribution in high temperature biotopes. Syst Appl Microbiol, 25, 551-554.

Huber, H., Hohn, M.J., Rachel, R., Fuchs, T., Wimmer, V.C. and Stetter, K.O. (2002) A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. Nature, 417, 63-67.

Hwang, J. and Inouye, M. (2006) The tandem GTPase, Der, is essential for the biogenesis of 50S ribosomal subunits in Escherichia coli. Mol Microbiol, 61, 1660-1672.

Itzen, A., Pylypenko, O., Goody, R.S., Alexandrov, K. and Rak, A. (2006) Nucleotide exchange via local protein unfolding--structure of Rab8 in complex with MSS4. Embo J, 25, 1445-1455.

Jindra, M., Gaziova, I., Uhlirova, M., Okabe, M., Hiromi, Y. and Hirose, S. (2004) Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in Drosophila. Embo J, 23, 3538-3547.

Kabe, Y., Goto, M., Shima, D., Imai, T., Wada, T., Morohashi, K., Shirakawa, M., Hirose, S. and Handa, H. (1999) The role of human MBF1 as a transcriptional coactivator. J Biol Chem, 274, 34196-34202.

Kaper, T., Brouns, S.J., Geerling, A.C., De Vos, W.M. and Van der Oost, J. (2002) DNA family shuffling of hyperthermostable beta-glycosidases. Biochem J, 368, 461-470.

Karbstein, K. (2007) The role of GTPases in ribosome assembly. Biopolymers.

Kawashima, T., Berthet-Colominas, C., Wulff, M., Cusack, S. and Leberman, R. (1996) The structure of the Escherichia coli EF-Tu.EF-Ts complex at 2.5 A resolution. Nature, 379, 511-518.

Koonin, E.V., Wolf, Y.I. and Aravind, L. (2001) Prediction of the archaeal exosome and its connections with the proteasome and the translation and transcription machineries by a comparative-genomic approach. Genome Res, 11, 240-252.

Kremer, W., Steiner, G., Beraud-Dufour, S. and Kalbitzer, H.R. (2004) Conformational states of the small G protein Arf-1 in complex with the guanine nucleotide exchange factor ARNO-Sec7. J Biol Chem, 279, 17004-17012.

Kuchner, O. and Arnold, F.H. (1997) Directed evolution of enzyme catalysts. Trends Biotechnol, 15, 523-530.

Lee, D., Walsh, J.D., Mikhailenko, I., Yu, P., Migliorini, M., Wu, Y., Krueger, S., Curtis, J.E., Harris, B., Lockett, S., Blacklow, S.C., Strickland, D.K. and Wang, Y.X. (2006) RAP uses a histidine switch to regulate its interaction with LRP in the ER and Golgi. Mol Cell, 22, 423-430.

Leipe, D.D., Koonin, E.V. and Aravind, L. (2003) Evolution and classification of P-loop kinases and related proteins. J Mol Biol, 333, 781-815.

Liu, Q.X., Nakashima-Kamimura, N., Ikeo, K., Hirose, S. and Gojobori, T. (2007) Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator MBF1 and TATA-box-binding protein. Mol Biol Evol, 24, 1458-1463.

Looger, L.L., Dwyer, M.A., Smith, J.J. and Hellinga, H.W. (2003) Computational design of receptor and sensor proteins with novel functions. Nature, 423, 185-190.

Luke, K.A., Higgins, C.L. and Wittung-Stafshede, P. (2007) Thermodynamic stability and folding of proteins from hyperthermophilic organisms. Febs J, 274, 4023-4033.

Makarova, K.S. and Koonin, E.V. (2005) Evolutionary and functional genomics of the Archaea. Curr Opin Microbiol, 8, 586-594.

Marteinsson, V.T., Hauksdottir, S., Hobel, C.F., Kristmannsdottir, H., Hreggvidsson, G.O. and Kristjansson, J.K. (2001) Phylogenetic diversity analysis of subterranean hot springs in Iceland. Appl Environ Microbiol, 67, 4242-4248.

McCormick, F. (1989) ras GTPase activating protein: signal transmitter and signal terminator. Cell, 56, 5-8.

Menetrey, J., Macia, E., Pasqualato, S., Franco, M. and Cherfils, J. (2000) Structure of Arf6-GDP suggests a basis for guanine nucleotide exchange factors specificity. Nat Struct Biol, 7, 466-469.

Moore, J.C. and Arnold, F.H. (1996) Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nat Biotechnol, 14, 458-467.

Moreau, A., Yotov, W.V., Glorieux, F.H. and St-Arnaud, R. (1998) Bone-specific expression of the alpha chain of the nascent polypeptide-associated complex, a coactivator potentiating c-Jun-mediated transcription. Mol Cell Biol, 18, 1312-1321.

Olsen, G.J. and Woese, C.R. (1997) Archaeal genomics: an overview. Cell, 89, 991-994.

Ostermeier, C. and Brunger, A.T. (1999) Structural basis of Rab effector specificity: crystal structure of the small G protein Rab3A complexed with the effector domain of rabphilin-3A. Cell, 96, 363-374.

Otten, L.G. and Quax, W.J. (2005) Directed evolution: selecting today's biocatalysts. Biomol Eng, 22, 1-9.

Ouhammouch, M., Dewhurst, R.E., Hausner, W., Thomm, M. and Geiduschek, E.P. (2003) Activation of archaeal transcription by recruitment of the TATA-binding protein. Proc Natl Acad Sci U S A, 100, 5097-5102.

Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. Science, 276, 734-740.

Pandit, S.B. and Srinivasan, N. (2003) Survey for g-proteins in the prokaryotic genomes: prediction of functional roles based on classification. Proteins, 52, 585-597.

Pasqualato, S., Renault, L. and Cherfils, J. (2002) Arf, Arl, Arp and Sar proteins: a family of GTP-binding proteins with a structural device for 'front-back' communication. EMBO Rep, 3, 1035-1041.

Pereira-Leal, J.B. and Seabra, M.C. (2001) Evolution of the Rab family of small GTP-binding proteins. J Mol Biol, 313, 889-901.

Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C. and Nyborg, J. (1996) Helix unwinding in the effector region of elongation factor EF-Tu-GDP. Structure, 4, 1141-1151.

Rak, A., Pylypenko, O., Durek, T., Watzke, A., Kushnir, S., Brunsveld, L., Waldmann, H., Goody, R.S. and Alexandrov, K. (2003) Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. Science, 302, 646-650.

Reiter, E. and Lefkowitz, R.J. (2006) GRKs and beta-arrestins: roles in receptor silencing, trafficking and signaling. Trends Endocrinol Metab, 17, 159-165.

Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. Nature, 431, 152-155.

Robinson, V.L., Hwang, J., Fox, E., Inouye, M. and Stock, A.M. (2002) Domain arrangement of Der, a switch protein containing two GTPase domains. Structure, 10, 1649-1658.

Roll-Mecak, A., Alone, P., Cao, C., Dever, T.E. and Burley, S.K. (2004) X-ray structure of translation initiation factor eIF2gamma: implications for tRNA and eIF2alpha binding. J Biol Chem, 279, 10634-10642.

Roll-Mecak, A., Cao, C., Dever, T.E. and Burley, S.K. (2000) X-Ray structures of the universal translation initiation factor IF2/eIF5B: conformational changes on GDP and GTP binding. Cell, 103, 781-792.

Ruzheinikov, S.N., Das, S.K., Sedelnikova, S.E., Baker, P.J., Artymiuk, P.J., Garcia-Lara, J., Foster, S.J. and Rice, D.W. (2004) Analysis of the open and closed conformations of the GTP-binding protein YsxC from Bacillus subtilis. J Mol Biol, 339, 265-278.

Salminen, A. and Novick, P.J. (1987) A ras-like protein is required for a post-Golgi event in yeast secretion. Cell, 49, 527-538.

Scheffzek, K., Ahmadian, M.R., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F. and Wittinghofer, A. (1997) The Ras-RasGAP complex: structural basis for GTPase activation and its loss in oncogenic Ras mutants. Science, 277, 333-338.

Scrima, A., Vetter, I.R., Armengod, M.E. and Wittinghofer, A. (2005) The structure of the TrmE GTP-binding protein and its implications for tRNA modification. Embo J, 24, 23-33.

Sprang, S.R. (1997) G protein mechanisms: insights from structural analysis. Annu Rev Biochem, 66, 639-678.

Stanier, R.Y. and Van Niel, C.B. (1962) The concept of a bacterium. Arch Mikrobiol, 42, 17-35.

Stemmer, W.P. (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature, 370, 389-391.

Stetter, K.O. (1999) Extremophiles and their adaptation to hot environments. FEBS Lett, 452, 22-25.

Takai, K. and Horikoshi, K. (1999a) Genetic diversity of archaea in deep-sea hydrothermal vent environments. Genetics, 152, 1285-1297.

Takai, K. and Horikoshi, K. (1999b) Molecular phylogenetic analysis of archaeal intron-containing genes coding for rRNA obtained from a deep-subsurface geothermal water pool. Appl Environ Microbiol, 65, 5586-5589.

Takemaru, K., Harashima, S., Ueda, H. and Hirose, S. (1998) Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation. Mol Cell Biol, 18, 4971-4976.

Takemaru, K., Li, F.Q., Ueda, H. and Hirose, S. (1997) Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. Proc Natl Acad Sci U S A, 94, 7251-7256.

Thomas, C., Fricke, I., Scrima, A., Berken, A. and Wittinghofer, A. (2007) Structural evidence for a common intermediate in small G protein-GEF reactions. Mol Cell, 25, 141-149.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. FEMS Microbiol Rev, 18, 159-171.

Unsworth, L.D., van der Oost, J. and Koutsopoulos, S. (2007) Hyperthermophilic enzymes - stability, activity and implementation strategies for high temperature applications. Febs J, 274, 4044-4056.

Valetti, F. and Gilardi, G. (2004) Directed evolution of enzymes for product chemistry. Nat Prod Rep, 21, 490-511.

Vetter, I.R. and Wittinghofer, A. (2001) The guanine nucleotide-binding switch in three dimensions. Science, 294, 1299-1304.

Waters, E., Hohn, M.J., Ahel, I., Graham, D.E., Adams, M.D., Barnstead, M., Beeson, K.Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G.G., Simon, M., Soll, D., Stetter, K.O., Short, J.M. and Noordewier, M. (2003) The genome of Nanoarchaeum equitans: insights into early archaeal evolution and derived parasitism. Proc Natl Acad Sci U S A, 100, 12984-12988.

Watson, J.D. and Crick, F.H. (1953a) Genetical implications of the structure of deoxyribonucleic acid. Nature, 171, 964-967.

Watson, J.D. and Crick, F.H. (1953b) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. Nature, 171, 737-738.

Wennerberg, K., Rossman, K.L. and Der, C.J. (2005) The Ras superfamily at a glance. J Cell Sci, 118, 843-846.

Wittinghofer, A., Scheffzek, K. and Ahmadian, M.R. (1997) The interaction of Ras with GTPase-activating proteins. FEBS Lett, 410, 63-67.

Woese, C.R. and Fox, G.E. (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc Natl Acad Sci U S A, 74, 5088-5090.

Woese, C.R., Kandler, O. and Wheelis, M.L. (1990) Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. Proc Natl Acad Sci U S A, 87, 4576-4579.

Zhang, J.H., Dawes, G. and Stemmer, W.P. (1997) Directed evolution of a fucosidase from a galactosidase by DNA shuffling and screening. Proc Natl Acad Sci U S A, 94, 4504-4509.

Zimmermann, P. and Pfeifer, F. (2003) Regulation of the expression of gas vesicle genes in Haloferax mediterranei: interaction of the two regulatory proteins GvpD and GvpE. Mol Microbiol, 49, 783-794.

Chapter 2

Engineering a selectable marker for hyperthermophiles

Stan J.J. Brouns, Hao Wu, Jasper Akerboom, Andrew P. Turnbull, Willem M. de Vos, and John van der Oost

J Biol Chem (2005) **280**: 11422-11431

ABSTRACT

Limited thermostability of antibiotic resistance markers has restricted genetic research in the field of extremely thermophilic Archaea and bacteria. In this study, we used directed evolution and selection in the thermophilic bacterium *Thermus thermophilus* HB27 to find thermostable variants of a bleomycin-binding protein from the mesophilic bacterium *Streptoalloteichus hindustanus*. In a single selection round, we identified eight clones bearing five types of double mutated genes that provided *T. thermophilus* transformants with bleomycin resistance at 77 degrees, while the wild-type gene could only do so up to 65 degrees. Only six different amino acid positions were altered, three of which were glycine residues. All variant proteins were produced in *Escherichia coli* and analyzed biochemically for thermal stability and functionality at high temperature. A synthetic mutant resistance gene with low GC content was designed that combined four substitutions. The encoded protein showed up to 17 degrees increased thermostability and unfolded at 85 degrees in the absence of bleomycin, whereas in its presence the protein unfolded at 100 degrees. Despite these highly thermophilic properties, this mutant was still able to function normally at mesophilic temperatures *in vivo*. The mutant protein was co-crystallized with bleomycin, and the structure of the binary complex was determined to a resolution of 1.5 Å. Detailed structural analysis revealed possible molecular mechanisms of thermostabilization and enhanced antibiotic binding, which included the introduction of an intersubunit hydrogen bond network, improved hydrophobic packing of surface indentations, reduction of loop flexibility, and alpha-helix stabilization. The potential applicability of the thermostable selection marker is discussed.

INTRODUCTION

Despite the vast amount of protein sequences and structures from micro-organisms that grow optimally at temperatures above 80 °C, improving a protein's thermal stability is still a challenging task. This is mainly because the laws governing protein stability are not easily extracted, since they are highly variable and complex (1,2). It seems generally accepted that the extreme stability of certain natural proteins results from the cumulative effect of small adaptations in protein architecture and amino acid composition. Although some of these stabilizing features, such as optimized surface ion pair networks (3), are unlikely to be engineered into a protein of interest, other

strategies like α-helix capping (4), and the introduction of disulfide bonds and prolines in β-turns (5) can be applied very successfully when carefully designed on the basis of a high resolution crystal structure. However, in many cases atomic resolution three dimensional information of a protein is unavailable. Directed evolution approaches, by contrast, do not require any structural information, and commonly rely on random mutagenesis and recombination followed by screening or selection schemes (1,6). Thermostability screens of mutant libraries are usually carried out by applying a thermal challenge at nonpermissive temperatures, after which the remaining functionality of the individual clones is tested (7,8). To explore sufficient sequence space, requires the testing of large numbers of mutant clones, which necessitates high throughput approaches such as the use of robotics. Conversely, efficient selection procedures allow the testing of a large set of variants while reducing the effort of finding improved ones to a minimum.

A convenient selection system for finding protein variants in a library with improved thermostability is based on *in vivo* screening in a thermophilic expression host. Cloning and selection in thermophilic micro-organisms such as *Geobacillus stearothermophilus* (30 to 60 °C) or *Thermus thermophilus* (50 to 80 °C), mimics natural evolution, but is only applicable when the gene of interest encodes a protein that is of biological relevance to growth or survival of the host organism (9,10). The selective pressure can be fine-tuned by raising the temperature of growth, enabling only hosts that bear thermo-adapted variants to grow on solid media. For instance, a combination of *in vitro* mutagenesis methods and *in vivo* selection schemes have led to a highly thermostable kanamycin nucleotidyltransferase gene that is able to function at temperatures up to 79 °C (11). Such mutant selection markers have permitted the development of genetic tools which are very useful in the study of gene-function relationships in thermophilic bacteria (12).

In contrast to thermophiles (optimum temperature for growth 60 - 80 °C), antibiotic-based genetic systems for hyperthermophilic bacteria and archaea (optimum temperature for growth > 80 °C) are still in their infancy. This is primarily due to the absence of thermostable antibiotics and their corresponding resistance factors, since most known antibiotic producing micro-organisms are mesophilic bacteria and fungi. Often, the common antibiotics cannot be used, since many of them are unstable at high temperatures, or hyperthermophiles are simply insensitive to them (13). The

glycopeptide bleomycin is an exception, since it is a highly thermostable molecule and effective against many aerobic micro-organisms and eukaryotic cell-lines (14,15). The bleomycin family of antibiotics, including phleomycin and tallysomycin, are DNA- and RNA-cleaving glycopeptides that are produced by the actinomycetes *Streptoalloteichus hindustanus* and *Streptomyces verticillus*. As little as a few hundred bleomycin molecules can effectively kill aerobic cells (16). For this reason, bleomycin is currently clinically employed as an antitumor agent against squamous cell carcinomas and malignant lymphomas (17). Resistance against bleomycin-like antibiotics is conferred by N-acetylation, deamidation and sequestration of the molecule (15). The latter mechanism involves bleomycin binding proteins (BBPs) which have been found only in mesophilic Bacteria. Two proteins, Shble and BlmA, provide self-immunity for bleomycin-producers *St. hindustanus* and *S. verticillus*, respectively (14,18), and may be involved the transport and excretion of the molecule (19). Two genes, *blmT* and *blmS*, are located on the *Klepsiella pneumoniae* transposon Tn*5* (20) and on the *Staphylococcus aureus* plasmid pUB110 (21), respectively. All four proteins are highly negatively-charged cytoplasmic proteins of around 14 kDa, which form homodimers that bind two positively charged antibiotic molecules at a hydrophobic subunit interface cleft (19,22,23). The small protein size and the wide applicability of the drug have made both *shble* and *blmT* popular dominant selection markers in vector systems for lower and higher eukaryotes, bacteria and halophilic archaea (15,24,25). This prompted us to investigate whether we could thermostabilize Shble and BlmS to allow for its application in aerobic thermophiles and hyperthermophiles.

In this study, we have performed directed evolution using selection in the thermophilic bacterium *T. thermophilus*, and obtained various mutant proteins which could operate at highly thermophilic growth conditions. Their enhanced performance at high temperature was analyzed biochemically and possible stabilizing effects were identified.

RESULTS & DISCUSSION

Selection of stabilized variants of Shble

Randomly mutated *shble* genes were introduced in the *E. coli* - *T. thermophilus* shuttle vector pMK18 under the control of the promoter of the surface layer protein A (*slpA*) from *Thermus thermophilus* HB8 (26). This promoter is known to drive efficient transcription of the single selection marker in both bacteria. An error-prone library of approximately 20.000 functional clones was generated in *E. coli* HB101. Colonies appeared of similar size and there was no difference between the mutant and wild-type *shble* phenotype. The plasmid library was harvested and transformed into *Thermus thermophilus* HB27, making use of its high natural competence (27). *Thermus* clones appeared on bleomycin-containing plates up to 65 °C after transformation with the wild-type *shble* shuttle vector, whereas wild-type *blmS* was unable to generate a resistant phenotype at either 50 or 65 °C. The transformation efficiency of the *shble* shuttle vector was approximately five times lower at 65 °C than the kanamycin-based vector pMK18 (28). This difference might be due to the lethal effect of bleomycin and the non-catalytic nature of its elimination, which requires at least one protein molecule per bleomycin molecule.

Upon increasing the temperature of selection, a dramatic decrease in the number of colonies was observed after transformation of 8 μg of mutant library DNA. While 1200 colonies appeared at 67 °C, this number decreased to 800 at 69 °C, 600 at 70 °C, 106 at 75 °C and 8 at 77 °C. No colonies appeared at 78 and 80 °C. Plating efficiencies at these temperatures have been reported to be severely reduced, which complicates selection up to 85 °C, the maximum temperature of growth (11). The eight *Thermus* clones found at 77 °C (termed 77-1 to 77-8) were grown overnight in selective media at 70 °C and their plasmids were isolated, transformed into *E. coli* HB101, and subsequently re-isolated and their inserts were sequenced. This revealed that all variants were double mutants bearing, in total, 6 different amino acid substitutions and 3 silent mutations (Table 1). Five types of double mutants could be distinguished at the protein level and two sets of double mutants were identical. Remarkably, 3 out of 6 mutations found were glycine substitutions, of which glycine 98 was replaced by either a valine or a serine. The fact that only double mutants were found, seems to be a clear indication of the high stringency that was employed during selection. Interestingly, some substitutions, such as Leu63Gln, had occurred in combination with either Gly18Glu, Asp32Val or Gly98Val, which may point to the independent effects of the different mutations. A multiple sequence alignment of BBPs

and the position of the mutations are shown in Figure 1. To assess the reason why these mutants performed better at elevated temperatures *in vivo*, we produced and purified wild-type Shble and all double mutants and studied their biochemical behavior *in vitro*. Furthermore, a synthetic quadruple mutant gene with low GC content was designed by combining mutation Gly18Glu, Asp32Val, Leu63Gln and Gly98Val. The protein, designated HTS (High Temperature Shble), was produced, purified and biochemically analyzed. The HTS protein was crystallized in complex with bleomycin A2 and its structure determined.

Table 1. Nucleotide and amino acid substitutions of Shble variants

	Residue number					
	18	**31**	**32**	**40**	**63**	**98**
wild-type	Gly	Arg	Asp	Gly	Leu	Gly
77-1 / 77-2 / 77-8					Gln T188A	Val G293T
77-3[a]			Val A95T		Gln T188A	
77-4 / 77-6[b]		Leu G92T		Ala G119C		
77-5[c]		Leu G92T				Ser G292A
77-7	Glu G53A				Gln T188A	

Additional silent mutations: [a]C360A, [b]G63A, [c]G30A.

Thermal unfolding

Shble variants were subjected to temperature-induced equilibrium unfolding experiments in the presence and absence of bleomycin. The protein was found to unfold largely irreversible, since only 40% of the native folded signal was regained after slow cooling of the thermally unfolded protein. Therefore only apparent midpoint temperatures of unfolding (T_m) could be calculated. The results are summarized in Table 2.

In the absence of the antibiotic, wild-type Shble appears to be a very stable protein. This is remarkable, since *S. hindustanus* grows optimally at 28 °C (29). It is often found, however, that proteins for which low biological turnover is beneficial for a host, are prone to little local unfolding and hence are less susceptible to proteolytic attack (32). Structurally, Shble which serves a function of self-immunity, might well be adapted to meet these criteria by its compactness, relatively high secondary structure

content, high surface charge and by its embedded N and C-termini (19). The unfolding

Table 2. Apparent thermal unfolding midpoints (^{0}C) of Shble variants

		WT	HTS	77-1	77-3	77-4	77-5	77-7
apo	CD[a]	70.8 ± 0.5	84.7 ± 0.6	72.2 ± 0.5	79.5 ± 0.6	66.8 ± 0.5	71.1 ± 0.6	70.3 ± 0.5
	FS[b]	67.9 ± 0.4	78.7 ± 0.5	67.2 ± 0.4	69.1 ± 0.5	63.8 ± 0.4	64.6 ± 0.5	65.9 ± 0.4
	DSC[c]	67.4 ± 0.6	85.1 ± 0.7	N.D.	N.D.	N.D.	N.D.	N.D.
bleomycin	DSC[c]	94.7 ± 0.8	100.3 ± 0.8	N.D.	N.D.	N.D.	N.D.	N.D.

[a]CD: circular dichroism spectroscopy ($\lambda_{205\ nm}$), [b]FS: fluorescence spectroscopy ($\lambda_{ex\ 295\ nm}$, $\lambda_{em\ 315\ nm}$), [c] DSC: differential scanning calorimetry.

data also clearly show the strong stabilizing effect of ligand binding on the thermostability of the BBP, since the apparent unfolding midpoint temperature increases 27.3 °C upon bleomycin binding. This effect has also been recognized in other ligand binding proteins, such as streptavidin and avidin, which become extremely thermostable in the presence of biotin (33). In the absence of

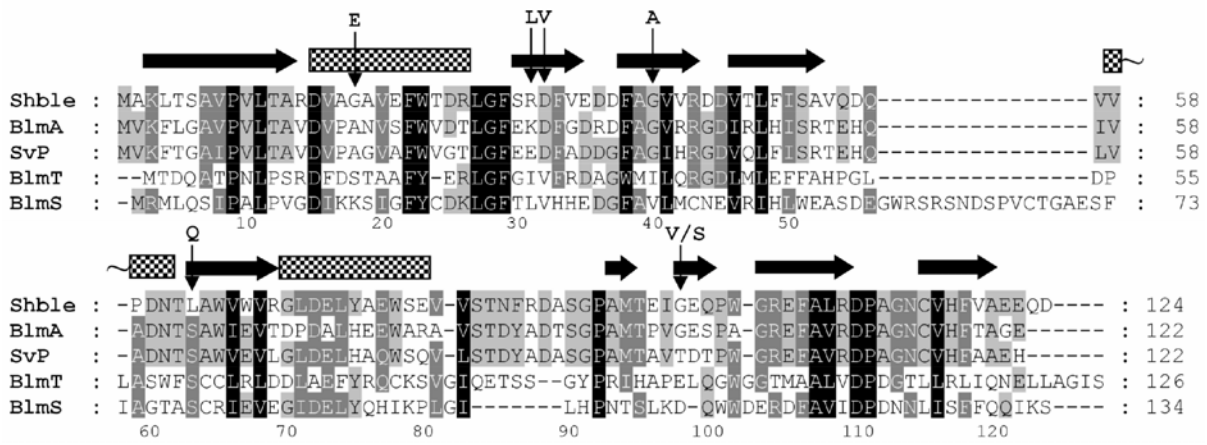


Figure 1. **Structural alignment of bleomycin binding proteins from different microbial sources.** Shble (PDB-ID: 1BYL) from *St. hindustanus* (19), BlmA (PDB-ID: 1QTO) from *S. verticillus* ATCC15003 (30), SvP from *S. verticillus* ATCC21890, BlmT (PDB-ID: 1ECS) from *K. pneumoniae* transposon Tn*5* (22) and BlmS from *S. aureus* plasmid pUB110. The alignment was created by backbone superimposition of the three structures and expanded with the SvP and BlmS sequences by realignment using ClustalX v1.81 (31) while maintaining the original gaps. The HTS structure was used for residue numbering and topology assignment (black arrows: β-strand, checkered boxes: α-helix). Mutations are indicated by arrows.

bleomycin, the stability of the various mutants is rather different. Of the double mutants, only 77-3 (Asp32Val, Leu63Gln) seems to have a marked increase in T_m as observed with circular dichroism spectroscopy (CD) and fluorescence spectroscopy (FS), while numbers 77-1 (Leu63Gln, Gly98Val), 77-5 (Arg31Leu, Gly98Ser) and 77-7 (Gly18Glu, Leu63Gln) remain virtually unchanged. Surprisingly, mutant 77-4 (Arg31Leu, Gly40Ala) displays significantly lower T_m values compared to the wild-type. Quadruple mutant HTS, which combines non-redundant mutations found in 77-1, 77-3 and 77-7, displays a profound increase of 13.9, 10.8 and 17.7 °C in stability in the absence of the antibiotic as found by CD, FS and differential scanning calorimetry (DSC), respectively. In its presence, the complex becomes hyperthermostable, unfolding at a temperature of just over 100 °C, 5.6 °C higher compared to the wild-type protein.

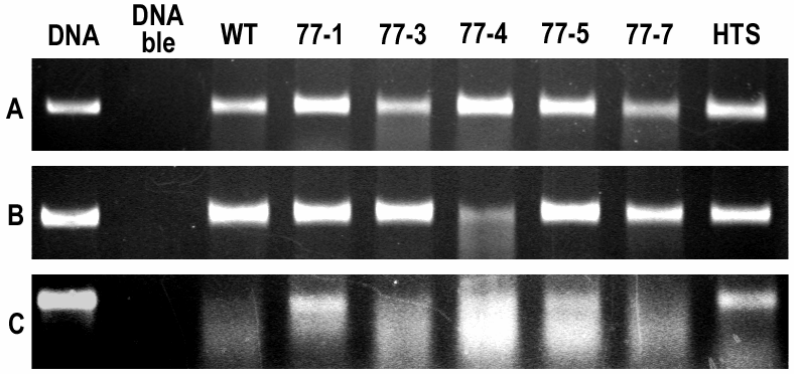


Figure 2. **DNA protection assay**. Digital photographs of 1% agarose gels showing the degree of DNA protection by Shble variants against the strand scission action of bleomycin A2. **A)** assay at 25 °C for 10 minutes. **B)** assay at 25 °C for 10 minutes after protein preincubation at 85 °C for 30 minutes. **C)** assay at 85 °C for 10 minutes.

To our surprise, the double mutants 77-1, 77-5 and 77-7 had almost unchanged apparent melting temperatures compared to the wild-type. This can be understood by realizing that *in vivo*, some amino acid changes may prevent instances of local protein unfolding, and therefore may avoid further unfolding and subsequent proteolytic attack. However, this is not necessarily reflected in its *in vitro* melting temperature, which is a measure of its global stability. Only when the weakest point of a structure was compensated (Asp32Val and Leu63Gln in 77-3), an increase of its melting temperature from 70.8 to 79.5 °C with CD, and 67.9 to 69.1 °C with FS was observed. Adding mutation Gly98Val from 77-1 and Gly18Glu from 77-7 to 77-3, giving rise to

HTS, further increased its melting temperature as one would expect. This observation is analogous to the findings of extensive work which has been conducted with the neutral protease from *Geobacillus stearothermophilus*, where interactions close to the N-terminus were found to be limiting the global stability (5).

Mutants improve DNA protection against bleomycin at high temperature

In vitro DNA protection assays were performed with the various Shble mutants in order to test whether the resistant phenotype of *T. thermophilus* at 77 °C was due to improved protection against the DNA degrading capability of bleomycin. The result of this is shown in Figure 2. At 25 °C, no significant differences in band intensities are observed. A 30-minute thermal pre-incubation of the protein at 85 °C, however, revealed a drastic loss of function in mutant 77-4. Differences between the wild-type and mutants became pronounced when bleomycin binding capabilities were tested at 85 °C. At this temperature, the DNA was protected best by 77-1 and HTS, followed by 77-4, 77-5, 77-3, 77-7 and the wild-type. Surprisingly, mutant 77-4, which displayed a low temperature unfolding midpoint and high thermal inactivation at 85 °C, apparently bound bleomycin effectively at high temperature conditions. So although the global stability of this mutant was decreased, it had improved bleomycin binding characteristics, which in itself, stabilizes the protein dramatically as observed by DSC measurements for the wild-type. These results indicate that some of the double mutants have improved the bleomycin binding properties compared to the wild-type, which confirms the findings of the *in vivo* selection procedure in *Thermus thermophilus*. Possible structural explanations for the improved functionality at higher temperature are discussed below.

Overall structure description

The quadruple mutant HTS was crystallized in the presence of bleomycin A2 and its structure was determined to 1.50 Å resolution. The crystals grown belong to space group $P2_1$ with unit cell parameters *a*= 44.0 Å, *b*= 66.6 Å, *c*= 47.2 Å and β= 117.4° and a dimer in the asymmetric unit (Figure 3A and 3B). Representative electron density is shown in Figure 3C. The structure forms a compact, homodimeric α/β protein of 121 amino acids (Met1, Gln123 and Asp124 are disordered) in which two bleomycin A2

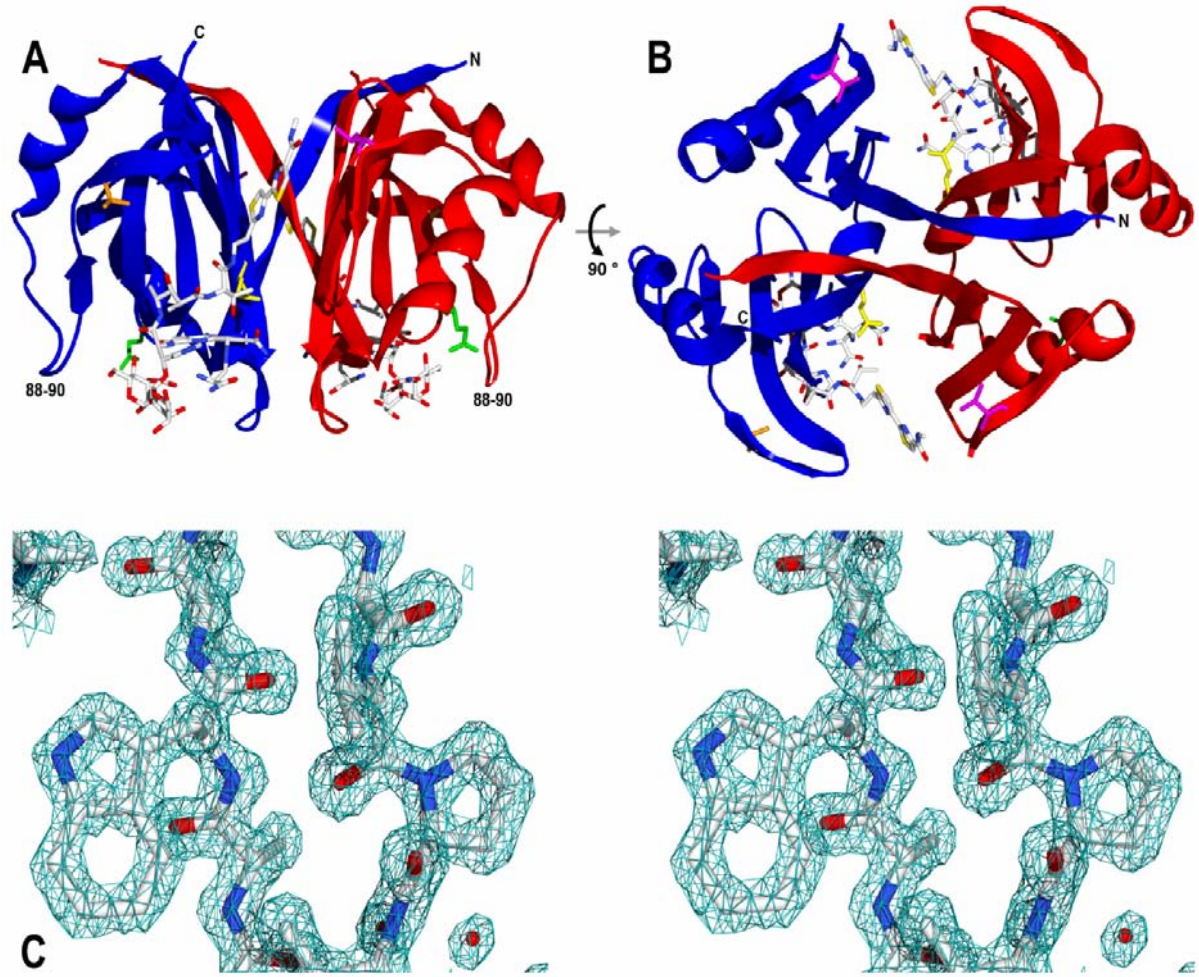


Figure 3. **Structure and electron density of HTS in complex with bleomycin A2**. Ribbon diagram showing the dimeric structure of the fourfold mutant Shble in complex with bleomycin A2. Mutations are indicated by stick representations. Chain A in blue, chain B in red, Gly18Glu in green, Asp32Val in pink, Leu63Gln in yellow and Gly98Val in orange. **A)** side view. **B)** viewed from the N- and C-terminal side (top view) **C)** stereo view of the electron density around residue Pro9 and Trp65 contoured at 2σ. Residues are colored according to the CPK color scheme, water molecules are represented by red spheres.

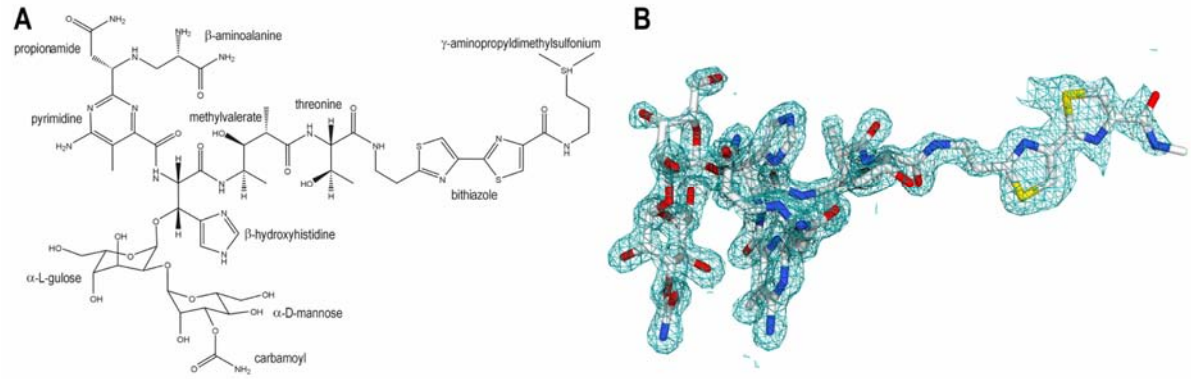


Figure 4. Chemical diagram and electron density of bleomycin A2. A) Schematic representation of bleomycin A2. **B)** Electron density around bleomycin A2 contoured at 1.5σ. The diagram indicates the missing electron density around the γ-aminopropyldimethylsulfonium moiety suggesting a disordered conformation.

molecules are accommodated in binding pockets at the dimer interface. These pockets consist of a hydrophilic concavity which runs into a hydrophobic intersubunit crevice. The dimer is maintained by alternate N-terminal β_1-strand hydrogen bonding between both monomers and by Van der Waals interactions at the largely hydrophobic subunit contact (19,30). Three sulphate ions are present at the surface of the dimer of which two form ion pairs with Arg104 of both chains. The presence of a dimer in the asymmetric unit allowed the identification of certain symmetry deviations between both monomers. A backbone superimposition of both chains (RMSD: 0.38 Å, Table 3) only revealed large differences in a random coil region comprising of residues Asp88, Ala89 and Ser90 (Figure 3A and 3B), which is spatially close to the carbamoyl group of the D-mannose moiety of bleomycin (Figure 4A). Their respective C_α atoms deviate 2.0, 5.1 and 1.6 Å in position, while giving rise to almost oppositely pointing amino acid side chains. In contrast to the bleomycin bound and unbound BlmA structure, backbone B-factors in this region are only marginally higher compared to the average

Table 3. Crystal structures of bleomycin binding proteins

Source/description	PDB ID	Form	Resolution (Å)	RMSD (Å)	Reference
Shble *St.hindustanus*	1BYL	apo	2.3	0.63[a]	(19)
	1XRK	bleomycin A2	1.5	0.38[b]	this study
BlmA *S. verticillus*	1QTO	Apo	1.5	0.77[a]	(30)
	1JIE	bleomycin A2	1.8	0.73[a]	(23)
	1JIF	Cu^{2+}-bleomycin A2	1.6	0.75[a]	
BlmT *K. pneumoniae*	1ECS	apo	1.7	1.18[a]	(22)
	1EWJ	bleomycin A2	2.5	1.15[a]	

[a]Averaged backbone superimposition RMSD values compared to HTS chain a and b, [b]Backbone superimposition RMSD values of HTS chain a to b

value, suggesting a rigid conformation (23,30). The difference in orientation of this loop might therefore be the result of sequential binding of two bleomycin molecules. Unlike BlmA, no symmetry related differences were observed in the region between amino acids 100 to 103. The topology of the HTS protein complex and the mode of bleomycin binding are similar to other BBPs. An overview of available structures is given in Table 3.

The structure of the HTS mutant in complex with bleomycin completes the list of structural information of three BBPs with and without their ligands, hereby contributing to our understanding of these proteins in general. Moreover, it has revealed several molecular features, which can account for increased protein stability and improved functionality at higher temperature *in vivo* and *in vitro*.

Structural effects of mutations

Introduction of an intersubunit hydrogen bond network

The structure of the dimer shows that each of the two bleomycin A2 molecules is bound by the concerted action of 21 amino acids. Due to its intersubunit location, both binding sites are composed of residues from either subunit. These include Val32, Phe33, Glu35, Phe38, Ser51, Ala52 and Val53 of one subunit and Pro59, Asp60, Asn61, Thr62, Gln63, Trp65, Phe86, Ala89, Trp102, Ala107, Arg109, Gly113, Cys115 and His117 of the other. The crystal structure clearly reveals the central role of mutation Leu63Gln which was found in 3 out of 5 different double mutants. Gln63 is involved in an extensive hydrogen bond network at the bottom of the bleomycin binding concavity (Figure 5A). It is noteworthy that the carbonyl side chains (Oε1) of both Gln63 residues in the dimer act as terminal hydrogen bond acceptors of a five-molecule water channel present at the dimer interface. A second hydrogen bond is accepted from the side chain hydroxyl group (Oγ) of Ser51 of the adjacent subunit. The amide side chain (Nε2) of Gln63 forms a hydrogen bond with one of two water molecules trapped between the bleomycin and the surface of the protein. The presence of a leucine at position 63 would most likely not have allowed for a hydrogen-bond network of this size. The advantage of an amino acid compatible with hydrogen bonding at position 63 is also evident from the alignment, which indicates that without exception the other four BBPs have a serine at this specific site (Figure 1). Although the mutant structure without bleomycin is not available, we speculate that an

intersubunit hydrogen bond between Gln63 and Ser51 can persist even without the antibiotic bound, giving rise to a beneficial interaction that might stabilize the dimer at high temperatures.

Hydrophobic packing of surface indentations

In wild-type Shble and BlmA, Asp32 is located on the edge of the intersubunit binding groove for the bithiazole moiety and tail-region of bleomycin (Figure 4A) (19,23,30). In bleomycin A2, B2 and in phleomycin D1 the tail is positively charged which suggests involvement of Asp32 in electrostatic stabilization or ligand recognition. From NMR studies it has become clear that in the bound state no strong interactions occur between the protein and the positively charged tail of bleomycin (34). These data are supported by the absence of electron density for the γ-aminopropyldimethylsulfonium moiety of bleomycin A2 in the binary complex structure of BlmA and HTS, suggesting a disordered conformation of the tail end (Figure 4B) (23). This might have allowed for an amino acid substitution to valine, which extends the hydrophobic bithiazole binding cleft at the dimer interface fitting nicely within a highly hydrophobic environment consisting of Phe33, Phe38, Val42, Thr47 and Phe49 (Figure 5B). In addition, both BlmT and BlmS sequences also contain a valine at the corresponding position (Figure 1). From a thermodynamic point of view, a mutation introducing surface hydrophobicity is generally believed to be unfavorable and has therefore rarely been investigated in directed mutagenesis studies. Nevertheless, some studies have reported significant improvements in protein stability by placing bulky hydrophobic amino acids at the surface of a neutral protease from *Geobacillus stearothermophilus* (35). Recent findings using *Bacillus licheniformis* α-amylase have clearly indicated that hydrophobic surface residues can indeed be extremely stabilizing by improving hydrophobic packing of surface indentations, hereby reinforcing subsurface secondary structure elements (36). This might explain the enhanced secondary structure preservation at high temperatures as inferred from CD of mutant 77-3. Strikingly, modeling of Gly40Ala into the wild-type structure (not shown) revealed close spatial proximity to Asp32 (5 Å between C_α and 3.5 Å between C_β atoms), which may also underline a similar need for hydrophobicity in this part of the protein. Mutation Arg31 to Leu, which occurred in two types of double mutants came as a surprise, since it is involved in a surface ion pair with Asp25 in wild-type Shble. Apparently this electrostatic interaction

does not counterweight beneficial effects of improved hydrophobic packing among residues Val20, Thr24, Val34, and Val41.

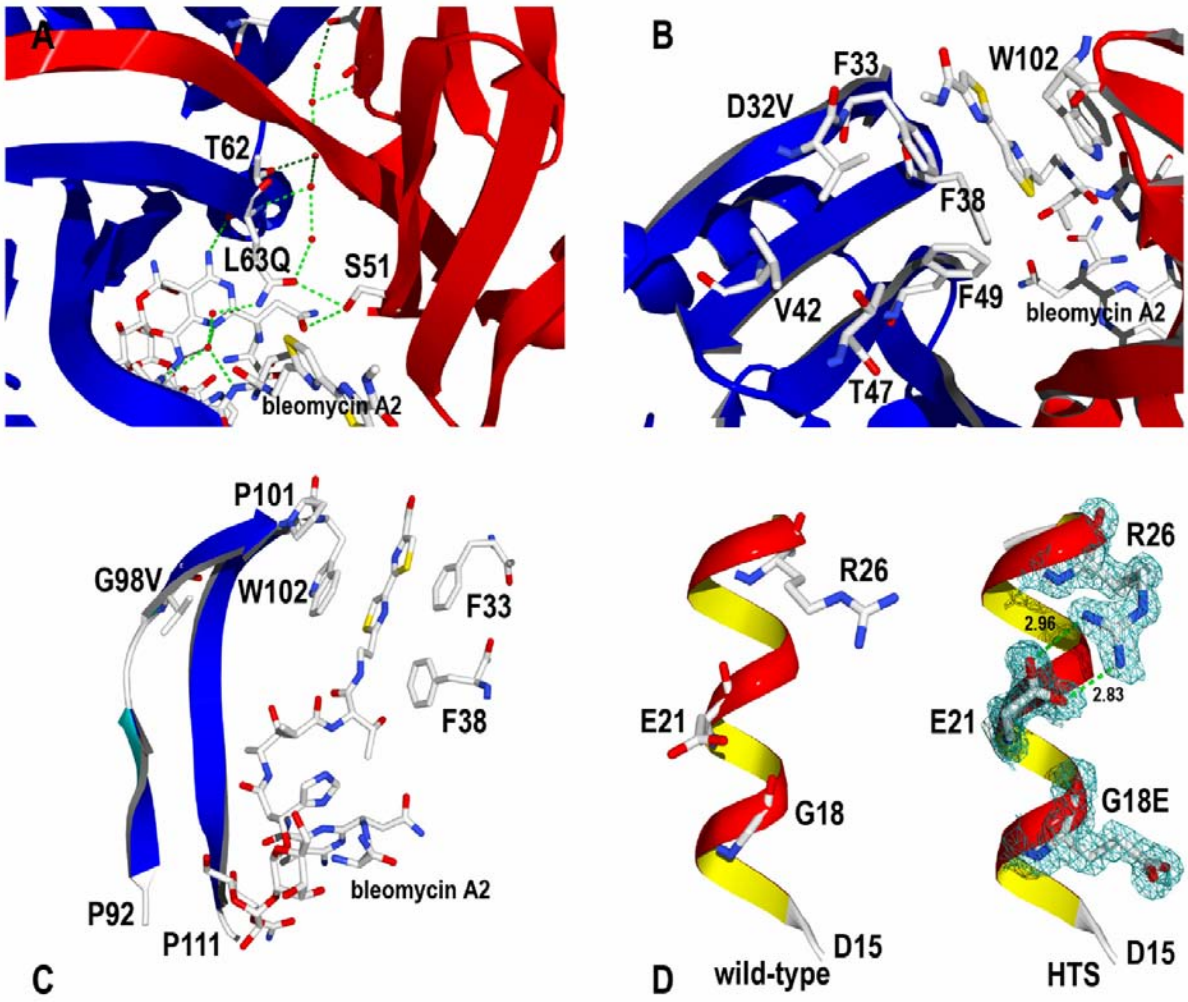


Figure 5. **Structural effects of the individual mutations. A)** Leu63Gln. Ribbon diagram showing the hydrogen bond network at the dimer interface. Thr62, Gln63, Ser51 and bleomycin A2 are shown together with the intersubunit water channel. **B)** Asp32Val. Ribbon diagram showing the hydrophobic intersubunit bleomycin tail binding crevice. Val32 may be involved in improved hydrophobic packing of this surface indentation among amino acids Phe33, Phe38, Val42, Thr47 and Phe49. **C)** Gly98Val. Ribbon representation showing a loop between Pro92 and Pro111 which is involved in bleomycin binding. Val98 is located at a former hinge region which enables Trp102 to stack the bithiazole tail against Phe33 and Phe38. The electron density revealed two alternative sidechain rotamers for Val98 in chain A (not shown) and a single side chain conformation in chain B. **D)** Gly18Glu. Side-by-side comparison of α-helix one formed between Asp15 and Leu27 in the wild-type and mutant crystal structures of Shble. The existence of a surface ion-pair between Glu21 and Arg26 is visible in the electron density (contoured at 1.5σ). Interatomic distances are indicated in Å. CPK color coding was used for amino acids and bleomycin A2. Chain A is indicated in blue, chain B in red. Water molecules are represented by red spheres. Hydrogen bonds are depicted by green dotted lines.

Reduction of surface loop flexibility

From previous crystallographic and NMR studies of BBPs, it has become clear that the loop following Gly98 in Shble will change its conformation upon binding of the antibiotic (22,23,30,34). This conformational change enables the tryptophan at position 102 to stack optimally with the hydrophobic bithiazole moiety of bleomycin (Figure 4A), packing both thiazole rings tightly against Phe33 and Phe38 of the adjacent subunit. In both BlmA and Shble, Gly98, located on the edge of a small β-strand leading towards the binding loop, seems to have a hinge function (Figure 5C). The bending motion of the backbone is also clearly reflected in large Phi and Psi torsion angle changes of more than 20° upon the binding of bleomycin. At high temperatures, however, this flexibility might have caused problems leading to local unfolding or a decreased bleomycin binding ability. A substitution for either a valine or serine as observed, would increase the rigidity of the loop and could therefore restore the binding capacity at high temperature.

α-Helix stabilization

Mutation Gly18Glu introduces a glutamate at position N3 in the first turn of the largest α-helix of the protein. Statistical analysis as well as experimental studies have shown that glutamates are energetically highly favored over glycines at the third position in an α-helix (37,38). This effect is most likely caused by the stabilizing effect of the negatively charged side chain on the helix macro dipole. To our surprise, chain A of the crystal structure revealed the formation of a genuine i, i + 5 α-helix surface ion-pair between Glu21 and Arg26, which was absent in the wild-type structure (Figure 5D). This new ion pair may have been the result of repulsion of anionic glutamate side chains of position 18 and 21, directing the latter towards the C-terminal arginine. Although i, i + 5 α-helical surface ion pairs do not give rise to strong ionic interactions at ambient temperatures (39), they might be more favorable at higher temperatures. Theoretical models have indicated that the energetic cost of desolvating charged groups is much less at 100 °C, due to a drop in the dielectric constant of water (40). This is currently the best explanation for the fact that proteins from extreme thermophiles have large ion pair networks at their surfaces, which are thought to be involved in maintaining structural integrity (3). Additionally, a minor beneficial effect of this mutation could be the introduction of additional negative surface charge, which

enhances electrostatic attraction of the cationic antibiotic under physiological conditions.

Laboratory versus natural evolution of thermostability

In this study, several possible mechanisms of adaptation to high temperature were identified, such as the introduction of a hydrogen bond network, improved hydrophobic packing of surface indentations, reduction of loop flexibility and α-helix stabilization. Remarkably, half of all mutations found were glycine replacements, which could point to protein stabilization by decreasing the entropy of the unfolded state (41). Although this could be a general strategy of stabilization, proteins from hyperthermophiles do not have a lower glycine content than their mesophilic counterparts, but rather a slightly increased one (42). Their predicted proteomes do have an increased propensity for charged (Arg, Lys and Glu) and bulky aliphatic (Ile and Val) amino acids, which has mostly come at the cost of polar residues (Asn, Gln, Ser and Thr) (42). This is fully in agreement with the requirements for the elevated numbers of surface salt bridges and improved hydrophobic core packing that has generally been recognized in these types of proteins. These are just two of a multitude of mechanisms that proteins from hyperthermophilic micro-organisms have employed to deal with extreme temperatures (2,43,44).

Recently, several other random mutagenesis studies have also reported large improvements in thermostability by applying directed evolution approaches. A mesophilic xylanase of family 11 was stabilized by over 35 °C by combining nine mutations found separately after extensive screening. The activity of this mutant was optimized by saturation mutagenesis of all mutated positions, yielding an enzyme variant with highly enhanced properties for high-temperature applications (8). In another study, a highly thermostable esterase containing seven mutations was evolved in six rounds of random mutagenesis, recombination and screening (7). The resulting enzyme was crystallized and its structure determined (45). The structure revealed that improved stability was due to altered core packing, α-helix stabilization, the introduction of surface salt bridges and reduction of flexibility in surface loops. From these and many other directed evolution and site-directed mutagenesis studies, it has become apparent that (i) proteins can be stabilized substantially by small numbers of mutations, (ii) these mutations are often located at the protein surface and

(iii) their effects are usually additive. As few as 2 out of 12 amino acid differences between a mesophilic and thermophilic cold shock protein turned out to be responsible for the difference in thermostability (46). The remaining variation in sequence might just have occurred as a result of neutral sequence drift or specific properties required by the host, such as solubility, turnover and molecular interactions (47).

Despite the fact that only a small number of mutations are required to render a protein thermostable, finding those mutations remains a difficult task. Apart from screening vast numbers of random mutants in microtiter plates, *in vitro* and *in vivo* selection schemes offer great advantages in order to reduce the effort to encounter improved variants. Currently, two main strategies are available.

One of these thermostabilizing selection methods is called Proside (Protein stability increased by directed evolution) which is based on the empirically derived inverse correlation between protein thermostability and proteolytic susceptibility (48). In a phage-display-like procedure, a library of mutants is fused between two domains of *E. coli* filamentous phage Fd gene-3-protein (G3P) which is then subjected to proteases while inducing local unfolding of the target protein by means of temperature or chemical denaturants. The resulting instable fusion protein variants are cleaved, causing only the surviving, more stable phages to be found after infection. It was found that in an ionic denaturant, non-polar surface interactions were optimized, whereas at elevated temperature variants with improved surface electrostatics were selected (47).

Cloning and selection in a thermophile, as conducted in this study, is a second directed evolution strategy which can lead to rapid improvements in thermostability. Despite its simplicity, very few studies using this technique have been reported. This is most likely due to its limited applicability, since the gene of interest rarely confers any biologically relevant function for growth or survival of the thermophile. Nonetheless, functionally stabilized mutants have been reported up to 79 °C for a kanamycin nucleotidyltransferase (9,11) and 58 °C for a chloramphenicol acetyltransferase (49). Auxotrophic knockout strains for leucine biosynthesis were complemented with thermolabile counterpart genes from *Bacillus subtilis* and *Saccharomyces cerevisiae* and adapted to higher temperature by serial accumulation of beneficial mutations in the *in-trans* introduced 3-isopropylmalate dehydrogenase genes (50,51). A hybrid α-galactosidase consisting of *B. stearothermophilus* and *T. thermophilus* peptide regions

was adapted to function at 67 °C by selection for growth on melibiose as the sole carbon and energy source (52).

Concluding remarks

In this study, several double mutants of a bleomycin binding protein were isolated with enhanced performance at high temperature both *in vivo* and *in vitro*. Structural analysis showed that the mutations gave rise to different means of stabilization in four parts of the protein. A combined mutant gene with a low GC content was created, which can serve as an antibiotic resistance marker for aerobic and micro-aerophilic mesophiles, thermophiles and hyperthermophiles. This may allow the development of efficient shuttle-vectors and knockout strategies for hyperthermophilic archaea and bacteria based on positive selection schemes. Moreover, the high GC content double mutant genes will now permit multigene knockout strategies in thermophiles such as *Thermus thermophilus*, allowing further exploration and exploitation of thermophilic microbial sources.

EXPERIMENTAL PROCEDURES

All chemicals were of analytical grade and purchased from Sigma. Primers were obtained from MWG Biotech AG (Ebersberg, Germany). Polymerase chain reactions were performed with *Pfu* TURBO (Stratagene) unless stated otherwise. Bleomycin A2 (Bleocin, Calbiochem) was used for all selections. *Escherichia coli* HB101 (*F- hsdS20* (r_B*-,* m_B*-*) *ara-14 galK2 lacY1 leuB6 mcrB mtl-1 proA2 recA13 rpsL20 supE44 thi-1 xyl-5* (*Str*^{R})) (53) was used for cloning purposes and routinely transformed by electroporation.

Generation of a bleomycin based shuttle vector

Bacillus subtilis 168 8G5 carrying pUB110 was kindly provided by Dr. S. Bron (University of Groningen, The Netherlands) and the plasmid was isolated by Qiagen Miniprep according to the manufacturer's instructions. The *blmS* gene was PCR amplified with primers BG1407 (sense) 5'-GGAGGTGCATATGAGAATGTTACAGTC TATCCC-3' and BG1240 (antisense) 5'-CGCGTCTAGATTAGCTTTTTATTTGTTG

AAAAAAG-3' (*Nde*I and *Xba*I sites underlined). Chromosomal DNA of *Streptoalloteichus hindustanus* (ATCC 31158) was prepared according to standard procedures (54) and used for PCR amplification of the *shble* gene with primers BG1410 (sense) 5'-TGAGGCATATGGCCAAGTTGACCAGTG CCG-3' and BG1411 (antisense) 5'-GATCCTCTAGATTAGTCCTGCTCCTCGGC CACG-3' (*Nde*I and *Xba*I sites underlined). PCR products were digested and ligated into *E. coli* - *T. thermophilus* shuttle-vector pMK18 (28) (Biotools, Madrid, Spain) thereby replacing the kanamycin nucleotidyltransferase gene. Ligation mixtures were transformed into *E. coli* HB101 and transformants were plated on 1.5% LB agar plates supplemented with 3 μg/ml bleomycin. Both *blmS* and *shble* provided resistance against the antibiotic, giving rise to the 4434 bp plasmid pWUR111 and the 4404 bp plasmid pWUR112, respectively.

Mutant library construction

Error-prone PCR was carried out using two different polymerases, namely *Taq* (Amersham) and Mutazyme (Genemorph kit, Stratagene). This approach was chosen to complement the transition and transversion bias of each enzyme to provide a more complete mutational spectrum of the PCR product. For the error-prone amplification, flanking primers BG1412 (sense) 5'-CGACCCTTAAGGAGGTGTGAGGCAT**ATG**-3' and BG1408 (antisense) 5'-CGAGCTCGGTACCCGGGGATCCTCTAGA**TTA**-3' (*Nde*I and *Xba*I site underlined) were designed to allow variation throughout the entire coding sequence, between the start- and stop-codon (indicated in boldface). *Taq* polymerase based PCR reactions were performed as previously described (55). A 50 μl PCR reaction contained 5 ng of pWUR112, 5 pmol of each primer, 0.2 mM of dATP and dGTP, 1 mM of dCTP and dTTP, 5 U of polymerase, 3 mM $MgCl_2$ and three concentrations of $MnCl_2$ (0.1, 0.3 and 0.5 mM). The mixture was thermocycled as follows: 95 °C (4 min), 30 cycles of 94 °C (30 sec), 55 °C (45 sec) and 72 °C (25 sec) post-dwelled for 4 min at 72 °C. Mutazyme PCR reactions were prepared according to the manufacturer's instructions and thermocycled as above using an elongation time of 50 seconds.

Randomly mutated PCR products were cloned into vector pMK18 and transformed into *E. coli* HB101. A total of approximately 10.000 *Taq* and 10.000 Mutazyme derived clones were resuspended in 50 ml LB media supplemented with

antibiotic and grown in 1 L media to early stationary phase. Plasmids were subsequently harvested using a Miniprep plasmid isolation kit (Qiagen).

Selection in *Thermus thermophilus*

Thermus thermophilus HB27 was kindly provided by Dr. J. Berenguer (Autonomous University of Madrid, Spain). Cells were routinely cultivated at 70 °C in a Ca^{2+} (3.9 mM) and Mg^{2+} (1.9 mM) rich media (28) containing 8 g/l tryptone, 4 g/l yeast extract and 3 g/l NaCl dissolved in Evian mineral water (pH 7.7, after autoclaving) (Evian-les-Bains, France). Transformation of *T. thermophilus* was essentially performed by the method of Koyama (27). Frozen cell aliquots were resuspended in 25 ml of media and grown at 150 rpm to an OD_{600} of 0.8. The culture was then diluted 1:1 in preheated media and incubated for another hour. Next, plasmids were added to 0.5 ml of culture and the mixture was incubated for 2 - 3 hours at 70 °C with occasional shaking before being plated on 3% agar plates (Beckton Dickinson) supplemented with 30 μg/ml kanamycin or 15 μg/ml bleomycin (Calbiochem) for selection. Colonies appeared within 36 hours at 60 - 70 °C. At temperatures above 70 °C, 1% Gelrite plates (Roth, Karlsruhe, Germany) were used, supplemented with 100 μg/ml kanamycin or 20 μg/ml bleomycin for selection. Colonies were grown overnight in liquid media containing 30 μg/ml kanamycin or 5 μg/ml bleomycin. *T. thermophilus* plasmid DNA was prepared using a plasmid Miniprep-kit (Qiagen) after a 2 mg/ml pre-incubation with lysozyme for 30 minutes at 37 °C.

Gene cloning, overexpression and protein purification

Wild-type and double mutant *shble* genes were PCR amplified from their respective pWUR112 plasmids using primers BG1503 (sense) 5'-GATGGCCATGGCCAAGTTGACCAGTGC-3' and BG1504 (antisense) 5'-GCCGCAAGCTTAGTCCTGCTCCTCGGCC-3' (*Nco*I and *Hin*dIII site underlined). PCR products were cloned into vector pET26b (Novagen) and fused to an *Erwinia carotovora* pectate lyase (*pelB*) signal sequence allowing periplasmic protein overexpression in *E. coli* BL21(DE3) (Novagen). Periplasmic fractions of 1 L cultures were prepared by osmotic shock according to the manufacturers instructions and dialyzed overnight against 20 mM Tris-HCl (pH 7.5). Samples were loaded onto a MonoQ HR 5/50 connected to a FPLC system (Amersham) and eluted using a 1 M

NaCl gradient. Shble containing fractions were pooled and dialyzed against a 10 mM NaP_i buffer (pH 7.0) supplemented with 50 mM NaCl and subsequently purified by size exclusion chromatography using a Superdex 200 HR 10/30 column (Amersham).

Synthetic gene construction

A synthetic mutant *shble* gene based on archaeal codon usage was constructed by oligonucleotide assembly PCR (56). This gene contains the point mutations Gly18Glu, Asp32Val, Leu63Gln and Gly98Val, and has a GC content of 40.8% compared to 70.2 % of wild-type *shble*. The synthetic gene was denominated HTS (High Temperature Shble). The sequence has been deposited to Genbank (accession number AY780486).

Assembly PCR mixtures contained 10 oligonucleotides (BG1542 - BG1451, ONLINE Suppl. Table) with an overlap of 20 bases. Both flanking primers were 40 bases in length whereas the 8 central primers consisted of 80 to 90 bases. The assembly PCR mixture contained 2.5 μM of each primer, 0.2 mM of dNTP's and 0.05 U/μl of Pfu polymerase. The mixture was thermocycled at 94 °C (30 sec), 55 °C (30 sec) and 72 °C (60 sec) for 40 cycles. The PCR products were purified over a PCR purification column (Qiagen) and diluted 1:1 in fresh PCR mix containing only both flanking primers BG1542 and BG1551 at 0.1 μM concentration and were thermocycled according to standard procedures. PCR products of the expected size were isolated from agarose gel using Qiaex II gel extraction kit (Qiagen), digested with *Nde*I and *Bgl*II and cloned into vector pET26b. This allowed for efficient cytoplasmic protein overproduction in *E. coli* BL21(DE3)-RIL (Novagen). Positive clones were picked from LB agar plates containing 3 μg/ml bleomycin and 50 μg/ml chloramphenicol. A 4 L culture was grown at 37 °C until OD_{600} 0.5, induced with 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and was incubated for another 5 hours. Cells were harvested, resuspended in 20 mM Tris-HCl (pH 7.5) and sonicated. Cell extracts were clarified by centrifugation (30 min, 26,500xg, 4 °C) and applied to a 70 ml Q-Sepharose Fast Flow (Amersham) anion exchange column. Proteins were eluted by a 1 M NaCl gradient, High Temperature Shble (HTS) containing fractions were pooled and concentrated over a YM10 filter (Amicon) and further purified by size exclusion chromatography as described above.

DNA sequencing

Inserts of plasmids used in this study were sequenced by Westburg Genomics (Wageningen, The Netherlands).

Protein quantitation

Protein concentrations were determined by using a Bradford assay (57) (Bio-Rad). Purified proteins were quantified from A_{280} measurements using a protein extinction coefficient of 29000 $M^{-1}cm^{-1}$ (58).

DNA protection assay

Protein functionality assays were essentially performed as described elsewhere (14), employing a 10-fold excess molar concentration of protein over bleomycin A2 (Calbiochem). In each assay, 0.2 µg of *Pst*I linearized plasmid, pUC19, was used. Assays were performed by first incubating DNA and protein shortly at 85 °C, after which bleomycin A2, dithiothreitol (DTT), and Fe^{2+} were sequentially added to the reaction mixture.

Circular dichroism spectroscopy

CD experiments were performed on a Jasco J-715 spectropolarimeter (Jasco, Tokyo, Japan) equipped with a PTC-348WI Peltier temperature control system. Far-UV CD measurements were conducted with Suprasil quartz cuvettes (Hellma Benelux, Rijswijk, The Netherlands) with a 1 mm cell length. During all experiments, the sample cell chamber was purged by dry N_2 gas at a flow rate of 10 L/min. In temperature-induced unfolding experiments, the cuvette containing 1.7 µM of protein sample in degassed 10 mM NaP_i (pH 7.0) and 50 mM NaCl, was heated from 25 to 95 °C at 0.4 °C/min and subsequently cooled to 25 °C at the same rate. The ellipticity at 205 nm was measured every 0.5 °C with a 2 second response time to monitor the loss of β-sheet and β-turn secondary structure elements. The bandwidth of the measurement was set to 1.0 nm and the sensitivity to 100 mdeg. Data were corrected for the temperature dependent ellipticity of a blanc without protein. Averaged data of two independent scans were fit according to a two state model of unfolding and the apparent temperature unfolding midpoint (T_m) was derived from van 't Hoff plots.

Fluorescence spectroscopy

Fluorescence experiments were performed on a Varian Cary Eclipse fluorimeter (Varian, Middelburg, The Netherlands) equipped with a four-cuvette Peltier multicell holder and PCB-150 waterbath. All measurements were performed in 3 ml Suprasil quartz cuvettes (Hellma Benelux, Rijswijk, The Netherlands) with a 1 cm pathlength. A magnetic stirring bar ensured a homogeneous sample temperature. The temperature of the sample was recorded by a temperature probe inside one of the four samples. Spectra and thermal unfolding curves were recorded of 1.7 μM protein solutions in degassed 10 mM NaP_i (pH 7.0) buffer supplemented with 50 mM NaCl. Tryptophans were excited at 295 nm and fluorescence was recorded from 300 to 550 nm with both the excitation and the emission slits set to 5 nm. During temperature-induced unfolding and refolding studies, fluorescence emission intensities were monitored at 315 nm from 27 to 92 °C at a heating and cooling rate of 0.4 °C/min. Data were corrected for the fluorescence emission of corresponding blanc solutions. Data of two independent scans were treated and fit as described above.

Differential scanning calorimetry (DSC)

DSC measurements were performed on a Microcal III system (Setaram, Caluire, France). Degassed protein samples of 0.28 mg/ml (20 μM) in 10 mM NaPi (pH 7.0) and 50 mM NaCl in the presence and absence of an 8-fold molar excess bleomycin A2, were heated from 20 to 120 °C at 0.5 °C/min. Midpoint temperatures of unfolding were determined by curved baseline analysis from two independent scans.

Protein crystallization, data collection and processing

The HTS protein was extensively dialyzed against 10 mM NaPi, pH 7.0 and was subsequently crystallized by the sitting drop method of vapor diffusion at 20 °C and a protein concentration of 3.3 mg/ml in the presence of a 10 fold molar excess of bleomycin A2 - HCl (Calbiochem). Crystals grew optimally using 2.0 M Ammonium-Sulfate as the precipitant in 0.1 M Sodium-Acetate buffer, pH 4.6. Data were collected from a single flash-frozen native crystal (100K) to 1.5 Å resolution using a MAR345 imaging plate at the Protein Structure Factory beamline BL14.2 of the Free University of Berlin at the BESSY synchrotron source (Berlin, Germany) (Table 4). All data were reduced with DENZO and SCALEPACK (59). The crystal used for data collection had

unit cell parameters *a*= 44.0 Å, *b*=66.6 Å and *c*=47.2 Å and *β*=117.4° and belonged to space group $P2_1$ with a dimer in the asymmetric unit.

Table 4. Data collection and refinement statistics

X-ray data collection statistics	
Wavelength (Å)	0.90830
Resolution (Å)	30 – 1.5 (1.53 – 1.50)
Total observations	87740 (3602)
Unique observations	36444 (1714)
Completeness (%)	94.4 (88.6)
$I / \sigma(I)$	20.3 (3.9)
R_{sym} [a]	0.043 (0.189)
Refinement statistics	
Resolution (Å)	30.0 - 1.5
R_{work} [b]	0.174
R_{free} [c]	0.194
RMSD bond distances (Å)	0.01
RMSD bond angles (°)	1.664
Total number of non-H atoms	1883
Av. protein *B* value (Å^2)	13.3
Number of solvent molecules	288
Av. solvent *B* value (Å^2)	27.0
Number of bleomycin atoms	182
Av. ligand *B* value (Å^2)	22.8
Number of Sulphate ion atoms	15
Av. Sulphate ion B value (Å^2)	31.4

[a] $R_{sym} = \Sigma_{hkl} \Sigma_i |I-\langle I_i \rangle| / \Sigma_{hkl} \Sigma_i I_i$, where I_i is the intensity of a given measurement and the sums are over all measurements and reflections. Values in parentheses refer to the highest resolution shell.

[b] $R_{work} = \Sigma ||F(obs)| - |F(calc)|| / \Sigma |F(obs)|$ for the 95% of the reflection data used in refinement.

[c] $R_{free} = \Sigma ||F(obs)| - |F(calc)|| / \Sigma |F(obs)|$ for the remaining 5%.

The structure of HTS was determined by molecular replacement using the program MOLREP (60) and the *Streptomyces verticillus* BlmA dimer (PDB-ID: 1JIE) (23) as the search model. The initial phases were improved using the free-atom refinement method together with automatic model tracing in ARP/wARP (61). TLS parameters were determined and TLS-restrained refinement was performed using REFMAC (62). Several rounds of iterative model building and refinement followed and

water molecules were added using ARP/wARP (61). The final model (comprising 241 amino acids, 288 water molecules, 2 bleomycin molecules and 3 sulphate ions), refined using data between 30 and 1.5 Å resolution, has an R- and free R-factor of 17.4 and 19.4 %, respectively, with good geometry. Residues Met1, Glu122, Gln123 and Asp124 in chain A and Met1, Gln123 and Asp124 in chain B are not visible in the electron density map and therefore have been excluded from the model. Additionally, the side chains of two residues in chain A (Asp36 and Arg87), four side chains in chain B (Glu21, Asp36, Arg87 and Glu122), and the γ-aminopropyldimethylsulfonium moiety of bleomycin have been truncated in the final model. The stereochemical quality of the model and the model fit to the diffraction data were analyzed with the programs PROCHECK (63) and SFCHECK (64).

The coordinates and experimental structure factors have been deposited in the Protein Data Bank with accession number 1XRK. Figures were prepared with Swiss-PDBviewer v3.7 SP5 (65) and rendered with POV-Ray v3.6.

ACKNOWLEDGEMENTS

The authors thank Dr. J. Berenguer for helpful suggestions and Anton Korteweg for technical assistance with DSC. This work was supported by a grant from the European Union in the framework of the SCREEN project (contract QLK3-CT-2000-00649).

REFERENCES

1. Arnold, F. H., Wintrode, P. L., Miyazaki, K., and Gershenson, A. (2001) *Trends Biochem Sci* **26**(2), 100-106
2. Vieille, C., and Zeikus, G. J. (2001) *Microbiol Mol Biol Rev* **65**(1), 1-43
3. Karshikoff, A., and Ladenstein, R. (2001) *Trends Biochem Sci* **26**(9), 550-556
4. Serrano, L., and Fersht, A. R. (1989) *Nature* **342**(6247), 296-299
5. Van den Burg, B., Vriend, G., Veltman, O. R., Venema, G., and Eijsink, V. G. (1998) *Proc Natl Acad Sci U S A* **95**(5), 2056-2060
6. Stemmer, W. P. (1994) *Nature* **370**(6488), 389-391
7. Giver, L., Gershenson, A., Freskgard, P. O., and Arnold, F. H. (1998) *Proc Natl Acad Sci U S A* **95**(22), 12809-12813
8. Palackal, N., Brennan, Y., Callen, W. N., Dupree, P., Frey, G., Goubet, F., Hazlewood, G. P., Healey, S., Kang, Y. E., Kretz, K. A., Lee, E., Tan, X., Tomlinson, G. L., Verruto, J., Wong, V. W., Mathur, E. J., Short, J. M., Robertson, D. E., and Steer, B. A. (2004) *Protein Sci* **13**(2), 494-503
9. Liao, H., McKenzie, T., and Hageman, R. (1986) *Proc Natl Acad Sci U S A* **83**(3), 576-580
10. Tamakoshi, M., Yamagishi, A., and Oshima, T. (1995) *Mol Microbiol* **16**(5), 1031-1036
11. Hoseki, J., Yano, T., Koyama, Y., Kuramitsu, S., and Kagamiyama, H. (1999) *J Biochem (Tokyo)* **126**(5), 951-956
12. Friedrich, A., Prust, C., Hartsch, T., Henne, A., and Averhoff, B. (2002) *Appl Environ Microbiol* **68**(2), 745-755
13. Noll, K. M., and Vargas, M. (1997) *Arch Microbiol* **168**(2), 73-80
14. Gatignol, A., Durand, H., and Tiraby, G. (1988) *FEBS Lett* **230**(1-2), 171-175
15. Sugiyama, M., and Kumagai, T. (2002) *J Biosci Bioeng* **93**(2), 105-116
16. Poddevin, B., Orlowski, S., Belehradek, J., Jr., and Mir, L. M. (1991) *Biochem Pharmacol* **42 Suppl**, S67-75
17. Sikic, B. I. (1986) *Cancer Surv* **5**(1), 81-91
18. Sugiyama, M., Thompson, C. J., Kumagai, T., Suzuki, K., Deblaere, R., Villarroel, R., and Davies, J. (1994) *Gene* **151**(1-2), 11-16
19. Dumas, P., Bergdoll, M., Cagnon, C., and Masson, J. M. (1994) *Embo J* **13**(11), 2483-2492
20. Genilloud, O., Garrido, M. C., and Moreno, F. (1984) *Gene* **32**(1-2), 225-233
21. Semon, D., Movva, N. R., Smith, T. F., el Alama, M., and Davies, J. (1987) *Plasmid* **17**(1), 46-53
22. Maruyama, M., Kumagai, T., Matoba, Y., Hayashida, M., Fujii, T., Hata, Y., and Sugiyama, M. (2001) *J Biol Chem* **276**(13), 9992-9999
23. Sugiyama, M., Kumagai, T., Hayashida, M., Maruyama, M., and Matoba, Y. (2002) *J Biol Chem* **277**(3), 2311-2320
24. Drocourt, D., Calmels, T., Reynes, J. P., Baron, M., and Tiraby, G. (1990) *Nucleic Acids Res* **18**(13), 4009
25. Nuttall, S. D., Deutschel, S. E., Irving, R. A., Serrano-Gomicia, J. A., and Dyall-Smith, M. L. (2000) *Biochem J* **346 Pt 2**, 251-254
26. Fernandez-Herrero, L. A., Olabarria, G., and Berenguer, J. (1997) *Mol Microbiol* **24**(1), 61-72
27. Koyama, Y., Hoshino, T., Tomizuka, N., and Furukawa, K. (1986) *J Bacteriol* **166**(1), 338-340
28. de Grado, M., Castan, P., and Berenguer, J. (1999) *Plasmid* **42**(3), 241-245
29. Tomita, K., Nakakita, Y., Hoshino, Y., Numata, K., and Kawaguchi, H. (1987) *Int J Syst Bacteriol* **37**, 211-213
30. Kawano, Y., Kumagai, T., Muta, K., Matoba, Y., Davies, J., and Sugiyama, M. (2000) *J Mol Biol* **295**(4), 915-925
31. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) *Nucleic Acids Res* **25**(24), 4876-4882
32. Jaswal, S. S., Sohl, J. L., Davis, J. H., and Agard, D. A. (2002) *Nature* **415**(6869), 343-346
33. Gonzalez, M., Argarana, C. E., and Fidelio, G. D. (1999) *Biomol Eng* **16**(1-4), 67-72
34. Vanbelle, C., Brutscher, B., Blackledge, M., Muhle-Goll, C., Remy, M. H., Masson, J. M., and Marion, D. (2003) *Biochemistry* **42**(3), 651-663
35. Van den Burg, B., Dijkstra, B. W., Vriend, G., Van der Vinne, B., Venema, G., and Eijsink, V. G. (1994) *Eur J Biochem* **220**(3), 981-985

36. Machius, M., Declerck, N., Huber, R., and Wiegand, G. (2003) *J Biol Chem* **278**(13), 11546-11553
37. Iqbalsyah, T. M., and Doig, A. J. (2004) *Protein Sci* **13**(1), 32-39
38. Penel, S., Hughes, E., and Doig, A. J. (1999) *J Mol Biol* **287**(1), 127-143
39. Smith, J. S., and Scholtz, J. M. (1998) *Biochemistry* **37**(1), 33-40
40. Elcock, A. H. (1998) *J Mol Biol* **284**(2), 489-502
41. Van den Burg, B., and Eijsink, V. G. (2002) *Curr Opin Biotechnol* **13**(4), 333-337
42. Cambillau, C., and Claverie, J. M. (2000) *J Biol Chem* **275**(42), 32383-32386
43. Kumar, S., and Nussinov, R. (2001) *Cell Mol Life Sci* **58**(9), 1216-1233
44. Vogt, G., Woell, S., and Argos, P. (1997) *J Mol Biol* **269**(4), 631-643
45. Spiller, B., Gershenson, A., Arnold, F. H., and Stevens, R. C. (1999) *Proc Natl Acad Sci U S A* **96**(22), 12305-12310
46. Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) *Nat Struct Biol* **7**(5), 380-383
47. Martin, A., Sieber, V., and Schmid, F. X. (2001) *J Mol Biol* **309**(3), 717-726
48. Sieber, V., Pluckthun, A., and Schmid, F. X. (1998) *Nat Biotechnol* **16**(10), 955-960
49. Turner, S. L., Ford, G. C., Mountain, A., and Moir, A. (1992) *Protein Eng* **5**(6), 535-541
50. Akanuma, S., Yamagishi, A., Tanaka, N., and Oshima, T. (1998) *Protein Sci* **7**(3), 698-705
51. Tamakoshi, M., Nakano, Y., Kakizawa, S., Yamagishi, A., and Oshima, T. (2001) *Extremophiles* **5**(1), 17-22
52. Fridjonsson, O., Watzlawick, H., and Mattes, R. (2002) *J Bacteriol* **184**(12), 3385-3391
53. Boyer, H. W., and Rouland-Dussoix, R. (1969) *J Mol Biol* **41**(3), 459-472
54. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Ed. Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
55. Shafikhani, S., Siegel, R. A., Ferrari, E., and Schellenberger, V. (1997) *Biotechniques* **23**(2), 304-310
56. Stemmer, W. P., Crameri, A., Ha, K. D., Brennan, T. M., and Heyneker, H. L. (1995) *Gene* **164**(1), 49-53
57. Bradford, M. M. (1976) *Anal Biochem* **72**, 248-254
58. Rondeau, J. M., Cagnon, C., Moras, D., and Masson, J. M. (1989) *J Mol Biol* **207**(3), 645-646
59. Otwinowski, A., and Minor, W. (1997) *Meth Enzymol* **276**, 307-326
60. Vagin, A. A., and Teplyakov, A. (1997) *J Appl Cryst* **30**, 1022-1025
61. Perrakis, A., Morris, R., and Lamzin, V. S. (1999) *Nat Struct Biol* **6**(5), 458-463
62. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr D Biol Crystallogr* **53**, 240-255
63. Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J Appl Cryst* **26**, 283-291
64. Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) *Acta Crystallogr D Biol Crystallogr* **55 (Pt 1)**, 191-205
65. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**(2714-23)

Chapter 3

Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Stan J.J.Brouns, Nicole Smits, Hao Wu, Ambrosius P.L.Snijders, Phillip C. Wright, Willem M. de Vos, and John van der Oost

J Bacteriol (2006) **188**: 2392-2399

ABSTRACT

Sulfolobus solfataricus is an aerobic crenarchaeon that thrives in acidic volcanic pools. In this study, we have purified and characterized a thermostable alpha-galactosidase from cell extracts of *S. solfataricus* P2 grown on the trisaccharide raffinose. The enzyme, designated GalS, is highly specific for alpha-linked galactosides, which are optimally hydrolyzed at pH 5 and 90^0C. The protein consists of 74.7-kDa subunits and has been identified as the gene product of open reading frame Sso3127. Its primary sequence is most related to plant enzymes of glycoside hydrolase family 36, which are involved in the synthesis and degradation of raffinose and stachyose. Both the galS gene from *S. solfataricus* P2 and an orthologous gene from *Sulfolobus tokodaii* have been cloned and functionally expressed in *Escherichia coli*, and their activity was confirmed. At present, these *Sulfolobus* enzymes not only constitute a distinct type of thermostable alpha-galactosidases within glycoside hydrolase clan D but also represent the first members from the *Archaea*.

INTRODUCTION

α-Galactosidases (α-Gals) (EC 3.2.1.22) are a widespread class of enzymes that liberate galactose from the nonreducing end of sugars. In bacteria, yeasts, and fungi, these enzymes are usually involved in the degradation of various plant saccharides, which can then serve as a carbon and energy source for growth. Plants synthesize α-galactosides such as raffinose and stachyose as the major energy storage molecules in leaves, roots, and tubers. Moreover, these oligosaccharides have been associated with cold and desiccation tolerance of seeds (28). Both raffinose and stachyose are produced by two specialized synthases (raffinose synthase [EC 2.4.1.82] and stachyose synthase [EC 2.4.1.67]) that use galactinol as a galactosyl donor (39). The oligosaccharides are degraded during seed germination by the action of two distinct types of α-Gals that differ in their optimal pH of catalysis. While the acid α-Gal type is most likely active in the acidic environment of the vacuole and the apoplasm, the alkaline α-Gal type probably catalyzes galactose release in the more neutral or alkaline cytoplasm (4, 17, 31). Mammals express an α-Gal and an α-N-acetylgalactosaminidase (α-NAGal) in lysosomal bodies to degrade glycolipids, glycoproteins, and oligosaccharides. In humans, mutations in the X-chromosomal α-Gal gene can lead to an accumulation of these α-linked galactosides in

tissues, which results in a recessive disorder called Fabry disease (12). Mutations in the related α-NAGal gene, which is located on chromosome 22, lead to either Schindler or Kanzaki disease.

Glycoside hydrolases (GHs) have been classified into families based on their primary sequence similarities (23). Except for some rare cases (34), the majority of α-Gals can be found in GH clan D (GH-D), which comprises families 27 and 36. The two families share a fairly conserved catalytic domain and hydrolyze the glycosidic bond with retention of configuration of the liberated D-galactose (3). Substantial insight into the molecular mechanisms of substrate recognition and catalysis of these enzymes was acquired when the crystal structures of the human, rice, and fungal α-Gals (16, 18, 20), as well as the chicken α-NAGal (19), were obtained.

In this study, we have purified an unusual intracellular α-Gal from the hyperthermophilic crenarchaeon *Sulfolobus solfataricus* P2, an aerobic microorganism that lives in terrestrial volcanic pools of high acidity (75 to 85°C, pH 2 to 4) (2).

MATERIALS AND METHODS

All chemicals were of analytical grade and purchased from Sigma, unless stated otherwise. Primers were obtained from MWG Biotech AG (Ebersberg, Germany). PCRs were performed with Pfu TURBO polymerase (Stratagene). The chromogenic substrate X-α-Gal (5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside) was purchased from Glycosynth (Warrington, United Kingdom).

Growth of *Sulfolobus* species. *S. solfataricus* P2 (DSM1617) and *S. tokodaii* (JCM10545) were grown aerobically at pH 3.5 in a rotary shaker at 80°C. The medium contained 2.5 g/liter $(NH_4)_2SO_4$, 3.1 g/liter KH_2PO_4, 203.3 mg/liter $MgCl_2 \cdot 6\ H_2O$, 70.8 mg/liter $Ca(NO_3)_2 \cdot 4\ H_2O$, 2 mg/liter $FeSO_4 \cdot 7\ H_2O$, 1.8 mg/liter $MnCl_2 \cdot 4\ H_2O$, 4.5 mg/liter $Na_2B_4O_7 \cdot 2\ H_2O$, 0.22 mg/liter $ZnSO_4 \cdot 7\ H_2O$, 0.06 mg/liter $CuCl_2 \cdot 2\ H_2O$, 0.03 mg/liter $Na_2MoO_4 \cdot 2\ H_2O$, 0.03 mg/liter $VOSO_4 \cdot 2\ H_2O$, and 0.01 mg/liter $CoCl_2 \cdot 6\ H_2O$ and was supplemented with 3 g/liter carbon source and Wollin vitamins. The Wollin vitamin stock (100×) contained 2 mg/liter D-biotin, 2 mg/liter folic acid, 10 mg/liter pyridoxine-HCl, 10 mg/liter riboflavin, 5 mg/liter thiamine-HCl, 5 mg/liter nicotinic acid, 5 mg/liter DL-Ca-pantothenate, 0.1 mg/liter vitamin B_{12}, 5 mg/liter *p*-aminobenzoic acid, 5 mg/liter lipoic acid.

Purification of the α-Gal from *S. solfataricus* P2 extracts. A 3.6-liter culture of *S. solfataricus* P2 was grown on raffinose to an A_{600} of 0.8, after which the cells were centrifuged (10 min, 6,000 × *g*, 4°C). The cell

pellet (8.8 g [wet weight]) was resuspended in 10 ml of 50 mM Tris-HCl (pH 7.5) and frozen at −80°C. After thawing, the cells were sonicated and diluted 1:1 in the same buffer containing 1 mM phenylmethylsulfonyl fluoride. The suspension was clarified by centrifugation (60 min, 19,000 × *g*, 4°C). All subsequent chromatographic steps were performed on an ÄKTA fast protein liquid chromatography system (Amersham Biosciences) at room temperature.

First, the supernatant was applied to a 70-ml Q-Sepharose Fast Flow (Amersham Biosciences) anion exchange column which was equilibrated with 50 mM Tris-HCl (pH 7.5). After extensive washing, the proteins were eluted by a linear gradient of buffer with 0.5 M NaCl. Active fractions were pooled and dialyzed overnight against 10 mM NaP_i buffer (pH 6.8). The dialyzed fraction was then loaded onto a hydroxyapatite column, CTH5-I (Bio-Rad), and eluted by a linear gradient of 500 mM NaP_i buffer (pH 6.8). Fractions that contained α-Gal activity were pooled, dialyzed overnight against 50 mM NaP_i buffer (pH 6.8), and applied to a MonoQ 5/50 GL column (Amersham Biosciences). The proteins were eluted by a linear gradient of 50 mM NaP_i buffer (pH 6.8) supplemented with 0.5 M NaCl. Subsequently, the pooled MonoQ fractions were purified by gel filtration chromatography using a Superdex 200 HR 10/30 column (Amersham Biosciences) and 50 mM NaP_i buffer (pH 6.8) supplemented with 100 mM NaCl.

Protein identification. The protocol to identify the α-Gal was slightly modified from that of Snijders et al. (42, 43). Coomassie-stained protein bands were excised with a scalpel from a sodium dodecyl sulfate (SDS)-polyacrylamide gel and transferred to low-adhesion Eppendorf tubes. Next, they were destained twice with 200 mM NH_4HCO_3 in 40% acetonitrile (ACN) for 30 min at 37°C. After this, the gel pieces were shrunk with ACN and dried in a vacuum centrifuge, and the proteins were reduced with 10 mM dithiothreitol (DTT) (30 min, 56°C) and alkylated with 50 μl of 55 mM iodoacetamide in 50 mM NH_4HCO_3 (20 min at room temperature in the dark). After this, gel pieces were washed with 50 mM NH_4HCO_3 and shrunk with ACN. Next, the trypsin solution was added, consisting of 8 μg of trypsin in 100 μl of 9% ACN and 50 mM NH_4HCO_3 (overnight at 37°C). The next day, peptides were extracted in four sequential extraction steps: (i) 50 μl of 50 mM NH_4HCO_3 (10 min at room temperature), (ii) 75 μl of ACN (15 min at 37°C), (iii) 75 μl of 5% formic acid (FA), and (iv) 75 μl of ACN (15 min at 37°C). Extracts were pooled and dried down to completeness in a vacuum centrifuge, and peptides were redissolved in 0.1% formic acid and 3% ACN. The peptide mixture was separated on a PepMap C_{18} RP capillary column (LC Packings, Amsterdam, The Netherlands) and eluted directly onto a QStarXL electrospray ionization-quadrupole time-of-flight (ESI qQ-TOF) tandem mass spectrometer (Applied Biosystems/MDS Sciex). Gradients and data acquisition were set up as previously

described (42,43). A database search was performed with Mascot 2.0 (www.matrixscience.com) against the Mass Spectrometry Protein Sequence DataBase. The peptide tolerance was set to 1.0 Da, and the tandem mass spectrometry tolerance was 0.6 Da. Carbamidomethyl modification of cysteine was set as a fixed modification, and methionine oxidation was set as a variable modification. A maximum of one missed cleavage site by trypsin was allowed.

Protein quantitation. Protein concentrations were determined by using the bicinchoninic acid protein assay (Pierce) and the Bradford assay (Bio-Rad) according to the supplied protocol.

Enzyme assays. Standard activity assays were performed with 1.0 mM pNPG (*para*-nitrophenol-α-D-galactopyranoside) in a 50 mM citric acid-Na_2HPO_4 buffer (pH 5.0) at 80°C. The reaction was stopped by adding 0.7 volumes of cold 1 M Na_2CO_3, after which the sample was put on ice. The amount of released *para*-nitrophenol (pNP) was measured at 420 nm and calculated with an extinction coefficient of 0.0135 μM^{-1} cm^{-1} (33). One unit of activity was defined as the amount of enzyme required to convert 1 µmol of pNP per minute. Thermal inactivation assays were performed by incubating 45 µg/ml of enzyme at 70, 80, and 90°C and drawing aliquots at regular intervals during 2.5 h. The residual activity was then determined by the standard assay. The optimal pH of the enzyme was determined with the McIlvaine citrate-phosphate buffer system (36) in a pH range of 2.2 to 8.0. Several pNP-substituted hexoses and pentoses were tested as a substrate at a concentration of 20 mM under standard conditions.

The activity toward the disaccharide melibiose [D-galactose-α(1,6)-D-glucose], the trisaccharide raffinose [D-galactose-α(1,6)-D-glucose-(α1,β2)-D-fructose], and the tetrasaccharide stachyose [D-galactose-α(1,6)-raffinose] (Fluka) was tested in a continuous assay at 80°C using 20 mM of substrate and 0.4 mM $NADP^+$ in a 50 mM NaP_i buffer (pH 7.0). The amount of liberated D-galactose was measured by adding 20 U of *Thermoplasma acidophilum* glucose dehydrogenase (Sigma) and calculated with an extinction coefficient of 6.22 mM^{-1} cm^{-1} for NADPH (33).

The effects of different divalent metal ions, EDTA, and DTT were tested at a concentration of 10 mM in a standard assay. Sugars such as D-galactose, sucrose, L-arabinose, and D-fucose were analyzed for inhibitory effects at concentrations of 5, 10, and 25 mM.

Gene cloning and mutagenesis. The genomic fragments corresponding to Sso3127 and St2554 were PCR amplified from genomic DNA that was prepared according to the method of Pitcher et al. (40), and the amplified genes were cloned into vector pET24d or pMAL-CT, respectively, using *Escherichia coli* HB101 as

TABLE 1. Strains, plasmids, and primers used in this study

Strains, plasmid, or primer	Entry code / genotype/description, or sequence(5' - 3')[a]	Reference, source, or restriction site/codon change
S. solfataricus P2	DSM1617 wild-type	46
S. tokodaii strain 7	JCM 10545 wild-type	44
E. coli HB101	*F- hsdS*20 (r_B-, m_B-) *ara*-14 *galK*2 *lacY*1 *leuB*6 *mcrB mtl*-1 *proA*2 *recA*13 *rpsL*20 *supE*44 *thi*-1 *xyl*-5 (Str^R)	1
E. coli BL21(DE3)-RIL	*hsdS gal* (λ*cIts*857 *ind*1Sam7 *nin*5 *lac*UV5-T7 gene *1*)	Novagen
Plasmids		
pET24d	T7 RNA polymerase expression system	Novagen
pMAL-CT	N-terminal fusion-tag of *E. coli* MBP	New England Biolabs
pWUR269	*galS* (Sso3127) in pET24d (BspHI/NcoI – XhoI)	This study
pWUR270	*galSt* (St2554) in pET24d (NcoI – BamHI)	This study
pWUR271	*galS* (Sso3127) in pMAL-CT (BamHI – PstI)	This study
pWUR272	pWUR271 with substitution D367G	This study
pWUR273	pWUR271 with substitution D425G	This study
Primers		
galS-fw (Sso3127)	CGTGATCATGATTTGGATAGAAGACGAGAATGGG	BspHI
galS-rv (Sso3127)	GCTCACTCGAGTCATTCTATAGTAACTGGGATTCC	XhoI
galSt-fw (St2554)	GGGCGCCATGGCTATTTGGATATATGATGAAAATGGG	NcoI
galSt-rv (St2554)	GCCCGGGATCCTTACTCGATACTAACTATTTCTTCAGC	BamHI
MBP-*galS*-fw (Sso3127)	CGCGGATCCATGATTTGGATAGAAGACGAG	BamHI
MBP-*galS*-rv (Sso3127)	CCGCGCTGCAGTCATTCTATAGTAACTGGGATTCC	PstI
galS-fw D367G[a]	GTAATCAATGGGTAATTCACGC	GAT → GGT
galS-rv D367G[a]	CAACCTTAACGAGATCGAAG	GAT → GGT
galS-fw D425G[b]	GAGGAATTCTATAGGCTACGTACCCTTC	GAC → GGC
galS-rv D425G[b]	GAAGGGTACGTAGCCTATAGAATTCCTC	GAC → GGC

[a]Restriction sites are underlined, and nucleotide mismatches are indicated in boldface.
[b]Excite mutagenesis primer (Stratagene).
[c]QuikChange mutagenesis primer (Stratagene).

a host (Table 1). Mutations were introduced by employing either the Excite or QuikChange protocols (Stratagene). Inserts of plasmids used in this study were sequenced by Westburg Genomics (Leusden, The Netherlands).

Recombinant protein overexpression and purification. Plasmids pWUR269 and pWUR270, containing the *galS* gene from *S. solfataricus* P2 and the *galSt* gene from *S. tokodaii*, respectively, were transformed into

E. coli BL21(DE3)-RIL. Transformants were grown overnight at 37°C and used to inoculate 1-liter cultures that were incubated to an A_{600} of approximately 0.8. The cultures were then placed on ice to induce cold shock proteins which may prevent inclusion body formation (7). After 1 h, recombinant protein expression was induced by adding 0.5 mM IPTG (isopropyl-β-D-thiogalactopyranoside). The cultures were allowed to grow for 4 h at room temperature, after which the cells were centrifuged.

Plasmids pWUR271, pWUR272, and pWUR273, containing wild-type and mutant α-Gal genes translationally fused to the *E. coli* maltose binding protein (MBP), were transformed into *E. coli* HB101, and the protein was overexpressed as described above. Frozen cell pellets were resuspended in 20 mM Tris-HCl (pH 7.5) buffer which was supplemented with 150 mM NaCl, 10% (vol/vol) glycerol, and complete protease inhibitors (Roche). Cell suspensions were then sonified (Branson sonifier) while the sample was kept in an ice-ethanol bath at −10°C. Next, insoluble cell matter was removed by centrifugation (60 min, 26,500 × *g*, 4°C). Clarified cell extracts were then loaded onto equilibrated amylose resin (New England Biolabs). After extensive washing with the same buffer without glycerol and protease inhibitors, the fusion protein was eluted in buffer supplemented with 10 mM maltose. To concentrate the fusion protein and to remove contaminating maltose, the sample was diluted five times in NaCl-free buffer and loaded onto a MonoQ 5/50 GL column (Amersham Bioscience). Pure fusion proteins were eluted by a linear gradient of the same buffer with 1 M NaCl. MBP was removed by thrombin cleavage according to the instructions from the manufacturer (Sigma), after which the recombinant GalS was reisolated by MonoQ as described above.

Sequence analysis of α-Gals. A total of 171 α-Gal sequences belonging to PF0265 and PF05691 were obtained from the Pfam database. A small subset of sequences, including the ones for which a crystal structure is available and selected members of the different subgroups of GH-D, was then aligned using the TCoffee program (38). Additional sequences were subsequently added to the alignment using the ClustalX profile alignment mode. Extensive manual alignment editing was performed with the BioEdit software package, after which highly similar sequences were discarded. Neighbor-joining trees of 43 representative sequences were then calculated and bootstrapped in ClustalX, while correcting for multiple substitutions. Trees were drawn with the program Treeview. Sequence motifs comprising the catalytic nucleophile and acid/base were made using the Weblogo server (8).

RESULTS

Purification and identification of the α-Gal from *S. solfataricus*. *S. solfataricus* P2 is capable of using α-linked galactosides such as melibiose and raffinose as a sole carbon and energy source. The capability to do so requires the presence not only of a suitable sugar uptake system but also of an intracellular α-Gal and an efficient metabolic pathway for step-wise oxidation of the released monosaccharides. Activity assays using extracts of *S. solfataricus* grown on several sugars, as well as tryptone, confirmed the presence of a constitutively expressed enzyme which is capable of hydrolyzing pNPG. The α-Gal was purified from 8.8 g of cells, which were harvested from a 3.6-liter late-exponential-phase

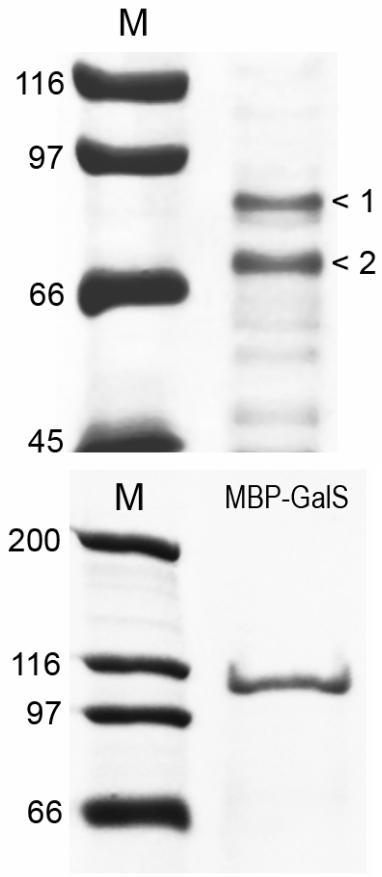


FIG. 1. (Top) Purification of the native α-Gal from *S. solfataricus*. A digital photograph of an 8% SDS-PAGE gel loaded with 5 µg total protein of the α-Gal enriched fraction after gel filtration chromatography is shown. The gel was stained with Coomassie brilliant blue G250. The sizes of the broad-range protein marker are shown in kilodaltons. Arrows 1 and 2 indicate the two protein bands that were identified by mass spectrometry.

(Bottom) Digital photograph of a 6% SDS-PAGE gel loaded with 2 µg GalS fusion protein after affinity and anion exchange chromatography.

culture grown on raffinose. In four subsequent chromatographic steps (Q-Sepharose, hydroxyapatite, MonoQ, and Superdex), the specific activity was enriched 23-fold (Table 2). The enriched fractions were found to contain two major protein bands, of approximately 73 and 87 kDa, by SDS-polyacrylamide gel electrophoresis (PAGE) (Fig. 1,

top). Size exclusion chromatography showed that the α-Gal eluted in fractions corresponding to proteins of a native molecular mass of 225 ± 15 kDa, suggesting a trimeric oligomerization (data not shown). Separations of the α-Gal using columns based on hydrophobic interactions resulted in an almost complete loss of activity. Similarly, affinity purification using immobilized D-galactose (Pierce) proved unsuccessful, since the enzyme did not bind to the resin. Therefore, the protein could not be purified to apparent homogeneity. To identify their corresponding genes, both major protein bands were excised and treated with trypsin, after which the peptides were eluted from the gel slice, desalted, concentrated, and analyzed by ESI qQ-TOF tandem mass spectrometry. Twenty-one different peptides (40% sequence coverage, MOWSE score of 831) were found that matched band 1 with the gene product of Sso2760 (*cutA*-5), an 84.9-kDa protein which is orthologous to the 80.5-kDa α-subunit of the aldehyde oxidoreductase of *S. acidocaldarius* (26). In band 2, 17 unique peptides were found that matched the gene product of Sso3127 (Uniprot, Q97U94), a 74.6-kDa hypothetical protein that shares sequence similarity with raffinose synthases and seed imbibition proteins of GH36. The MOWSE score for this protein was 517, and a sequence coverage of 34% was achieved. This protein was deemed likely to be responsible for the thermostable α-Gal activity and was termed GalS.

Characteristics of native GalS. (i) Substrate specificity. The enzyme was able to hydrolyze pNPG but showed no activity toward pNP-α-substituted hexoses such as D-glucose, D-mannose, L-rhamnose, or N-acetyl-D-galactosaminide. pNP-β-substituted hexoses such as D-galactose, D-glucose, and D-mannose did not support catalysis, nor did the pentose pNP-β-D-xylopyranoside. Of the tested artificial substrates, only pNP-β-L-arabinopyranoside was hydrolyzed by the enzyme. However, the Michaelis-Menten constant for this substrate was very high (K_m, 37.4 ± 2.2 mM) compared to that of pNPG (K_m, 0.08 ± 0.01 mM). Of the natural substrates tested, the α-Gal was most active on the trisaccharide raffinose (2.9 U/mg), followed by the disaccharide melibiose (2.6 U/mg), and the tetrasaccharide stachyose (1.2 U/mg).

(ii) Catalytic and stability properties. Activity assays in a range of pH values indicated that the enzyme is more than 50% active between pH 4.1 and 6.7 and that it has a sharp optimum at pH 5.0 (Fig. 2, top). When the assay temperature was varied, optimal α-Gal

TABLE 2. Purification of the native *S. solfataricus* α-Gal

Step	Protein (mg)	Activity (U)	Specific activity (U mg^{-1})	Purification factor	Recovery (%)
Cell extract	174	31.5	0.18	1	100
Q-sepharose	49	19.5	0.4	2	62
Hydroxyapatite	7	7.8	1.16	6	25
MonoQ	3	6.4	2.31	13	20
Superdex	0.26	1.1	4.25	23	4

activity was found at 90°C, while the enzyme was 50% active at a more physiological temperature of 75°C (Fig. 2, middle). GalS was completely inactive at temperatures below 50°C. At its optimal temperature for catalysis, the enzyme showed a half-life of 30 min (Fig. 2, bottom). At 70 or 80°C, the enzyme did not show any significant decrease in activity during 2.5 h of incubation.

(iii) Inhibition of activity. The α-Gal was tested for inactivation by divalent cations at a concentration of 10 mM. Most metal ions completely inhibited the activity of the enzyme (Ag^{2+}, Ca^{2+}, Cd^{2+}, Co^{2+}, Cu^{2+}, Hg^{2+}, Mn^{2+}, Ni^{2+}, and Zn^{2+}), but Mg^{2+} and Mo^{2+} had no effect. Assays in the presence of the divalent cation chelator EDTA or the reducing agent DTT did not alter the activity. Enzyme activity was also unaffected in the presence of several saccharides, such as D-galactose, L-arabinose, D-fucose, and sucrose, up to concentrations of 20 mM.

Recombinant GalS overexpression. The *galS* gene (Sso3127; Uniprot, Q97U94) was cloned in a T7 RNA polymerase-based vector and overexpressed *E. coli* BL21(DE3)-RIL. The heat-treated soluble fraction confirmed the presence of α-Gal activity at 80°C in the *galS* extract. Activity of *E. coli* transformants was also observed on selective plates containing 20 μg/ml of the chromogenic substrate X-α-Gal after a short incubation at 60°C. However, the soluble expression levels were too low to conduct any detailed experimental studies. Therefore, an orthologous gene from *S. tokodaii* (*galSt*, St2554, Q96XG2) was cloned and expressed, yielding an enzyme with characteristics and expression levels similar to the *S. solfataricus* α-Gal. Several reported strategies were then employed to obtain a sufficient amount of soluble recombinant protein. These

included monitoring the protein overexpression levels over time (1, 2, 5, and 8 h and overnight induction); lowering the temperature during overnight protein overexpression (15, 20, and 25°C); decreasing the IPTG concentration (0.01, 0.05, and 0.1 mM); and changing the cell lysis buffer, the pH, and the ionic strength (HEPES-KOH, Tris-HCl, NaP_i, pH 5 to 10; ionic strength, 0 to 0.5 M NaCl), but none of these strategies prevented the formation of inclusion bodies. However, since it is known that the *E. coli* MBP promotes the solubility of proteins to which it is fused (25), this strategy was also tested. The MBP was translationally fused to the N terminus of GalS, giving rise to a 116-kDa fusion protein. Overexpression in *E. coli* HB101 now yielded much more soluble fusion protein, which could easily be isolated by amylose affinity purification and subsequent anion exchange chromatography (Fig. 1, bottom).

The catalytic properties of the *E. coli*-produced GalS were comparable to those determined for the native enzyme. The Michaelis-Menten constants K_m and V_{max} for hydrolysis of pNPG were 0.085 ± 0.009 mM and 48.3 ± 1.1 U · mg^{-1}, respectively. The catalytic efficiency of the *E. coli*-produced GalS was 703.6 ± 15.7 $s^{-1} \cdot mM^{-1}$.

Phylogeny of α-Gals and prediction of their catalytic amino acids. GHs and transglycosidases are classified into families based on their sequence similarities (23). Family 27, which mainly contains α-Gals (EC 3.2.1.22) and α-N-acetylgalactosaminidases (EC 3.2.1.49), is related to family 36, which additionally consists of stachyose synthases (EC 2.4.1.67) and raffinose synthases (EC 2.4.1.82). Both families make up GH-D, and their catalytic domain adopts a common $(\beta\alpha)_8$-barrel fold (23). This prediction was recently confirmed by the elucidation of four structures of members of GH27 (16, 18-20). In order to analyze the diversity of both families and to position the *Sulfolobus* α-Gal sequences in a phylogenetic tree, we aligned the catalytic domains of 43 representative members of GH-D (for the alignment, see Fig. S1 in the supplemental material). A phylogenetic tree, which is shown in Fig. 3 (top), was constructed from this alignment. The tree indicates that α-Gals of clan GH-D can be divided into at least three major types: the eukaryal type (GH27), the bacterial type (GH36b), and a type consisting of mainly plant enzymes (GH36p). The latter type additionally comprises uncharacterized sequences from the *Archaea*, intestinal bacteria, and a fungus. The two *Sulfolobus* α-Gal sequences form the deepest branch of this plant subfamily and, as yet, constitute the only

archaeal sequences of clan GH-D. Moreover, the *Sulfolobus* sequences are clearly distinct from other thermostable α-Gals that are produced by thermophilic bacteria, such as those from the genera *Thermotoga* (Uniprot, O33835) (33), *Thermus* (Q746I3) (15), and *Geobacillus* (Q9LBD1) (14).

The catalytic domain of GH-D members comprises only 241 of the 648 amino acids of the total GalS sequence (amino acids 225 to 466). It is encompassed by a 26-kDa N-terminal domain and a 21-kDa C-terminal domain of unknown functions. These extra domains are partly shared by members of GH36p (PF05691, raffinose synthase domain) but not by GH27 or GH36b members (PF02065, melibiase domain).

Figure 3B depicts the invariant aspartic acid residues that are predicted to be involved in catalysis as either the nucleophile or the acid/base (A/B) of the double displacement reaction mechanism of α-galactosyl hydrolysis. The prediction is based on the alignment combined with experimental and structural evidence for the nucleophile and A/B, which are available for GH27 (16, 18-20, 22, 35). While the sequence motif for the catalytic nucleophile is fully conserved within GH-D (K[Y/V/L/W]**D** [catalytic Asp in boldface]), the motif comprising the A/B aspartic acid is much less conserved (RXXX**D**) and has therefore been missed in previous analyses. The mesophilic bacterial α-Gal group seems to deviate most from the main A/B consensus but appears to contain another motif (**D**XXD) which is mutually shared with GH27. Yet another motif that is strongly conserved among α-Gals comprises two invariant aspartic acids that are involved in substrate binding of the C-4 and C-6 OH groups of D-galactose (16, 20). These residues have long been assumed to be involved in catalysis as the A/B residues. Interestingly, members of GH27 and GH36 can be distinguished on the bases of either a DDC or a DDG motif at this position, respectively. The cysteine of GH27 is fully conserved, since it is involved in the formation of a disulfide bond within the active site (see Fig. S1 in the supplemental material).

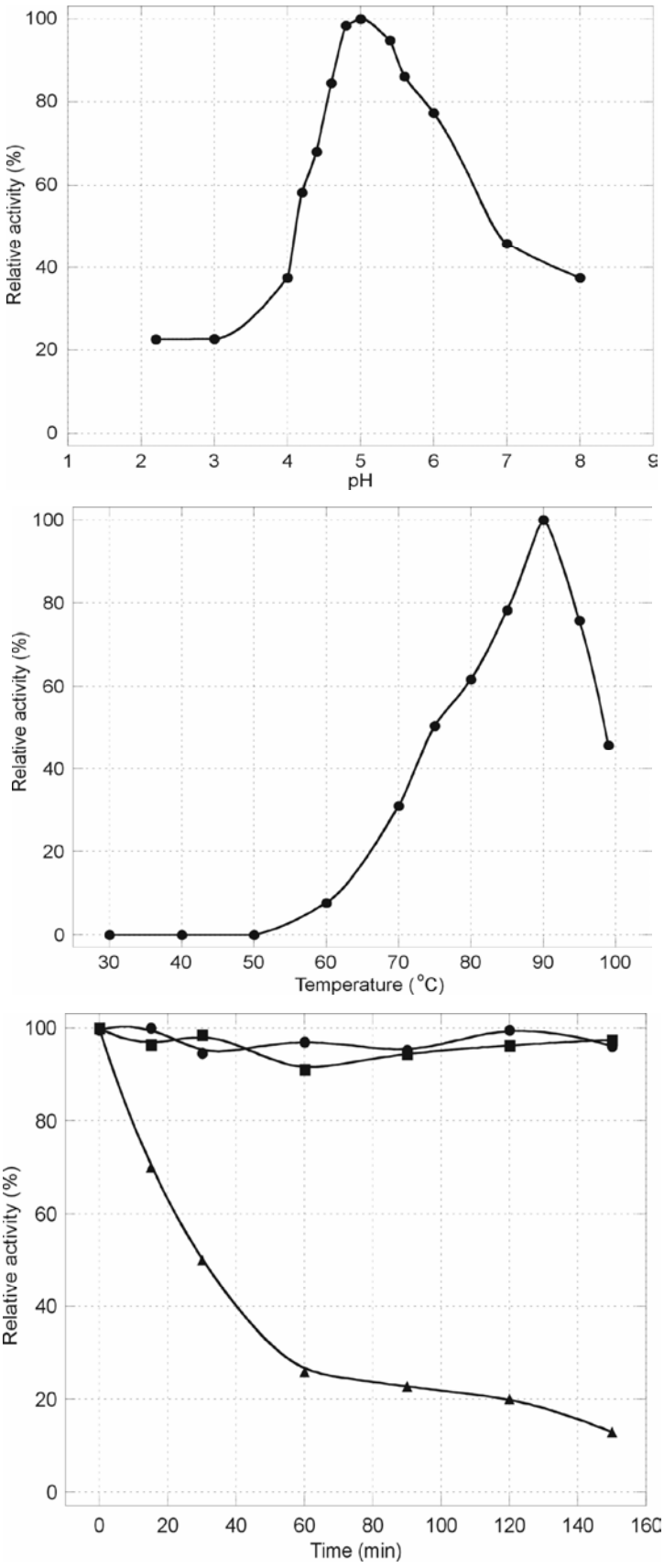


FIG. 2. Properties of the native GalS. Relative reaction rates as a function of pH (top) and temperature (middle).

(Bottom)Thermal inactivation curves. Residual activity is given after preincubation at 70°C (•), 80°C (■), and 90°C (▲).

Verification of the predicted catalytic amino acids. The predicted nucleophilic and A/B aspartic acids of GalS were changed into glycines, yielding mutants D367G and D425G, respectively. The wild-type and both mutant enzymes were produced and purified. Activity assays using various substrate concentrations indicated that the activity of mutant D367G was below the detection limit (<1 × 10^{-3} times that of the wild type), whereas mutant D425G showed approximately 5 × 10^{-3} times the activity of the wild type (data not shown). Molar concentrations of sodium azide or sodium formate were unable to restore or increase the activity of the mutant enzymes (data not shown).

DISCUSSION

In the present study we have identified a novel thermostable α-Gal in cell extracts of the hyperthermophilic crenarchaeon *S. solfataricus* and studied the biochemical properties of the enzyme. Sequence analysis revealed an unusual phylogenetic position within the widely distributed class of GH-D α-Gals.

Properties of GalS. GalS and GalSt belong to the most thermoactive α-Gals known to date. Despite the fact that their primary sequence is very different from the α-Gals of thermophilic bacteria, their catalytic properties, such as the optimal pH of 5 and the optimal temperature of catalysis of 90°C, are similar. Interestingly, most related enzymes belonging to GH36p have neutral or slightly alkaline pH optima. This might be due to the fact that these enzymes contain an aspartic acid residue that is in juxtaposition to the predicted A/B residue, which may cause an increase in the pK_a of the A/B. Catalysis can proceed only when the correct protonation state of the A/B is achieved, which may thus be at slightly higher pH. Since GalS contains an isoleucine residue at the corresponding position, its optimum activity may therefore be at a slightly acidic pH. GalS is specific for α-linked D-galactosides and β-linked L-arabinosides. However, the Michaelis-Menten constant K_m, which is roughly the inverse measure of the enzyme's substrate affinity, is approximately 460 times lower for pNPG than for pNP-β-L-arabinoside, which probably excludes a physiological role of GalS as an arabinosidase. The activity on β-linked L-arabinoside is not surprising since it has identical configurations of the hydroxyl groups at C-1, C-2, C-3, and C-4 but lacks the CH_2OH at its C-5 atom in comparison with galactose (11). Activity assays using the natural substrates melibiose, raffinose, and stachyose

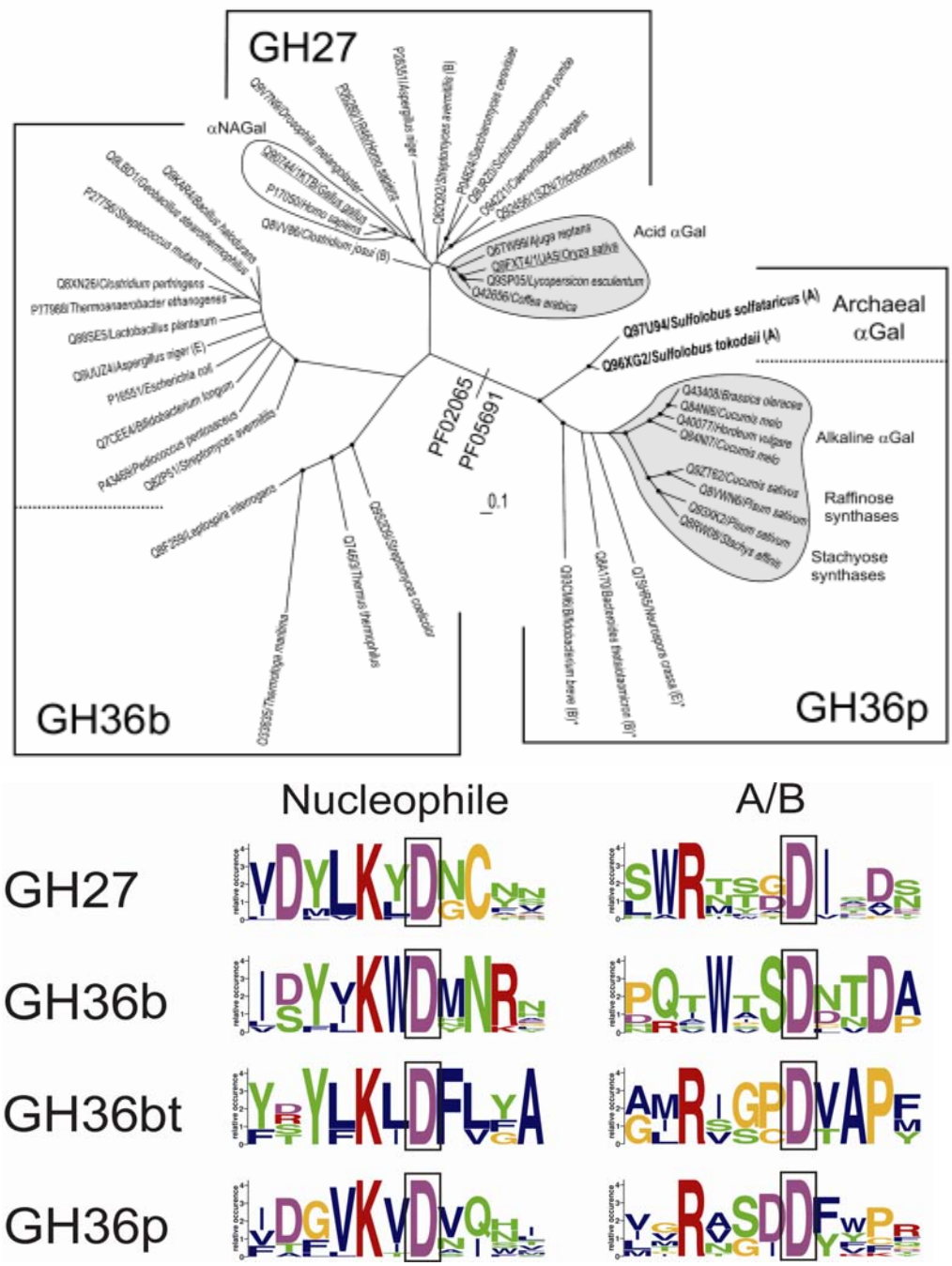


FIG. 3. (Top) Unrooted neighbor-joining tree of the catalytic domain of a representative set of α-Gals belonging to GH-D. Nodes with bootstrap probabilities of >70% are indicated by dots. The scale bar indicates an evolutionary distance of 0.1 substitutions per position. Gray fields indicate sequences from a plant origin. Abbreviations: E, eukaryal sequences; B, bacterial sequences; A, archaeal sequences. Asterisks indicate uncharacterized sequences. (Bottom) Weblogo representations of the predicted catalytic amino acid sequence motifs of the different classes of GH-D. Character sizes indicate the relative occurrences of amino acids in the sequences encompassing the catalytic aspartic acids (boxed). The thermophilic bacterial sequences from the *Thermus* and *Thermotoga* subgroup are indicated by GH36bt; the other groups are defined as depicted in panel A.

indicated that the enzyme can release D-galactose from oligosaccharides up to at least four sugar moieties, with the maximum catalytic rates on the trisaccharide. Several related alkaline α-Gals were reported to prefer the longer raffinose-type oligosaccharides (4, 17, 31). The enzyme showed no product inhibition, but it was severely inhibited by several divalent metal ions. This strong inhibitory effect of transition metal ions, such as Zn^{2+}, Mn^{2+}, and Hg^{2+}, was also observed with the raffinose synthase from *Vicia faba* seeds (32). Similarly, the raffinose synthase from *Pisum sativum* (Uniprot, Q8VWN6) was only 30.2% active after 1 h of incubation with 1 mM $NiCl_2$ (39). This finding suggests the presence of an oxidizable group, such as a free cysteine, that is required for substrate binding or catalysis.

Mechanism of catalysis. Based on sequence similarity with GH27, GH36 members are assumed to hydrolyze the glycosidic bond with retention of configuration of the released D-galactose (3). Although this assumption has never been experimentally verified, it is highly likely that GH36 enzymes retain instead of invert enzymes, since only the covalent enzyme galactose intermediate of the retaining mechanism will allow transglycosylation reactions to occur. The raffinose and stachyose synthases of GH36p make use of this principle by coupling the galactose moiety of galactinol to an incoming sucrose or raffinose molecule in the second stage of the reaction, respectively (39).

Mutagenesis of the proposed nucleophilic and A/B aspartic acids of GalS gave rise to similar protein yields, indicating properly folded proteins but virtually inactive enzymes. Although many mutations may have a near lethal effect on enzymatic activity, this finding, combined with sequence alignment data, suggests that both substituted aspartic acids are involved in α-galactosyl hydrolysis. Unfortunately, reactivation experiments with external nucleophiles such as azide and formate failed to restore the activity of either the nucleophile or the A/B mutant enzymes. While β-glycosidases can sometimes be reactivated up to wild-type levels (37), for unknown reasons this strategy seems to be far less efficient for retaining α-glycosidases (6, 9).

Catalytic residues of glycosidases can also be identified by an approach that consists of covalent modification of either the nucleophile or A/B, followed by mass spectrometric analysis of tryptic digests of the modified enzyme. This method requires specifically designed enzyme inhibitor molecules that create a stable bond between the

enzyme and inhibitor. Using this strategy, the nucleophilic aspartic acid of the GH27 α-Gals from *Phanerochaete chrysosporium* and *Coffea arabica* (Uniprot, Q42656) was identified (22, 35).

Metabolism of α-linked galactosides. The ability to use α-linked galactosides such as melibiose and raffinose as a sole carbon and energy source implies that an organism should at least have a suitable sugar uptake system, an α-Gal, and an efficient catabolic pathway for the liberated monosaccharides. Within the *Sulfolobus* genus, the α-Gal described here was found only in the complete genome sequences of *S. solfataricus* (41) and *S. tokodaii* (27), not in *S. acidocaldarius* (5). This observation correlates well with the ability of *S. solfataricus* and *S. tokodaii*, but not *S. acidocaldarius*, to grow on melibiose and raffinose and also the monosaccharide D-galactose (21, 44). In addition, cellular growth on raffinose and D-galactose was also observed with *S. shibatae*, *S. yangmingensis*, and *S. tengchongensis*, which may imply that these species have an α-Gal gene in their genomes as well (21, 24, 45).

To date, two primary catabolic pathways for D-galactose have been reported. Most ubiquitous is the Leloir pathway, in which galactose is converted to glucose-1-phosphate by the action of a sugar kinase, uridylyltransferase, and an epimerase (13). Galactose can also be degraded according to a scheme that was originally proposed by De Ley and Doudoroff (10). In this pathway, D-galactose is oxidized to D-galactonate, subsequently dehydrated to 2-keto-3-deoxy-D-galactonate, phosphorylated at the C-6 position, and cleaved by an aldolase to yield pyruvate and D-glyceraldehyde-3-phosphate. *S. solfataricus* was recently demonstrated to use this pathway by employing a promiscuous enzyme set that can convert both D-galactose and D-glucose, as well as their derivatives (29, 30). Since the aldolase can convert nonphosphorylated substrates, the pathway may also occur nonphosphorylatively *in vivo*.

ACKNOWLEDGMENTS

This work was supported by a grant from the European Union in the framework of the SCREEN project (contract QLK3-CT-2000-00649). A.P.L.S. thanks the University of Sheffield and the United Kingdom's Engineering and Physical Sciences Research Council (EPSRC) for a scholarship. P.C.W. thanks the EPSRC for provision of an Advanced Research Fellowship (GR/A11311/01).

REFERENCES

1. **Boyer, H. W., and R. Rouland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Biol. **41:**459-472.

2. **Brock, T. D., K. M. Brock, R. T. Belly, and R. L. Weiss.** 1972. *Sulfolobus*: a new genus of sulfur-oxidizing bacteria living at low pH and high temperature. Arch. Mikrobiol. **84:**54-68.

3. **Brumer, H., III, P. F. Sims, and M. L. Sinnott.** 1999. Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main alpha-galactosidase. Biochem. J. **339:**43-53.

4. **Carmi, N., G. Zhang, M. Petreikov, Z. Gao, Y. Eyal, D. Granot, and A. A. Schaffer.** 2003. Cloning and functional expression of alkaline alpha-galactosidase from melon fruit: similarity to plant SIP proteins uncovers a novel family of plant glycosyl hydrolases. Plant J. **33:**97-106.

5. **Chen, L., K. Brugger, M. Skovgaard, P. Redder, Q. She, E. Torarinsson, B. Greve, M. Awayez, A. Zibat, H. P. Klenk, and R. A. Garrett.** 2005. The genome of *Sulfolobus acidocaldarius*, a model organism of the *Crenarchaeota*. J. Bacteriol. **187:**4992-4999.

6. **Cobucci-Ponzano, B., A. Trincone, A. Giordano, M. Rossi, and M. Moracci.** 2003. Identification of the catalytic nucleophile of the family 29 alpha-l-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. Biochemistry **42:**9525-9531.

7. **Constantinesco, F., P. Forterre, and C. Elie.** 2002. NurA, a novel 5′-3′ nuclease gene linked to rad50 and mre11 homologs of thermophilic Archaea. EMBO Rep. **3:**537-542.

8. **Crooks, G. E., G. Hon, J. M. Chandonia, and S. E. Brenner.** 2004. WebLogo: a sequence logo generator. Genome Res. **14:**1188-1190.

9. **Debeche, T., C. Bliard, P. Debeire, and M. J. O'Donohue.** 2002. Probing the catalytically essential residues of the alpha-l-arabinofuranosidase from *Thermobacillus xylanilyticus*. Protein Eng. **15:**21-28.

10. **De Ley, J., and M. Doudoroff.** 1957. The metabolism of d-galactose in *Pseudomonas saccharophila*. J. Biol. Chem. **227:**745-757.

11. **Dey, P. M., and J. B. Pridham.** 1972. Biochemistry of alpha-galactosidases. Adv. Enzymol. Relat. Areas Mol. Biol. **36:**91-130.

12. **Fabry, J.** 1898. Ein Beitrag Zur Kenntnis der Purpura haemorrhagica nodularis (Purpura papulosa hemorrhagica Hebrae). Arch. Dermatol. Syphilol. **43:**187-200.

13. **Frey, P. A.** 1996. The Leloir pathway: a mechanistic imperative for three enzymes to change the stereochemical configuration of a single carbon in galactose. FASEB J. **10:**461-470.

14. **Fridjonsson, O., H. Watzlawick, A. Gehweiler, and R. Mattes.** 1999. Thermostable alpha-galactosidase from *Bacillus stearothermophilus* NUB3621: cloning, sequencing and characterization. FEMS Microbiol. Lett. **176:**147-153. 15. **Fridjonsson, O., H. Watzlawick, A. Gehweiler, T. Rohrhirsch, and R. Mattes.** 1999. Cloning of the gene encoding a novel thermostable alpha-galactosidase from *Thermus brockianus* ITI360. Appl. Environ. Microbiol. **65:**3955-3963.

16. **Fujimoto, Z., S. Kaneko, M. Momma, H. Kobayashi, and H. Mizuno.** 2003. Crystal structure of rice alpha-galactosidase complexed with d-galactose. J. Biol. Chem. **278:**20313-20318.

17. **Gao, Z., and A. A. Schaffer.** 1999. A novel alkaline alpha-galactosidase from melon fruit with a substrate preference for raffinose. Plant Physiol. **119:**979-988.

18. **Garman, S. C., and D. N. Garboczi.** 2004. The molecular defect leading to Fabry disease: structure of human alpha-galactosidase. J. Mol. Biol. **337:**319-335.

19. **Garman, S. C., L. Hannick, A. Zhu, and D. N. Garboczi.** 2002. The 1.9 A structure of alpha-*N*-acetylgalactosaminidase: molecular basis of glycosidase deficiency diseases. Structure **10:**425-434.

20. **Golubev, A. M., R. A. Nagem, J. R. Brandao Neto, K. N. Neustroev, E. V. Eneyskaya, A. A. Kulminskaya, K. A. Shabalin, A. N. Savel'ev, and I. Polikarpov.** 2004. Crystal structure of alpha-galactosidase from *Trichoderma reesei* and its complex with galactose: implications for catalytic mechanism. J. Mol. Biol. **339:**413-422.

21. **Grogan, D. W.** 1989. Phenotypic characterization of the archaebacterial genus *Sulfolobus*: comparison of five wild-type strains. J. Bacteriol. **171:**6710-6719.

22. **Hart, D. O., S. He, C. J. Chany II, S. G. Withers, P. F. Sims, M. L. Sinnott, and H. Brumer III.** 2000. Identification of Asp-130 as the catalytic nucleophile in the main alpha-galactosidase from *Phanerochaete chrysosporium*, a family 27 glycosyl hydrolase. Biochemistry **39:**9826-9836.

23. **Henrissat, B.** 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:**309-316.

24. **Jan, R. L., J. Wu, S. M. Chaw, C. W. Tsai, and S. D. Tsen.** 1999. A novel species of thermoacidophilic archaeon, *Sulfolobus yangmingensis* sp. nov. Int. J. Syst. Bacteriol. **49:**1809-1816.

25. **Kapust, R. B., and D. S. Waugh.** 1999. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein Sci. **8:**1668-1674.

26. **Kardinahl, S., C. L. Schmidt, T. Hansen, S. Anemuller, A. Petersen, and G. Schafer.** 1999. The strict molybdate-dependence of glucose-degradation by the thermoacidophile *Sulfolobus acidocaldarius* reveals the first crenarchaeotic molybdenum containing enzyme—an aldehyde oxidoreductase. Eur. J. Biochem. **260:**540-548.

27. **Kawarabayasi, Y., Y. Hino, H. Horikawa, K. Jin-No, M. Takahashi, M. Sekine, S. Baba, A. Ankai, H. Kosugi, A. Hosoyama, S. Fukui, Y. Nagai, K. Nishijima, R. Otsuka, H. Nakazawa, M. Takamiya, Y. Kato, T. Yoshizawa, T. Tanaka, Y. Kudoh, J. Yamazaki, N. Kushida, A. Oguchi, K. Aoki, S. Masuda, M. Yanagii, M. Nishimura, A. Yamagishi, T. Oshima, and H. Kikuchi.** 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, Sulfolobus tokodaii strain7. DNA Res. **8:**123-140.

28. **Keller, F., and M. Pharr.** 1996. Metabolism of carbohydrates in sinks and sources: galactosyl-sucrose oligosaccharides, p. 157-183. *In* E. Zamski and A. A. Schafer (ed.), Photoassimilate distribution in plants and crops. Marcel Dekker, New York, N.Y.

29. **Lamble, H. J., N. I. Heyer, S. D. Bull, D. W. Hough, and M. J. Danson.** 2003. Metabolic pathway promiscuity in the archaeon *Sulfolobus solfataricus* revealed by studies on glucose dehydrogenase and 2-keto-3-deoxygluconate aldolase. J. Biol. Chem. **278:**34066-34072.

30. **Lamble, H. J., A. Theodossis, C. C. Milburn, G. L. Taylor, S. D. Bull, D. W. Hough, and M. J. Danson.** 2005. Promiscuity in the part-phosphorylative Entner-Doudoroff pathway of the archaeon *Sulfolobus solfataricus*. FEBS Lett. **579:**6865-6869.

31. **Lee, R. H., M. C. Lin, and S. C. Chen.** 2004. A novel alkaline alpha-galactosidase gene is involved in rice leaf senescence. Plant Mol. Biol. **55:**281-295.

32. **Lehle, L., and W. Tanner.** 1973. The function of myo-inositol in the biosynthesis of raffinose. Purification and characterization of galactinol:sucrose 6-galactosyltransferase from *Vicia faba* seeds. Eur. J. Biochem. **38:**103-110.

33. **Liebl, W., B. Wagner, and J. Schellhase.** 1998. Properties of an alpha-galactosidase, and structure of its gene galA, within an alpha- and beta-galactoside utilization gene cluster of the hyperthermophilic bacterium *Thermotoga maritima*. Syst. Appl. Microbiol. **21:**1-11.

34. **Lieshout, J. F. T., C. H. Verhees, T. J. G. Ettema, S. Van der Sar, H. Imamura, H. Matsuzawa, J. Van der Oost, and W. M. De Vos.** 2003. Identification and molecular characterization of a novel type of alpha-galactosidase from *Pyrococcus furiosus*. Biocatal. Biotransformation **21:**243-252.

35. **Ly, H. D., S. Howard, K. Shum, S. He, A. Zhu, and S. G. Withers.** 2000. The synthesis, testing and use of 5-fluoro-alpha-d-galactosyl fluoride to trap an intermediate on green coffee bean alpha-galactosidase and identify the catalytic nucleophile. Carbohydr. Res. **329:**539-547.

36. **McIlvaine, T. C.** 1921. A buffer solution for colorimetric comparison. J. Biol. Chem. **49:**183-186.

37. **Moracci, M., A. Trincone, G. Perugino, M. Ciaramella, and M. Rossi.** 1998. Restoration of the activity of active-site mutants of the hyperthermophilic beta-glycosidase from *Sulfolobus solfataricus*: dependence of the mechanism on the action of external nucleophiles. Biochemistry **37:**17262-17270.

38. **Notredame, C., D. G. Higgins, and J. Heringa.** 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. **302:**205-217.

39. **Peterbauer, T., L. Mach, J. Mucha, and A. Richter.** 2002. Functional expression of a cDNA encoding pea (*Pisum sativum* L.) raffinose synthase, partial purification of the enzyme from maturing seeds, and steady-state kinetic analysis of raffinose synthesis. Planta **215:**839-846.

40. **Pitcher, D. G., N. A. Saunders, and R. J. Owen.** 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. **8:**151-156.

41. **She, Q., R. K. Singh, F. Confalonieri, Y. Zivanovic, G. Allard, M. J. Awayez, C. C. Chan-Weiher, I. G. Clausen, B. A. Curtis, A. De Moors, G. Erauso, C. Fletcher, P. M. Gordon, I. Heikamp-de Jong, A. C. Jeffries, C. J. Kozera, N. Medina, X. Peng, H. P. Thi-Ngoc, P. Redder, M. E. Schenk, C. Theriault, N. Tolstrup, R. L. Charlebois, W. F. Doolittle, M. Duguet, T. Gaasterland, R. A. Garrett, M. A. Ragan, C. W. Sensen, and J. Van der Oost.** 2001. The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. Proc. Natl. Acad. Sci. USA **98:**7835-7840.

42. **Snijders, A. P., M. G. de Vos, B. de Koning, and P. C. Wright.** 2005. A fast method for quantitative proteomics based on a combination between two-dimensional electrophoresis and (15)N-metabolic labelling. Electrophoresis **26:**3191-3199.

43. **Snijders, A. P., M. G. de Vos, and P. C. Wright.** 2005. Novel approach for peptide quantitation and sequencing based on 15N and 13C metabolic labeling. J. Proteome Res. **4:**578-585.

44. **Suzuki, T., T. Iwasaki, T. Uzawa, K. Hara, N. Nemoto, T. Kon, T. Ueki, A. Yamagishi, and T. Oshima.** 2002. *Sulfolobus tokodaii* sp. nov. (*f. Sulfolobus* sp. strain 7), a new member of the genus *Sulfolobus* isolated from Beppu Hot Springs, Japan. Extremophiles **6:**39-44.

45. **Xiang, X., X. Dong, and L. Huang.** 2003. *Sulfolobus tengchongensis* sp. nov., a novel thermoacidophilic archaeon isolated from a hot spring in Tengchong, China. Extremophiles **7:**493-498.

46. **Zillig, W., K. O. Stetter, S. Wunderl, W. Schulz, H. Priess, and I. Scholz.** 1980. The *Sulfolobus*-"*Caldariella*" group: taxonomy on the basis of the structure of DNA-dependent RNA polymerases. Arch. Mikrobiol. **125:**259-269.

Chapter 4

Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*

Hao Wu, Lei Sun, Stan J. J.Brouns, Sheng Fu, Jasper Akerboom, Xuemei Li, and John van der Oost

Acta Crystallograph Sect F Struct Biol Crysta Commun (2007) **63**, 239-241

ABSTRACT

A predicted GTP-binding protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*, termed SsGBP, has been cloned and overexpressed in *Escherichia coli*. The purified protein was crystallized using the hanging-drop vapour-diffusion technique in the presence of 0.05 *M* cadmium sulfate and 0.8 *M* sodium acetate pH 7.5. A single-wavelength anomalous dispersion data set was collected to a maximum resolution of 2.0 Å using a single cadmium-incorporated crystal. The crystal form belongs to space group $P2_12_12_1$, with approximate unit-cell parameters a = 65.0, b = 72.6, c = 95.9 Å and with a monomer in the asymmetric unit.

INTRODUCTION

Guanosine triphosphate (GTP) binding proteins are widely distributed across the three domains of life and constitute the GTPase superclass (Leipe et al., 2002). A common feature of these proteins is the presence of a well conserved GTPase domain. The GTPase superclass is subdivided into several superfamilies and families, as the GTPase domains are often associated with different classes of (predicted) RNA-binding and/or protein-binding domains (Leipe et al., 2002). This variable domain architecture allows GTP-binding proteins to act as molecular switches in a wide range of biological processes, including protein synthesis, signal transduction and protein trafficking (Bourne et al., 1990). Biochemical and structural analysis of poorly characterized GTPase subfamilies is expected to provide insight into the control of numerous relevant biological processes.

Sulfolobus solfataricus is a model organism of the hyperthermophilic archaea that grows optimally at 353 K. Its complete genome sequence, genetic systems and functional genomics tools have been established (She et al., 2001). A putative GTP-binding protein (SsGBP) has recently been identified in the genome of *S. solfataricus*. The C-terminal half of SsGBP (residues 179–357) corresponds to a classical ‘GTPase domain’ [COG2262, as classified in the Clusters of Orthologous Groups (COGs) database; Tatusov et al., 1997]; the N-terminal domain has been described as a ‘glycine-rich segment’ (Leipe et al., 2003). Homologues of SsGBP are present in several archaea, bacteria and eukaryotes (Caldon & March, 2003). The best characterized SsGBP homologue is a GTPase from *Escherichia coli* named HflX (Brown, 2005); a BLAST search on SsGBP at NCBI (http://www.ncbi.nlm.ni

h.gov/BLAST/) reveals significant homology to *E. coli* HflX (Z score = 123, *E* value = 10_26) spanning the entire sequence. HflX is the prototype of a family within the Obg-HflX-like superfamily of GTPases (Leipe et al., 2002). The *E. coli* hflX gene is present in a locus that governs the lysis–lysogeny decision and has been proposed to be involved in controlling the proteolysis of the λ phage cII repressor (Noble et al., 1993). The molecular mechanism of the action of HflX is unknown and no three-dimensional structures of any members of the HflX subfamily are available.

In this communication, we report the cloning, purification, crystallization and preliminary X-ray analysis of SsGBP as an initial step towards the structural and functional characterization of this relatively unknown class of GTPases.

CLONING, OVEREXPRESSION AND PURIFICATION

The Sso0269 gene (gene ID 1455417) was PCR-amplified from genomic DNA using oligonucleotide primers BG1861 (50-GCGCGCTCATGAAAACAGCTGCTCTTTTTGTATC-30) and BG1837 (50-CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG-30). The PCR product of 1069 base pairs was purified using the Qiagen kit and digested with the restriction enzymes *BspHI* and *XhoI*. The restriction product was purified from agarose gel and ligated into an *NcoI–XhoI* pre-digested pET24d vector (Novagen), resulting in a 3' gene fusion to a six-histidine-tag encoding sequence. After transformation of the ligation mixture to *E. coli* HB101, a positive clone (pWUR335) was identified by PCR and restriction-fragment analysis. The sequence of pWUR335 has been verified by sequencing (AuGCT Biotechnology, Beijing, People's Republic of China).

The pWUR335 construct was transformed into *E. coli* BL21(DE3) and a single colony was used to inoculate an overnight culture in a rotary shaker at 310 K in 100 ml LB medium containing kanamycin (50 µg ml^{-1}). This 100 ml culture was used to inoculate two 1 l batches of selective LB medium. When these cultures reached an OD_{600} of approximately 0.5, isopropyl β-$_D$-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M*, after which the cultures were incubated under the same conditions for 5 h. Cells were harvested by centrifugation at 5000g for 30 min at 277 K. The pellets were frozen immediately in liquid nitrogen and stored at 253 K.

For purification of SsGBP, approximately 5 g cell paste was resuspended in 40 ml buffer *A* [20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 20 m*M* imidazole, 10%(v/v) glycerol] and the cells were lysed by three 15 s pulses of ultrasonication at 10 µm amplitude. After spinning down the cell debris, the resulting cell-free extract was incubated at 338 K for 25 min and centrifuged at 70,000g for 30 min at 277 K to effectively remove the majority of the contaminating *E. coli* proteins. The heat-stable supernatant was applied onto a Ni^{2+}-chelating column packed with 2 ml Ni–NTA His-Bind Resin (Novagen) and equilibrated with buffer *A*. SsGBP eluted in a linear gradient of imidazole in buffer B [20 m*M* Tris–HCl pH 8.0, 0.5 *M* NaCl, 1.0 *M* imidazole, 10%(v/v) glycerol] at an imidazole concentration of approximately 500 m*M*. Fractions containing SsGBP were combined and concentrated to a volume of 1.0 ml using an Amicon Ultra-15 centrifugal filter with a 10 kDa molecular-weight cutoff (Millipore). The concentrated sample was applied onto a 10/300 GL Superdex 200 column (Amersham Biosciences) equilibrated with buffer C (20 Mm HEPES pH 7.0 and 150 m*M* NaCl). SsGBP eluted as a single peak at an apparent molecular weight of 38 kDa, suggesting that SsGBP is a monomer in solution. Analytical ultracentrifugation (ProteomeLab XL-1) confirmed a monomeric state under the conditions used (38 ± 2 kDa; data not shown).

CRYSTALLIZATION AND PRELIMINARY X-RAY ANALYSIS

A preliminary crystallization screen was carried out by the hanging-drop vapour-diffusion technique (290 K) using Hampton Research Crystal Screen with a protein concentration of approximately10 mg ml^{-1} in buffer *C*. For crystallization screening, 16-well tissue-culture plates were used. Typically, 1 µl protein solution was mixed with 1 µl precipitant solution in the drop and equilibrated over 200 µl precipitant solution. After one week, crystals of SsGBP appeared in 0.05 *M* cadmium sulfate, 0.1 *M* HEPES pH 7.5 and 1.0 *M* sodium acetate (condition No. 34). Optimization revealed that crystals of SsGBP grew optimally in 0.05 *M* cadmium sulfate, 0.1 *M* HEPES pH7.5 and 0.8 *M* sodium acetate (Fig. 1). Crystals with typical dimensions of 0.06 x 0.07 x 0.18 mm were immersed in cryoprotectant (paraffin oil, Hampton Research) for 1 min, mounted into a nylon cryo-loop and flash-cooled to 100 K in a stream of nitrogen gas. Data were collected at 100 K using an in-house Rigaku MM007 rotating-anode Cu Kα X-ray generator operating at 45 kV and 45 mA (λ= 1.5418 Å)

with an R-AXIS IV^{++} image-plate detector. The beam was focused using Osmic mirrors. Single-wavelength anomalous dispersive (SAD) X-ray data were collected to a maximum resolution of 2.0 Å. The diffraction data were indexed, integrated and scaled using the *HKL*-2000 program package (Otwinowski & Minor, 1997). The diffraction data were processed

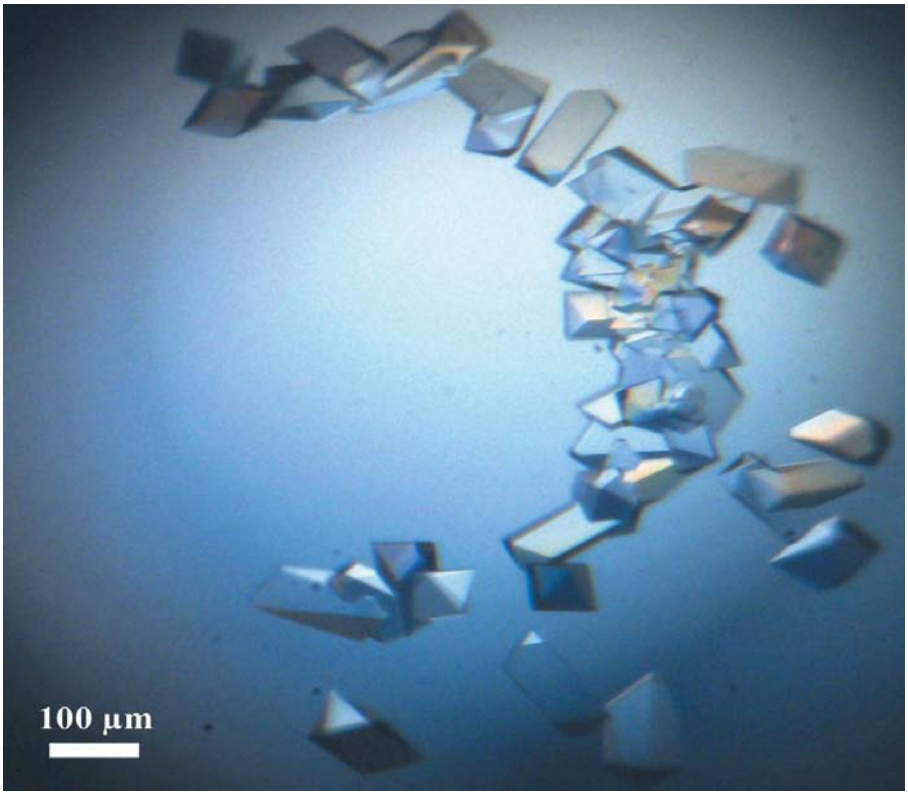


Figure 1. Crystals of *S. solfataricus* GBP grown and analyzed as described in the text

smoothly and the positions of three cadmium ions were located in the asymmetric unit. There is one protein monomer per asymmetric unit, with a V_M of 2.8 Å^3 Da^{-1} and 55% solvent content (Matthews, 1968). Data-collection statistics are summarized in Table 1.

Structure determination is currently in progress. Combined with biochemical analyses, we expect that this study will provide insights into the function of this relatively unknown subfamily of the GTPase superfamily in general and of the GBP of *S. solfataricus* in particular.

Table 1. X-ray data collection statistics for the cadmium-incorporated *S. solfataricus* GTPase crystal. Values in parentheses are for the highest resolution shell.

Space group	$P2_12_12_1$
Unit-Cell parameters (Å)	*a*= 65.0, *b*= 72.6, *c*= 95.9
Resolution range (Å)	50–2.0 (2.07–2.00) [a]
No. of unique reflections	30,618 (2,794)
Data completeness (%)	97.3 (89.6)
R_{merge} [¥] (%)	7.6 (25.6)
Redundancy	13.8 (14.0)
Average I/σ(I)	29.9 (9.5)

[¥]. $R_{merge} = \Sigma_{\mathbf{h}} \Sigma_l |I_{\mathbf{h}l} - \langle I_{\mathbf{h}} \rangle| / \Sigma_{\mathbf{h}} \Sigma_l \langle I_{\mathbf{h}} \rangle$, where I_l is the *l*th observation of reflection **h** and $\langle I_{\mathbf{h}} \rangle$ is the weighted average intensity for all observations *l* of reflection **h**.

This work was supported by grants from NWO Vici to JvdO and from KNAW to HW. We are grateful to Yi Han, ZhiyiWei, Shuang Li and Xiaoxia Yu for their excellent technical assistance and to Professor Zihe Rao for stimulating discussions.

REFERENCES

Bourne, H. R., Sanders, D. A. & McCormick, F. (1990). *Nature (London)*, **348**, 125–132.

Brown, E. D. (2005). Biochem. *Cell Biol.* **83**, 738–746.

Caldon, C. E. & March, P. E. (2003*). Curr. Opin. Microbiol*. **6**, 135–139.

Leipe, D. D., Koonin, E. V. & Aravind, L. (2003). *J. Mol. Biol*. **333**, 781–815.

Leipe, D. D., Wolf, Y. I., Koonin, E. V. & Aravind, L. (2002). *J. Mol. Biol*. **317**,41–72.

Matthews, B. W. (1968). *J. Mol. Biol*. **33**, 491–497.

Noble, J. A., Innis, M. A., Koonin, E. V., Rudd, K. E., Banuett, F. & Herskowitz, I. (1993). *Proc. Natl Acad. Sci. USA,* **90**, 10866–10870.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol*. **276**, 307–326.

She, Q. et al. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 7835–7840.

Tatusov, R. L., Koonin, E. V. & Lipman, D. J. (1997). *Science*, **278**, 631–637.

Chapter 5

Structural and functional analysis of a ubiquitous HflX-like GTPase

Hao Wu [#], Lei Sun [#], Fabian Blombach [#], Stan J.J. Brouns , Ambrosius P.L. Snijders , Kristina Lorenzen , Robert H.H. van den Heuvel , Albert J. R. Heck , Sheng Fu , Xuemei Li , Xuejun C. Zhang , Zihe Rao[*], and John van der Oost [*]

Submitted to *EMBO J.*

* Corresponding authors

[#] authors contributed equally

Data deposition: *The atomic coordinates and structure factors of SsGBP and SsGBP-GDP complex have been deposited in the RCSB Protein Data Bank (www.pdb.org) with PDB ID codes 2QTF and 2QTH, respectively.*

Summary

Guanine nucleotide-binding proteins (GBPs) of the HflX family are widely distributed across the three domains of life. Here, we present the first structural and functional information on an HflX homolog. Crystal structures of SsGBP from *Sulfolobus solfataricus* have been solved in both the nucleotide-free and the GDP-bound form. The structures reveal a two-domain arrangement with an unprecedented N-terminal domain and a classical GTPase (G) domain at the C-terminus. Biochemical analyses indicate GTP hydrolysis by a truncated G domain, and nucleic acid binding by a truncated N-terminal domain. In *S. solfataricus* co-transcription occurs of the clustered genes encoding SsGBP and Multiprotein Bridging Factor (SsMBF), a protein reported to be involved in translation fidelity. Evidence is provided here that both proteins interact *in vitro*, and that they co-purify with *S. solfataricus* ribosomal subunits. Based on these experimental results, and on analysis of the conserved genomic context of HflX-encoding genes in bacterial and archaeal genomes, we propose a role of the HflX-like GBPs in RNA maturation.

Introduction

Guanine nucleotide-binding proteins (GBPs), also referred to as P-loop GTPases, are ubiquitous in bacteria, archaea and eukaryotes (Bourne *et al.*, 1990; Leipe *et al.*, 2003; Vetter and Wittinghofer, 2001). A common feature of these proteins is the presence of a well conserved GTPase domain (G-domain) that generally acts as a molecular switch that cycles between a GTP-bound, a GDP-bound, and a nucleotide-free form. These GTPase domains are often associated with a variety of different domains (predicted to be) responsible for RNA-binding, protein-binding, or catalytic activity (Leipe *et al.*, 2003). This variable domain architecture (or subunit composition in protein complexes) allows GTP-binding proteins to control a wide range of biological processes, ranging from cell division, cell cycling, signal transduction, mRNA translation and ribosome assembly (Bourne *et al.*, 1990; Caldon and March, 2003; Karbstein, 2007).

The GTPases that resemble the archetype Translation Factors (TRAFAC) are sub-divided in 18 Families (Leipe *et al.*, 2003). Some of these families are very well characterized in terms of structure and function, and many are involved in a range of cellular processes in which RNA plays a key role. The classic translation factor family has been studied in great detail, revealing their important role in mRNA translation and protein synthesis, e.g. elongation factor Tu (EF-Tu/EF-1α) is responsible for delivery of

aminoacyl-tRNA molecules to the ribosomal A-site, whereas elongation factor G (EF-G/EF-2) catalyses the step-wise translocation of mRNA and peptidyl-tRNA in the ribosome (reviewed by (Nilsson and Nissen, 2005)). Another well-studied type of GTPase is the TrmE family (*Thermotoga maritima* TrmE), members of which have been demonstrated to be directly involved in modification of the anticodon loop of several tRNAs (Scrima *et al.*, 2005). In addition, structural insight in some GTPase-types has recently been reported on members of the Obg family (*Bacillus subtilis* (Buglino *et al.*, 2002); *Thermus thermophilus* (Kukimoto-Niino *et al.*, 2004)), the Era family (*E. coli* Era (Chen *et al.*, 1999) and the EngA family (*B. subtilis* YphC (Muench *et al.*, 2006)). The latter three GTPases play important roles in bacterial ribosome assembly (reviewed by (Karbstein, 2007)). On the bases of a typical active site architecture (switch I), the Obg family has been proposed to cluster with the HflX family in the Obg-HflX superfamily (Leipe *et al.*, 2003).

HflX (GenBank ID/gi: COG2262) is well conserved in many bacteria (Gram negatives and Gram positives), archaea (Crenarchaea and Euryarchaea), and eukaryotes (Invertebrates and Vertebrates) (Leipe *et al.*, 2003). The first member of the HflX family was detected in the *E. coli* hlfA locus, and reported to be indirectly involved in the lysis/lysogeny decision of phage lambda (Noble *et al.*, 1993); however, the actual biological function of HflX remains to be established. No archaeal homologs have hitherto been characterized. The genes encoding the human pseudoautosomal GTP-binding protein-like (PGPL) HflX-homolog have been reported to be linked to prostate cancer (Lau and Zhang, 2000); however, the molecular basis of this phenomenon is unknown.

To gain deeper insight into relevant features of the HflX-like GBP family, we set out to produce the GBP protein of the archaeon *Sulfolobus solfataricus* (SsGBP) to allow its biochemical analysis. We also determined the crystal structures of SsGBP and its complex with the cofactor GDP both at 2.0-Å resolution. In addition, an *in silico* analysis was performed of the conserved genomic neighbourhood of the prokaryotic *hflX* genes. Our results reveal that the HflX family constitutes a novel type of P-loop GTPase, and it is proposed that the HflX-like GBPs may be involved in tRNA processing and as such may contribute to translation fidelity.

Results

HflX family

HflX homologs (COG2262) are present in genomes of bacteria (149 out of 312 genomes), archaea (26/41) and eukaryotes (8/24) [http://img.jgi.doe.gov/cgi-bin/pub/main.cgi]. Sequence alignment revealed a relatively high conservation, especially in the predicted G domain at the C-terminus. Most homologs have rather variable extensions at their C-terminus, and sometimes also at their N-terminus. SsGBP represents a minimal size variant within the HflX family, lacking extensions at either side (Fig. 1).

Crystal structure of SsGBP

To get structural insights into the novel ubiquitous HflX-like GTPase family, we crystallized the full-length SsGBP both in the nucleotide-free and the GDP-bound form. The complex crystal of SsGBP with cofactor was obtained by soaking GDP into apo-SsGBP crystals. Both crystal structures were determined to 2.0 Å resolution by the Single-wavelength Anomalous Diffraction (SAD) method.

Both structures are composed of 312 residues out of the total of 356, with residues 123–143, 166–178, and 203–213 missing (Fig. 2). The two structures are identical within experimental error (RMSD 0.3 Å for all 312 Cα atoms). The SsGBP structure reveals a two-domain architecture (Fig. 2). The N-terminal domain (SsGBP-N) is subdivided into an α/β subdomain I (residues 1–99) that is composed of a four-stranded parallel β-sheet surrounded by four α-helices (Nβ1–Nβ4, and Nα1–Nα4), and a two α-helix subdomain II (residues 100–178; Nα5–Nα6) (Fig. 2). The link of the two subdomains is novel (see below), and consequently the combination of this domain with a canonical GBP-domain (SsGBP-G; discussed below) has not been described before.

N-domain of SsGBP

Structural comparison of the α/β subdomain I reveals only distant structural similarity to domains of distinct proteins (Fig. S1). Interestingly, however, the best hits includes the RNA-binding domain of Argonaute from *Archaeoglobus* fulgidus (PDB ID: 2BGG; Z-score: 5.5; RMSD: 2.8) (Parker et al, 2005) and *Aquifex aeolicus* (PDB ID: 1YVU; Z-score: 6.1; RMSD:3.3) (Yuan *et al.*, 2005). Limited homology was also observed with

yeast arginyl-tRNA synthetase (Delagoutte *et al.*, 2000) with SsGBP-N gave also a significant hit (PDB ID: 1F7U. Z-score: 6.0; RMSD: 3.1). Despite the poor conservation of amino acid residues between SsGBP-N and the two classes of proteins, similarities in overall fold (Fig. S1) and in positively charged surface residues (Fig. 3) may suggest an analogous role in RNA binding (see discussion below). Among the completely conserved residues (Fig. 1), Glu15 is involved in a salt bridge with Arg238 of the G domain (Fig. S2). In addition, Leu91, Phe94, Ala98 and Ala110 are involved in domain interaction (Fig. 1, 2 and 4). Both Leu19 and Glu102 lie in the cavity between the N and the G domain, and residue Asp85 is not located on the surface; therefore, it is unlikely that these three residues are involved in the interaction with polypeptides/nucleic acids. Gln106 is exposed, and does not appear to interact with other side chains; the fact that it is completely conserved may suggest it plays a functionally important role (see below).

The subdomain II has an anti-parallel coiled-coil fold that links subdomain I and the G-domain (Fig. S2). The two α-helices interact via a series of hydrophobic interactions: Met105, Leu109, Leu112, Leu116, and Ile119 (Nα5); and Ile146, Tyr149, Ile153, Leu156, and Leu160 (Nα6). Additional salt bridge between Glu108 and Arg152 further strengthens the tight interaction of the two helices. Many positively charged residues are exposed to the surface of the coiled-coil domain: Lys113, Lys120, Lys147, Lys150, Arg151, and Lys164 on the side facing the G domain; and Lys101, Lys104, Arg155, and Arg152 on the opposite side. The former forms a highly positive surface, together with Arg69 from Nα3 (Fig. 3A), providing a potential surface for binding nucleic acid (most likely RNA; see below). The overall positively charged, surface-exposed residues of subdomain II are well conserved in the HflX family (Fig. 1), supporting their functional significance.

A structural homolog of this coiled–coil subdomain has been frequently observed in other protein structures. Interestingly, the highest hits in a Dali-search include an effector of small GTPases: the coiled–coil domains of Rab effector Rabenosyn-5 (PDB ID: 1YZM. Z-score:7.5; RMSD 0.9 Å and 1Z0J. Z-score:7.8; RMSD 1.5 Å) (Eathiraj et al., 2005). The effector contacts both switch I and II of the activated GTPase (Bustelo et al., 2007; Grosshans et al., 2006). Since no structural information is available of SsGBP in the presence of GTP or an analogue (see below), a functional role of subdomain II remains to be established.

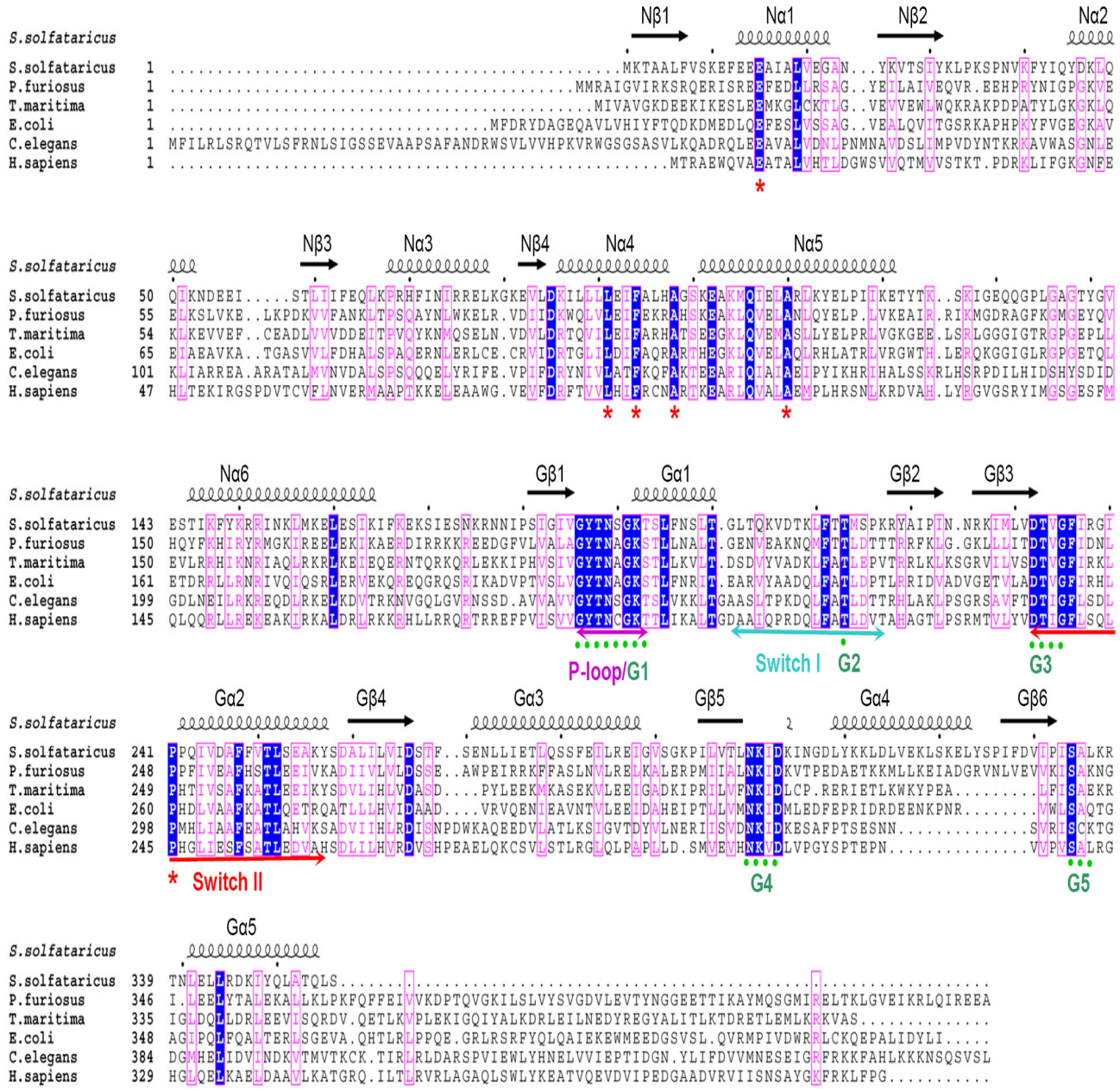


Figure1. Sequence conservation of HflX family. (A) Sequence alignment of GBP from *S.solfataricus* (gi:15897212), *P.furiosus* (gi:18977549), *T.maritima* (gi:15643293), *E.coli* (gi:16131995), *C.elegans* (gi:17561038) and *H.sapiens* (gi:6912588). Identical and relatively conserved residues are highlighted in blue and colored in purple, respectively. The identical residues involved in domain interaction are indicated by red asterisks. The nucleotide binding motifs (G1–G5) are indicated by green dots. The P-loop and switch regions in the G domain are indicated by underlines.

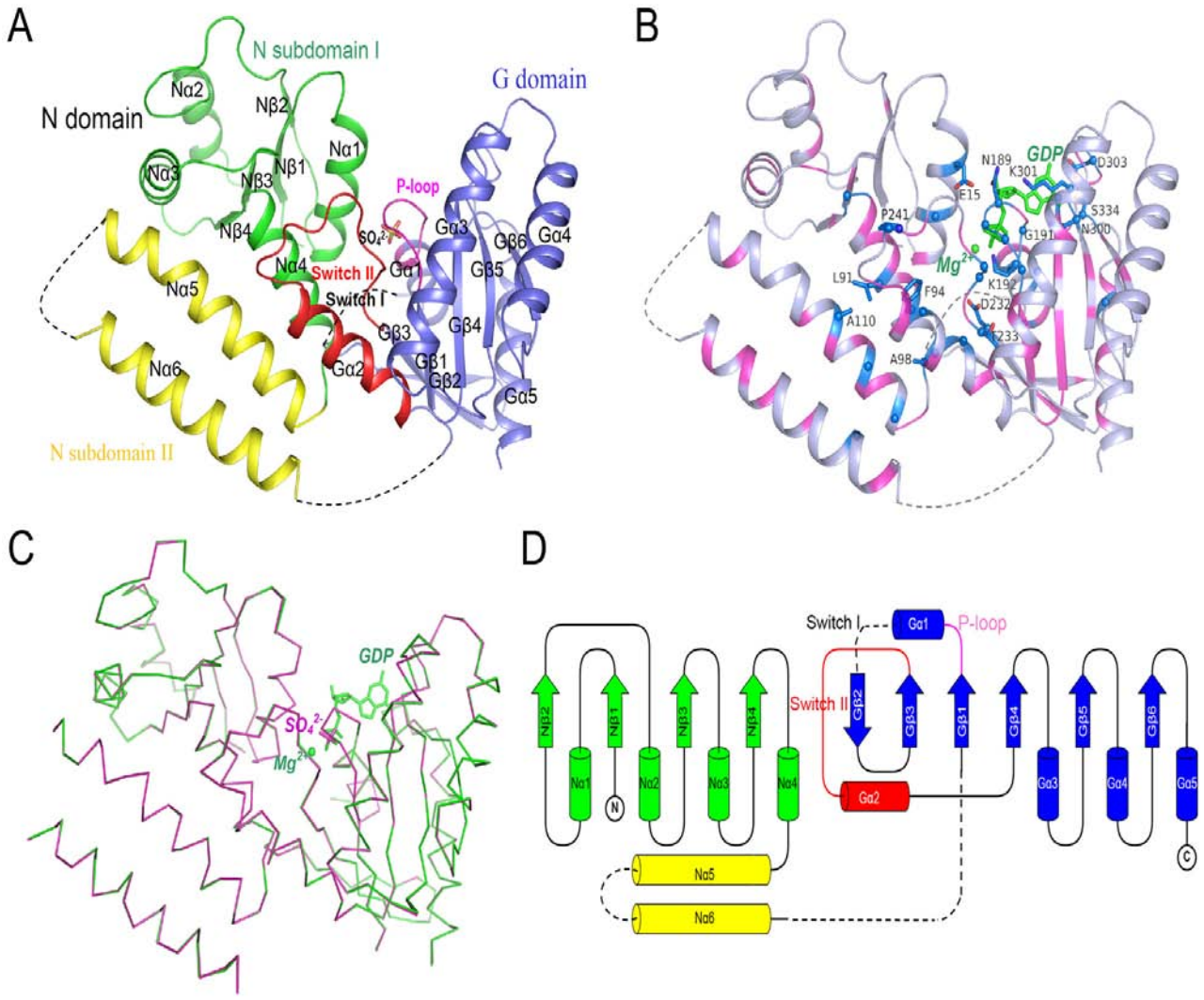
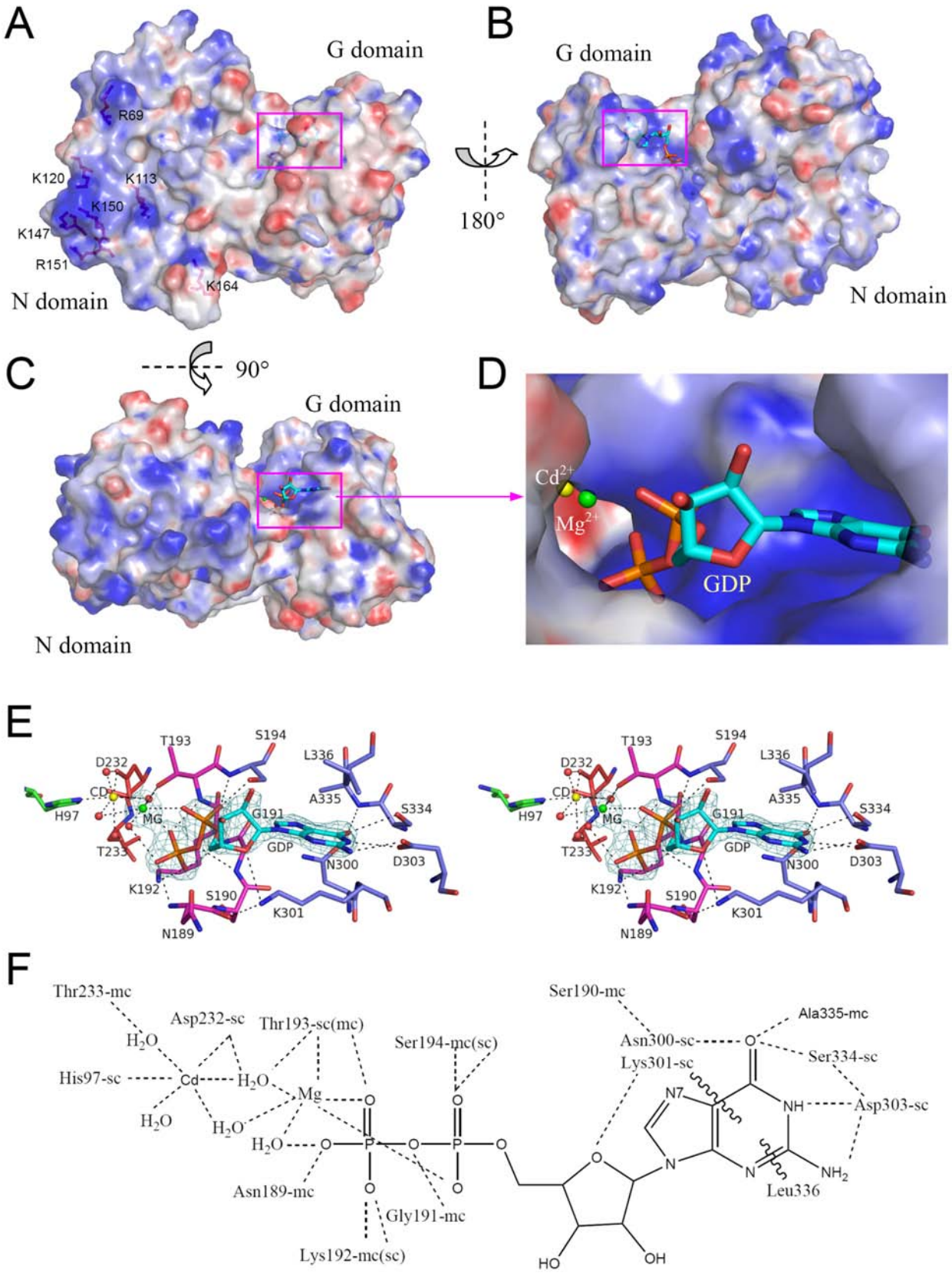


Figure 2. Overall structures of SsGBP. (A) The ribbon representation of SsGBP. Three domains are distinguished: the N1 subdomain (residues 1-99, green), the N2 subdomain (residues 100-165, yellow), and the G domain (residues 179-356, blue), the latter of which contains the P-loop region (magenta) and the switch II region (red). The sulfate ion is shown in a ball-and-stick representation. Regions missing in the final model are represented as dashed line. (B) The ribbon representation of SsGBP-GDP. The residues which show conservation in HflX family are distinguished: identical residues (blue) and relatively conserved residues (purple). The identical residues involving in the domain interactions and GDP-Mg^{2+} binding are shown in a ball-and-stick representation, and the Cα atoms of the others are shown in spheres. GDP is shown in a ball-and-stick representation. Mg^{2+} is shown as a green ball. (C) Backbone superposition of SsGBP (magenta) and SsGBP-GDP complex (green). (D) Topology diagram of GBP. The arrows show the directions of β-strands, whereas the α-helix is represented by a cylinder. Regions missing in the final model are represented as dashed line.

G-domain of SsGBP

The G-domain of SsGBP (residues 179-356) is composed of 6 ß-strands (Gß1-Gß6) and 5 α- helices (Gα1–Gα5) (Fig. 2A). It closely resembles the folding pattern of the small Ras-like GTPases, and the G-domain of Obg family GTPases (Fig. S3), representing the typical fold of the TRAFAC-GTPases. SsGBP contains all five conserved nucleotide-binding motifs: Gx_5GKS/T (G1) or P-loop, T (G2), Dx_2G (G3), NKxD (G4), and SAL/K (G5) (Bourne et al., 1990; Caldon and March, 2003; Karbstein, 2007; Leipe et al., 2003). In the GDP-bound form of SsGBP, the GDP molecule and Mg^{2+} ion are well defined in the electron density. The sites involved in the binding of GDP are located at the P-loop (G1: Asn189, Ser190, Gly191, Thr193, and Ser194), the switch II region (G3: Asp232 and Thr233), G4 (Asn300 and Lys301), and G5 motif (Ser334, Ala335, and Leu336). Most of these residues are highly conserved in the HflX family. In the nucleotide-free structure, we observed strong electron density at a position corresponding to the β-phosphate group of the GDP molecule in SsGBP, as well as of the GTP analogue in Ras (Pai et al., 1990). It was interpreted as a sulphate ion which was present in the crystallization buffer. Cadmium was present in the buffer as well and has been used for SAD phasing. A Cd^{2+} ion was observed in both nucleotide-free and GDP-bound forms of SsGBP. It is located at a distance of 4.5 Å from the Mg^{2+} ion and is coordinated by His97 from the Nα4 helix, and Asp232 in the Switch II region, as well as two bridging water molecules shared with Mg^{2+}, and another two water molecules (Fig. 3EF).

Figure 3. GDP binding site on SsGBP. (A, B and C) Electrostatic surface representation of the 'front', 'back' and 'top' views of the SsGBP-GDP complex. Red and blue surfaces represent negative and positive potentials (–15 kBT^{-1} to 15 kBT^{-1}), respectively. GDP is shown in a ball-and-stick representation. The green and yellow balls represent Mg^{2+} and Cd^{2+}, respectively. The positively charged residues located in the highly positive region are shown in a ball-and-stick representation. (D) Close view of the GDP binding surface. (E) Stereo view of the GDP binding site. GDP and the residues involved in the interaction are shown in a ball-and-stick representation, with nitrogen atoms in blue, oxygen atoms in red, and phosphorus atoms in orange. The dashed lines indicate hydrogen bonds. The superimposed ($2F_o - F_c$) electron density map was calculated with the GDP omitted from phasing and was contoured at 1.0σ with a 2.0-Å cover radius. The red balls are water molecules bound to GDP-Mg^{2+} or Cd^{2+}. The orientation is the same as in C. (F) Schematic diagram showing the hydrogen bonding interactions between SsGBP and GDP-Mg^{2+}. An 'sc' index after a residue number means that a side-chain atom is involved. 'mc' designates the involvement of a main-chain atom. H_2O indicates a water molecule. Waved lines symbolize stacking interactions.


A
G domain
R69
K120
K113
K150
K147
R151
K164
N domain
B
G domain
180°
N domain
C
90°
G domain
N domain
D
Cd2+
Mg2+
GDP
E
T193
S194
L336
D232
CD
A335
S334
H97
MG
G191
T233
GDP
N300
D303
K192
S190
K301
N189
F
Thr233-mc
Asp232-sc
Thr193-sc(mc)
Ser194-mc(sc)
Ser190-mc
Ala335-mc
Asn300-sc
Ser334-sc
Lys301-sc
His97-sc
Cd
Mg
N7
NH
Asp303-sc
Asn189-mc
Gly191-mc
Leu336
Lys192-mc(sc)
HO
OH

As in many GTPases, the switch I region of SsGBP is disordered in both nucleotide-free and GDP bound forms, and only weak electron density was visible. Generally, this flexible loop appears to be fixed only when GTP (or its analogue, *e.g.* GppNHp) is bound; attempts to obtain crystals of these complexes with SsGBP were not successful. In order to get some insight in the switch I region and the GTP-bound SsGBP structure, we compared the G domain of SsGBP to nucleotide-free Obg from *Thermus thermophilus* HB8 (PDB ID:1UDX, RMSD 1.7 Å), as well as to the GDP-bound form (PDB ID:4Q21, RMSD 2.8 Å) and the GTP-bound form of human Ras (PDB ID:1QRA, RMSD 2.7 Å) (Fig. S3AB). The main differences in the two nucleotide-bound states are within the switch I and switch II regions. The helices flanking the central β-sheet core on both sides show larger deviations. The positions of G4 and G5 regions (recognition and fixation of the guanine base) superimpose well with Ras and Obg. P-loop/G1 region, which mainly contacts the β- and γ- phosphate of the nucleotide, does not show much difference.

The comparison of selected G-domains reveals that, although the switch I region is not very well conserved, it can roughly be divided into two conformations: it takes a position that is either close to nucleotide binding site (as in Ras), or far away from it (as in Obg) (Fig. S3). The switch I region of SsGBP can not adopt a conformation like Ras switch I, because of steric hindrance of the Nα1 and Nα4 helices of the N-domain. Therefore, in its GDP-bound or free form, the switch I region of SsGBP is far away from the nucleotide binding sites, resembling the position of the Obg switch I region. Interestingly, on the basis of switch I conservation (a typical Phe close to the potentially catalytic Thr (G2); Fig. S3C) the Obg and HflX families have been proposed to form a superfamily (Leipe *et al.*, 2003).

The switch II region in SsGBP is one-residue longer than that from Obg, and 8-residues longer than that in Ras and interacts extensively with the N-domain. The switch II Gα2 in SsGBP forms a helix bundle with the Nα4 and Nα5 helices in the N-domain (Fig. 4), whereas in Obg, only one residue (arginine) forms two hydrogen bonds with two residues from another domain. Ras just contains G domain and therefore is independent of domain interaction. The differences in domain architecture explain the observed structural difference between Gα2 of SsGBP and Ras, and to some extent with Obg.

Ras-like small GBPs do not have intrinsic GTPase activity, but rather require a GTPase-activating protein (RasGAP) that forms a complex, and supplies a catalytic arginine residue (via an arginine-finger) assisting the glutamine (Gln61) in the Dx_2GQ motif that coordinates an attacking water molecule (Bos *et al.*, 1984; Milburn *et al.*, 1990). Another small GTPase Rap also misses the glutamine residue, and requires a dual finger activator (Rap1GAP), supplying both catalytic residues Arg and Gln (Pan *et al.*, 2006). At least several multi-domain GTPases have intrinsic GTPase activity and as such do not require assistance by an activating protein; the fact that the glutamine is replaced by a hydrophobic amino acid in the Obg-HflX superfamily (Fig. 1, S3C), indicates a distinct catalytic site. There is a single conserved glutamine residue in the HflX family located on Nα5 (Fig.1) which appears to be located, at least in the nucleotide-free and GDP-bound structures, too far from the active site. The invariant magnesium ion, as well as the conserved threonine in switch I (G2) of several multi-domain GTPases have been proposed to be involved in GTP hydrolysis; however, details of the catalytic mechanism remain to be revealed.

Domain interfaces

The N-terminal subdomains and the G-domain of SsGBP have an extensive interface (Fig. 2, 4). The Nα4, Nα5, Gα2 and the P-loop region are highly conserved in TRAFAC GTPases (Fig. S3). In the refined crystal structure, the two domains bury a total of 1,870 $Å^2$ solvent accessible surfaces (SAS, calculated using a 1.4 Å probe). On the N-domain side, the Nα1, Nα4, and Nα5 helices contribute to the entire G-domain binding site. On the G-domain site, the switch II region (residues 232 – 257) is the major interaction motif. The interface features include hydrophobic contacts surrounded by a few hydrogen-bonding and salt-bridge interactions. Nα4 and Nα5 together with Gα2 of switch II form an α-helices bundle. Nα1 interacts with P-loop and Switch II by a hydrogen bond (Glu14-Asn189) and a salt bridge (Glu15-Arg238) respectively. The conserved acidic residue Asp232 from the beginning of switch II region and semi-conserved basic positively charged His97 in Nα4 interact through common ligand binding (Fig. 3F) and might therefore play a critical role in the catalytic reaction or the stabilization of the N- and G- domains in SsGBP. Some of the residues involved in the domain interaction are completely conserved within the HflX family, including Glu15, Leu91, Phe94, Ala98, Ala110 and Asn189 (Fig. 1 and 2B), which implies that the HflX

family members may adopt a similar domain composition. Comparison of the nucleotide-free and GDP-bound structures does not reveal major changes at the domain interfaces. Only Nα6 of SsGBP-N slightly turns and approaches the switch II region, but they are not in direct contact.

Nucleotide binding and Hydrolysis

The interaction of SsGBP with different nucleotides has been analyzed by native Mass Spectrometry (Fig. S4AB). GDP or GTP were mixed with SsGBP in the presence Mg^{2+}. In case of GDP, no apo SsGBP was detected, 81% of SsGBP has one nucleotide and 19% appears to have two nucleotides bound (Fig. S4C). In the presence of GTP, 25% of SsGBP is present in the apo form, and equal amounts of the remainder complexes with either GTP or GDP (Fig. S4C). This suggests that the affinity of SsGBP for GDP is higher than that for GTP. Since we did not observe the second GDP binding site in the crystal structure, binding of more than one nucleotide is most likely due to non-specific association with the positively charged surface of the N-terminal domain of SsGBP. The latter assumption would be in agreement with the observation that the truncated SsGBP-N domain (residues 1–176) is co-purified with RNA/DNA from its *E. coli* host; in contrast, the truncated SsGBP-G domain (residues 178–356) is free of nucleic acid when produced similarly (see below).

GTPase activity of SsGBP has been studied by thin layer chromatography (TLC) using ^{32}P-labelled GTP (Fig. 5). Mg^{2+} dependent GTPase activity (hydrolysis of GTP to GDP) could be demonstrated with the full length SsGBP as well as with the truncated SsGBP-G domain. The molar ratios (GBP/α^{32}P-GTP) were 1.7 and 2.2 for SsGBP and SsGBP-G, respectively. No activity was detected with the truncated N-domain. The time series shows that under the experimental conditions the GTPase activity of SsGBP-G is higher than that of full-length SsGBP (Fig. S5). The kcat of SsGBP-G is 1.54 ± 0.01 min^{-1}. The Km-value is 12.9 ± 0.8 µM. Activity of the full-length SsGBP was more than ten times lower compared to SsGBP-G. These results are in agreement with SsGBP being a slow GTPase with regulatory role (Fig. S5).

Genomic Context

In prokaryotes, the conserved genomic context is a powerful tool in predicting function of genes (Ettema *et al.*, 2005). HflX-encoding genes in both bacteria and archaea appear to have a well conserved context, which may provide clues as to their

physiological function. The context of the SsGBP-encoding gene appears well conserved in the Crenarchaeal kingdom (Fig. 6A). There is a general clustering with the *mbf* gene encoding a homolog of the eukaryal Multiprotein Bridging Factor-1 (Aravind and Koonin, 1999). The anticipated co-expression has indeed been confirmed experimentally by Northern blot analysis (data not shown). The function of MBF has been deduced from *in vitro* protein-protein interaction studies and electrophoretic mobility shift assays; it was found that MBF associates with transcription regulators as well as with the transcription initiation machinery (Takemaru *et al.*, 1997). Interestingly,

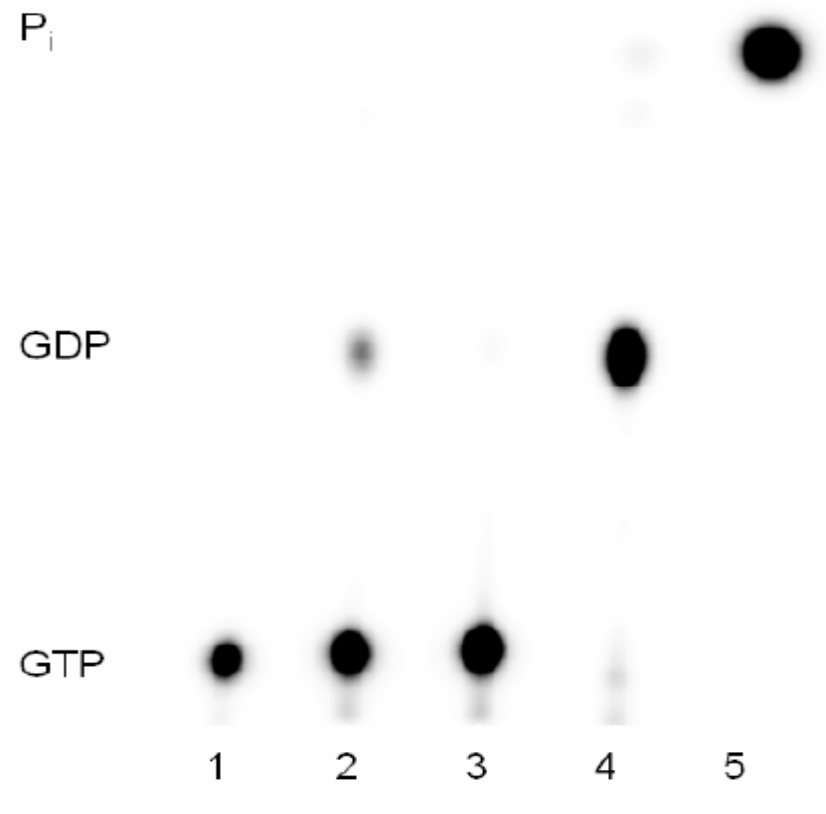


Figure 5. GTP hydrolysis. GTPase activity of SsGBP, SsGBP-N, and SsGBP-G measured by thin layer chromatography (TLC) using α-^{32}P-labelled GTP (in all lanes). Lane 1, no protein (negative control); Lane 2, SsGBP; Lane 3, SsGBP-N; Lane 4, SsGBP-G; Lane 5, Calf Intestine Alkaline Phosphatase (CIAP, a positive control).

a different role for MBF has been reported; an *mbf* mutant of yeast appeared to result in increased levels of frameshift suppression *in vivo* on the basis of which a role in translation fidelity has been proposed (Hendrick *et al.*, 2001). Another set of genes that are clustered with *mbf* and the *hflX*-homolog in Crenarchaea (to which *Sulfolobus* belongs) encode proteins that appear to be involved in RNA binding (COG1958, COG1370) and RNA modification (COG0343). The latter gene encodes queosine/archaeosine tRNA-ribosyltransferase; this enzyme catalyzes the modification of guanine at position 34 of a subset of bacterial tRNAs to a precursor of queuosine,

and of guanine at position 15 of approximately half of the archaeal tRNAs to an archaeosine-precursor (Sabina and Soll, 2006).

The context of *E. coli hflX* is very well conserved over a wide range of phylogenetically distinct bacterial genomes (Fig. 6B). Two genes cluster with *hflX* in phylogenetically distant bacterial genomes, including Proteobacteria (*E. coli, Pseudomonas*), Firmicutes (*Streptococcus, Listeria*), and the thermophilic bacterium *Thermotoga*: (i) *hfq,* an RNA chaperone involved in binding of various RNA molecules

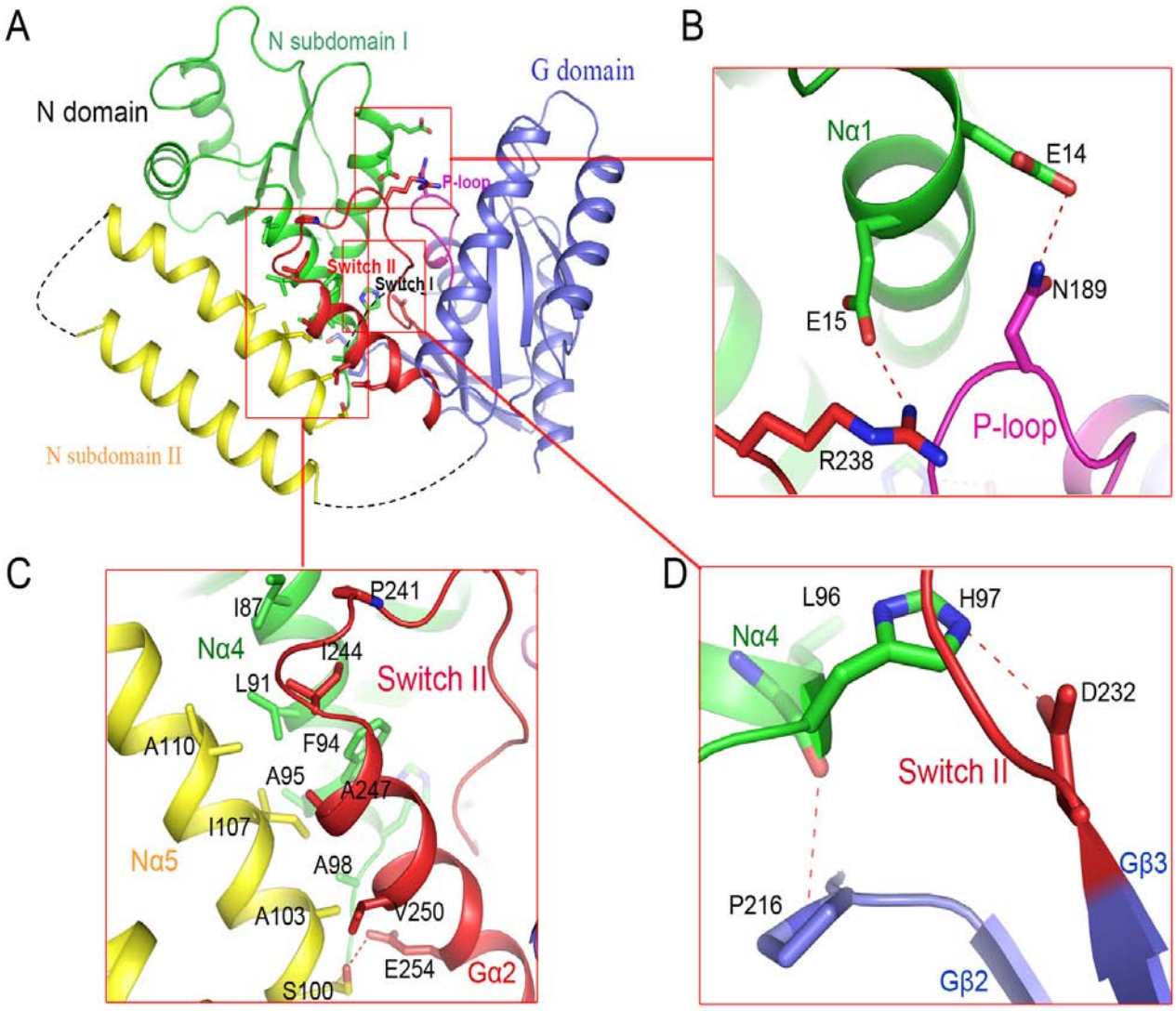


Figure 4. Interface of the N domain and the G domain. (A) Overall structure of SsGBP. The side chains of residues involved in the domain interaction are shown in a ball-and-stick representation. (B) Interaction between Nα1 and G domain. The dashed lines indicate hydrogen bonds and salt bridges. (C) Interaction between Nα4, Nα5 and Gα2. Except S100 and E254 (which form a hydrogen bond), all the other residues are involved in the hydrophobic interactions. (D) Interaction between Nα4 and G domain.

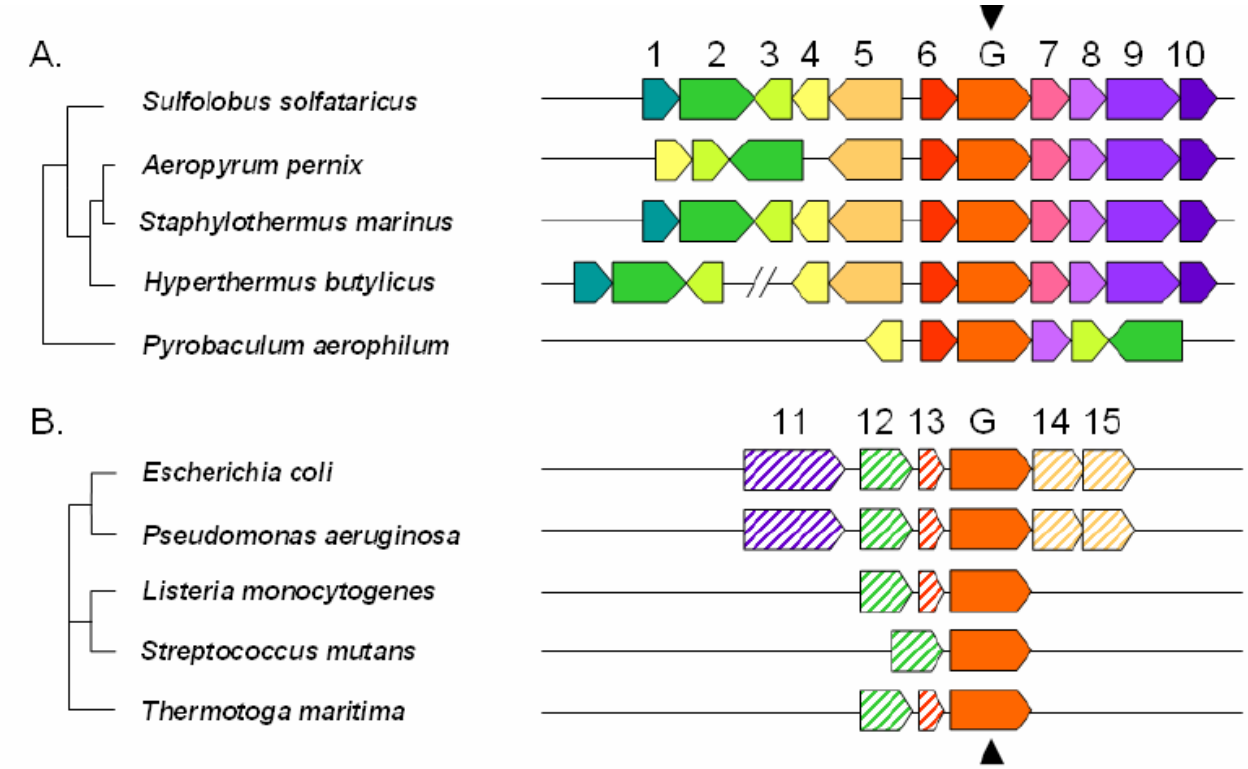


Figure 6. Genomic context of *hflX* homologs. (A) Archaea. In *Sulfolobus solfataricus* the SsGBP-encoding gene (indicated by arrow) resides in a large gene cluster that is conserved in distantly related Crenarchaea. COG classification, the *S. solfataricus* (Sso) gene identifier, and the COG name are provided. (1) COG1958 (Sso0276 - small nuclear RNP (RNA-binding protein)); (2) COG0343 (Sso0274 - tRNA ribosyl transferase (TGT)); (3) COG1938 (Sso0273 - DUF75; ATP-GRASP superfamily; ligase (C/N, C/S-CoA)); (4) COG1370 (Sso5542 - PUA domain, RNA binding; in many Euryarchaea fused to TGT (COG0343)); (5) COG1222 (Sso0271 - AAA+ type ATPase; Proteasome Activating Nucleotidase); (G) COG2262 (Sso0269 - GTP-binding protein; HflX-like GTPase); (7) COG1303 (Sso0267 - DUF12); (8) COG1675 (Sso0266 - TFE; transcription factor); (9) COG0270 (Sso0265 - Site-specific DNA methylase); and (10) COG4080 (Sso0264– RecB-family nuclease). **(B) Bacteria.** In Gram-negative bacteria, Gram positive bactyeria and *Thermotoga maritima*, the genes encoding HflX homologs display a conserved clustering. COG number/name and *E. coli* protein name are: (11) COG1370 (MutL - DNA mismatch repair protein; (12) COG0324 (MiaA - tRNA Δ^2-isopentenylpyrophosphate transferase; (13) COG1923 (Hfq - RNA binding protein; small RNA chaperone); (G) COG2262 (HflX - GTP-binding protein; GTPase); (14 & 15) COG0330 (HflK & HflC - Membrane protease subunits). *Notice:* (6) is COG1813 (Sso0270 - MBF).

(Moller *et al.*, 2002), and (ii) *miaA*, encoding tRNA-Δ^2-isopentenyl-pyrophosphate transferase, an enzyme that specifically modifies adenosine at position 37 in the anticodon-loop of some tRNAs (Zhao *et al.*, 2001). tRNA modification generally contributes to reading frame maintenance, and as such to translation fidelity (Urbonavicius *et al.*, 2001). An ortholog of the latter gene, Mod5 in yeast, has a similar tRNA modification activity; in a mutant of the latter protein translation fidelity is affected

(Benko *et al.*, 2000). Overall, the conserved gene context of *hflX*-like genes in bacteria and archaea suggests a link with RNA processing and translation.

Interaction with nucleic acid, MBF and ribosomes

Heterologous production of the full-length SsGBP and the truncated fragments SsGBP-N and SsGBP-G revealed that nucleic acid was co-purified with the SsGBP-N domain. Exposure to either RNaseA or DNaseI indicated that upon its production in *E. coli* SsGBP-N bound both RNA and DNA (data not shown).

A functional link between SsGBP and MBF has been suggested by their clustering in many crenarchaeal genomes, and by the demonstrated co-expression (data not shown). To investigate their physical interaction interaction, we performed His-tag mediated pull-down experiments. For this purpose, the truncated C-terminal helix-turn-helix domain of MBF (MBF-C, residues 57–165) was used, because the N-terminal domain (MBF-N) turned out to bind to the Ni-NTA resin (data not shown), most likely due to an endogenous metal binding site in a zinc-ribbon structural motif. The significantly increased amount of MBF-C in the presence of the GTPase (apparently independent of GTP and GDP under the conditions tested), does suggest an interaction of SsGBP and the truncated SsMBF-C (Fig. 7A). The identity of the separated SDS-PAGE fragments was confirmed by MS analysis of trypsin fragments and by anti-MBF immunodetection (Fig. 7A). Interaction between SsGBP and MBF-C could not be demonstrated with ESI-MS (data not shown).

Because of the proposed links of MBF as well as Obg-like GTPases with translation fidelity and ribosome assembly, respectively (Hendrick *et al.*, 2001; Karbstein, 2007), a possible interaction with ribosomes was tested. Immunological analysis of *S. solfataricus* ribosomal subunits that were fractionated on a sucrose gradient, indeed indicated the presence of SsGBP and SsMBF. The SsMBF was highly enriched in fractions that contained the 30S componenets (Fig. 7B), whereas the peak of SsGBP matched to the 50S large ribosomal subunit.

Discussion

HflX structure

SsGBP and its homologs from the HflX family display a new, two-domain configuration in the P-loop GTPases. SsGBP contains a G domain that is very well conserved in the

TRAFAC superclass, especially with the small GTPase Ras, and the multi-domain Obg (Fig. 5). Ras is a stand-alone G-domain, whereas Obg has a three-domain configuration: an N-terminal 'extended helical' domain (Gly rich), a central G domain, and a C-terminal RNA-binding domain (see below). Apart from the relatively well conserved G-domain

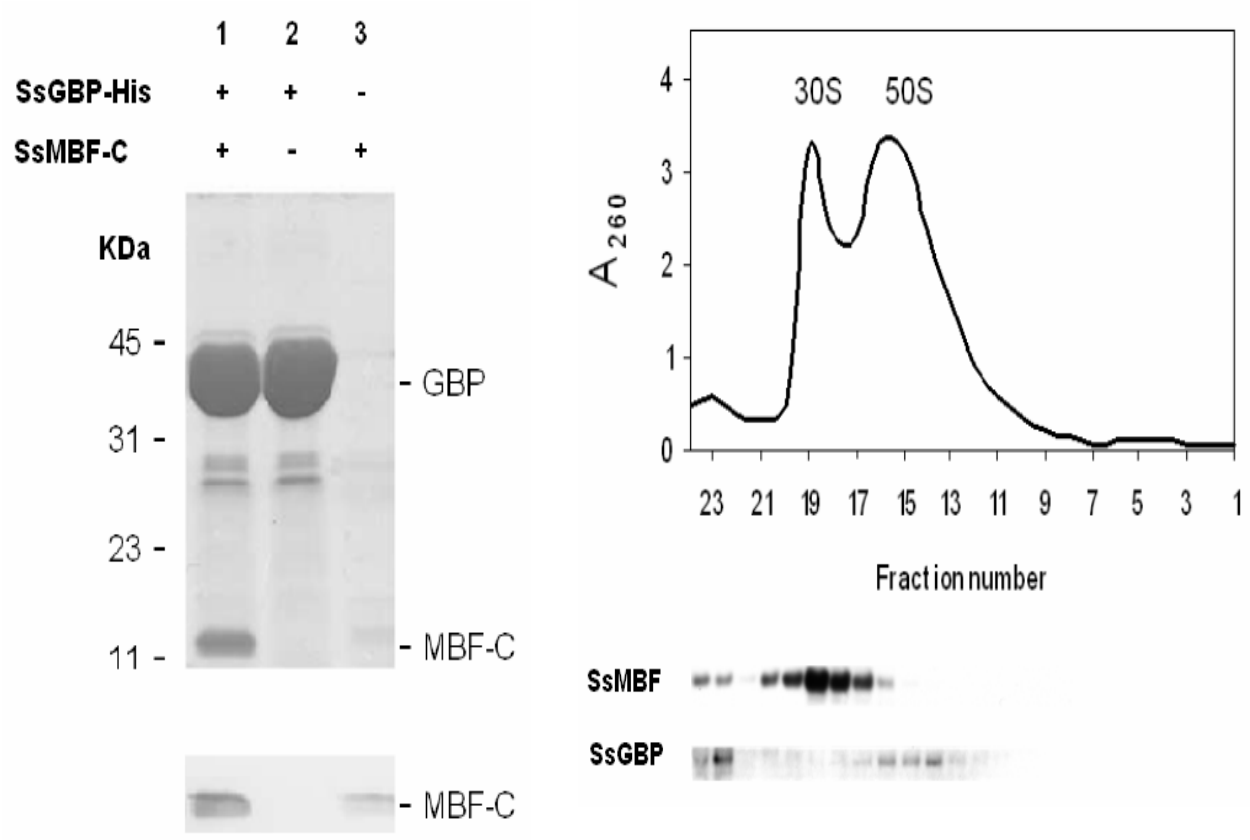


Figure 7. SsGBP interacts with MBF and ribosomes. (A) Interaction in vitro of SsGBP and SsMBF. Pull down of SsMBF-C by SsGBP-His in Ni-column. (B) Interaction in vivo of SsGBP and SsMBF with ribosomes. Immunological detection of SsGBP and the C-terminal fragment of SsMBF in enriched ribosome fractions from sucrose gradient. Ribosomal subunits form *S. solfataricus* dissociate during preparation (Londei *et al.*, 1986). 50S and 30S peaks are indicated. The upper 14 fractions containing the ribosomal subunits were analyzed by immunodetection of SsMBF and SsGBP.

with the typical FxT motif in switch I (Leipe *et al.*, 2003), there is little conservation in the overall structure between the Obg and HflX families. The SsGBP-N domain is not present in any of the currently available GTPase structures. The N- and C-termini in SsGBP are located on the surface of the protein, permitting the poorly conserved extensions at both peptide ends in the HflX family to adapt variable conformations (Fig. 1,2).

SsGBP function

Binding and hydrolysis GTP by SsGBP has been demonstrated by MS and TLC analysis, respectively. The TLC analysis suggests that the SsGBP-G domain has higher intrinsic GTPase activity than the full-length SsGBP. GTPases are generally molecular switches that use the free energy of GTP hydrolysis to drive a wide range of reactions and processes (Bourne *et al.*, 1990; Bourne *et al.*, 1991). On the basis of the structural analyses, the genomic context comparisons, and the biochemical experimental results, it is tempting to speculate on the physiological function of HflX-like GTPases in general, and SsGBP in particular.

The conserved genomic context in *Sulfolobus* species and related Crenarchaea suggested a related function of SsGBP and MBF. Experimental evidence has been provided that the genes are co-transcribed. Moreover, pull down analysis was performed that revealed significant interaction of SsGBP and the C-terminal fragment of SsMBF. The *in vitro* binding of both proteins was not influenced by the presence of nucleotides, and the GTPase activity of SsGBP was not influenced in the presence of SsMBF-C (data not shown).

Screening the PDB for structures resembling SsGBP-N did not result in highly similar structures. Interestingly, however, among the highest hits were several RNA-binding domains of distinct classes of enzymes involved in RNA processes, including the RNA-degrading Argonaut and a Class I tRNA-synthetase. Analysis of the coiled-coil of SsGBP-N revealed relatively high hits with helical subdomains of small GTPase effectors. Since proposed key amino acids of Argonaut or the tRNA-synthetase are not conserved in the SsGBP domains, no conclusions can be drawn on actual functionality. However, the overall positively charged surface of SsGBP-N (Fig. 3) would allow binding of nucleic acid. This is in good agreement with the analysis of recombinant SsGBP-N.

Analyses of purified recombinant SsGBP (full length) and SsMBF (full length) has revealed that both proteins form rather stable interactions with nucleic acid. More detailed analysis revealed that the truncated SsGBP-N fragment was responsible for binding both RNA and DNA; the SsGBP-G domain was not associated with nucleic acid when purified under the same conditions. Sequence analysis of nucleic acids association to SsMBF mainly revealed fragments of *E. coli* tRNA and rRNA (data not shown). Association with different classes of RNA has been demonstrated for many types of TRAFAC GTPases: including the well-characterized translation elongation

factors, as well as the TrmE-, Era-, EngA- and Obg-families. The Obg family contains two types of RNA domains: two variants of the latter domain have been described: TGS, *B. subtilis* (Buglino *et al.*, 2002); OCD, *T. thermophilus* (Kukimoto-Niino *et al.*, 2004).

Several GTPases have been reported to play a role in processing of RNA and/or ribosome assembly (reviewed by (Karbstein, 2007)). Because the results of both comparative and experimental analyses point to a similar role of SsGBP, its *in vivo* association with ribosomes from *S. solfataricus* was analyzed. Indeed, both SsGBP and SsMBF co-purify with the ribosomes, although they appear to associate with different subunits, 50S and 30S respectively. MS analysis has suggested that, under the experimental growth conditions, SsMBF only binds to ribosomes in sub-stoichiometric amounts.

Conclusions

A multi-disciplinary approach has been presented to analyze structural and functional features of a member of the HflX family. SsGBP has a canonical G-domain, with intrinsic GTPase activity, and apparently a tight GDP binding. The latter might suggest that GDP dissociation and GTP loading are facilitated by the interaction of GBP with its physiological partners (e.g. protein, RNA), or by a yet to be identified guanine-nucleotide exchange factor. An N-terminal domain has an overall positively-charged surface, which would agree well with the fact that its fold distantly resembles RNA binding domains of distinct proteins. Indeed, when produced in *E. coli*, SsGBP-N co-purifies with nucleic acid. *In vitro* analysis suggests that SsGBP associates with MBF, and immunological detection of *Sulfolobus* lysates indicates that SsGBP and SsMBF co-purify with ribosomes, apparently bound to different subunits. Conservation of the genomic context of SsGBP-like genes in archaea strongly suggests involvement in RNA modification/maturation; moreover, the bacterial orthologs (Proteobacteria, Firmicutes, *Thermotoga maritima*) cluster with genes encoding an RNA chaperone and a tRNA-modification enzyme. Overall, it is concluded that the HflX-like GTPases, at least in prokaryotes, appear to participate in RNA processing. Future studies are required to elucidate the enzyme kinetic properties, mechanism and regulation of GTP hydrolysis, and the relationship between GTPase activity and biological functions of SsGBP and its homologs.

Materials & Methods

Protein production & analysis

Recombinant protein of SsGBP was produced as described (Wu *et al.*, 2007). Truncated fragments of SsGBP were generated likewise, including N-terminal domain (SsGBP-N) and G-domain (SsGBP-G) (Table S1). In addition, full-length MBF and its C-terminal fragment (MBF-C) were produced similarly (Table S1). Western Blotting on Nytran nitrocellulose membrane (Schleicher and. Schuell, Dassel, Germany) was performed in 10 mM CAPS, pH 11.0, 10% methanol at 50 mA overnight in a Mini Trans-Blot Cell (BioRad). Immunodetection of SsGBP was performed using the blocking reagent I-block (Applied Biosystems) according to manufacturer's protocol with the following modifications: Washing steps were prolonged to 15 min. Antibodies against SsMBF-C and C-terminal His-tagged SsGBP were raised in rabbit at Eurogentec (Be). SsGBP-specific antiserum was used in 1:100 dilution in combination with 1:20,000 diluted ImmunoPure Goat Anti-Rabbit IgG [F(ab')2], Alkaline Phosphatase Conjugated (Pierce). For detection of SsMBF, partially purified antibodies were labeled 1:10 with Digoxigenin according to manufacturer's protocol (Roche) and used in 1:1000 dilution for detection, in combinations with 1:1500 diluted Anti-Digoxigenin-AP, Fab fragments (Roche). Chemiluminescence detection was carried out using 1:100 diluted CDP-Star reagent (New England Biolabs) and BioMax Light Chemiluminescence films (Kodak).

Crystallization, Data Collection, and Structure determination

SsGBP was crystallized by the hanging drop vapor diffusion method at 290K. Crystallization drops were prepared by mixing 1 μL of a 10 mg/mL protein solution with 1 μL of the reservoir solution containing 0.05 M cadmium sulfate, 0.1 M HEPES (pH 7.5), and 0.8 M Sodium Acetate. To obtain the SsGBP–GDP complex, the apo-enzyme crystals were soaked in a reservoir solution with 85 mM GDP and 10 mM $MgCl_2$ for three weeks. Crystals were soaked in cryo-protectant (Paraffin oil, Hampton Research) for 1 min before mounted into a cryo loop and flash-cooled to 100K in a stream of nitrogen gas. All data was collected using an in-house Rigaku MM007 rotating-anode CuKα X-ray generator (λ = 1.5418 Å) with an R-AXIS IV^{++} image plate detector. The beam was focused using Osmic mirrors. SAD data up to 2.0 Å resolution were collected from the apo-enzyme SsGBP and SsGBP–GDP complex respectively. Diffraction data

were indexed, integrated, and scaled using the *HKL2000* program package (Otwinowski and Minor, 1997). The crystal belongs to space group $P2_12_12_1$, with unit-cell parameters a = 65.0, b = 72.6, and c = 95.9 Å. There is one protein monomer per asymmetric unit with a V_M of 2.8 $Å^3$ Da^{-1} and 55% solvent content (Matthews, 1968).

The structure of apo-SsGBP was solved by cadmium-based SAD phasing. Four Cd^{2+} ions from the reservoir solution were identified by SHELXD. AutoSHARP(La Fortelle and Bricogne, 1997) was used for heavy atom refinement and phasing. To improve the quality of the electron density maps, we used SOLOMON (Abrahams and Leslie, 1996), run within autoSHARP, to perform a density modification based on solvent flattening. Arp/Warp (Perrakis *et al.*, 1999) automatically built approximately 80% of the polypeptide chains. The *Fo−Fc* difference *Fourier* electron density map and omit density map displayed clear density and were used to assign the a sulfate group and 2 acetic groups. The structure was then manually rebuilt in Coot (Emsley and Cowtan, 2004) and refined using CNS (Brünger *et al.*, 1998) and REFMAC (Murshudov *et al.*, 1997). The model was refined to a final R_{work} = 18.6% and R_{free} = 23.1%. The structure of SsGBP-GDP complex was resolved in the same way. The *Fo−Fc* difference map and omit map showed clear electron density at the canonical guanine-nucleotide binding site into which a GDP molecule was docked manually. The model was refined to a final R_{work} = 22.9% and R_{free} = 25.9%. There is no visible electron density for residues 123–143, 166–178, and 203–213. These residues are located at the surface of the molecule and presumably form flexible loops. The stereochemistry of the structure was analyzed with the programs PROCHECK (Laskowski *et al.*, 1993). Tyr42 has well defined electron density but has a distorted geometry in a disallowed region of the Ramachandran plot. Statistics of the data collection and refinement are summarized in Table 1.

Native ESI MS

The buffer of SsGBP was exchanged sequentially to 50 mM ammonium acetate (pH 6.8) using centrifugal filter units with a cut-off of 5 kDa (Millipore, England). The final concentration used for the mass spectrometry measurements was 10 µM. The samples were analysed on a LCT electrospray time-of-flight instrument. (Waters, Manchester, UK). Nanospray glass capillaries were used to introduce the samples into the Z-spray source. The source pressure was increased to 10 mbar to create increased collisional

cooling (Krutchinsky *et al.*, 1998; Tahallah *et al.*, 2001). Source temperature was set at 80°C and sample cone voltage varied from 80 to 125 V. Needle voltage was around 1300 V.

Thin layer chromatography (TLC)

GTPase activity assay of SsGBP was analyzed by TLC. Final protein concentrations: 8 µM SsGBP, 8.6 µM SsGBP-N, 9.9 µM SsGBP-G, 5000× diluted commercial product CIAP (Calf Intestine Alkaline Phosphatase; New England Biolab). Reactions were carried out at 50°C in the presence of 10 mM $MgCl_2$, 200 mM KCl, and 50 mM HEPES (pH 7.7). 4.5 µM of α-^{32}P-GTP (400 ci/ml, Amersham) was mixed with $MgCl_2$ and used to initiate reactions. At given time points reactions were quenched by the addition of an equal volume of stop buffer (2% SDS, 5mM EDTA) and put on ice. One microliter of the quenched reaction was spotted onto a Merck TLC plates 20 × 20 cm PEI cellulose F (1.05579). The TLC plate was developed in 1 M Acetic Acid, 0.8 M LiCl until eluent front reached at least 2/3 of the plate. After drying the plate, signals were quantified with a phosphorimager (Molecular Imager FX, BioRad)

Phosphate release assay

GTP hydrolysis by the truncated GTPase domain of SsGBP was measured by malachite-green assay as described by Esue et al.(Esue et al., 2005) with following modifications. All measurements were performed in 20 mM Tris/HCl pH 7.8; 200 mM NaCl, 5 mM $MgCl_2$, 5% glycerol in 50 µl volume. Absorption was measured at 690 nm in a microplate reader (iEMS Reader MF. Labsystems). In order to determine substrate concentration dependent activity, phosphate release was allowed over a 15 min time interval (during which it appeared to be linear) at 50°C. Enzyme concentration was 0.28 µM (0-100 µM GTP) or 1.38 µM (100-1000 µM GTP). Measurements were performed at least in triplicates. Values were corrected for background determined from controls without protein and controls without GTP. Inorganic phosphate concentrations were calculated using a phosphate standard in assay buffer ranging from 0-50 µM phosphate.

His-tag pull-down assay

E. coli strains for overproduction of either SsGBP+His or MBF-C were cultivated in 100 ml medium, as described before (Wu *et al.*, 2007). Cell pellets were resuspended in 1.0 ml lysis buffer (50 mM NaH_2PO_4 (pH 8.0), 10 mM imidazole, 300 mM NaCl, 5% glycerol (v/v), and when indicated supplemented with 5 mM GTP or GDP). After sonication, the solution is mixed as indicated (0.5 ml cell free extract containing SsGBP+His with 0.5 ml lysate with MBF-C), and incubated for 30 min at 72°C. Centrifuged at 13000 rpm for 30 min at 4^0C to get the supernatant. Subsequently a small scale purification by Ni-NTA resin spin column (Qiagen) was performed, with thorough washing with 5-10 column volumes wash buffer (50 mM NaH_2PO_4 (pH 8.0), 20 mM imidazole, 300 mM NaCl, 5% glycerol (v/v), and when indicated supplemented with 5 mM GTP or GDP. Bound proteins are eluted (Elution buffer: 500 mM imidazole; 50 mM NaH_2PO_4 (pH 8.0); 300 mM NaCl; 5% glycerol (v/v), and checked by SDS-PAGE and Western blot. Western blotting is executed as described above, with anti-SsMBF rabbit-raised primary antibody (1:1000) and (anti-rabbit IgG)-alkaline phosphatase conjugate secondary antibody (Promega) (1:3000). Blots were scanned with GS800 densitometer (Biorad).

Ribosome enrichment

Crude ribosomes were essentially prepared as described earlier (Londei *et al.*, 1986) with following modifications. All buffers were supplemented with 40 U/ml RNasin Ribonuclease Inhibitor (Promega). *S. solfataricus* cells were grown aerobically at 80°C on 0.4% sucrose in modified Brock medium (Brouns *et al.*, 2006) with modifications with regard to the use of Fe-citrate as described (Worthington *et al.*, 2003). Cells were grown until OD_{600} reached 1.1. Cells were disrupted by passing through a French Press Cell (Aminco) thrice at 16,000 psi. After preparation of crude ribosomes in a TFT65.13 rotor (Kontron Instruments), 300 pmol was loaded on a 15-30% sucrose gradient (Hasenohrl *et al.*, 2006) and centrifuged in a TST41.14 rotor (Kontron Instruments). Fractions of 500 µl were collected and absorption at 260 nm and 280 nm was measured to localize the two peaks corresponding to the ribosomal subunits. After TCA-precipitation, precipitates were dissolved in 2x SDS-PAGE sample buffer.

Table 1. Statistics of diffraction data and structure refinement of SsGBP

	Apo-SsGBP	SsGBP-GDP complex
Data collection		
Wavelength (Å)	1.5408	1.5408
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å) a, b, c	65.1, 72.6, 95.9	65.0, 72.4, 96.0
Resolution (Å)	50.0 – 2.00 (2.07–2.00)[c]	50.0 – 2.0 (2.07–2.00)
Unique reflections	30618 (2794)	30589 (2945)
Completeness	97.1 (89.6)	97.1 (94.3)
R_{merge}[a]	0.076 (0.256)	0.070 (0.451)
$\langle I/\sigma(I)\rangle$	29.9 (9.5)	16.1 (4.6)
Redundancy	13.8 (14.0)	6.7 (6.2)
Refinement		
Resolution range	30–2.0	20–2.0
R_{work}/ R_{free}[b] (%)	18.6/23.1	22.4/26.4
RMS deviation:		
bonds (Å)	0.010	0.011
angles (°)	1.60	1.80
Average B factor (Å^2):		
Protein	37.4	35.4
Water	43.2	38.0
Metal ions	34.9	37.5
Other ligand	33.4	50.1
Ramachandran plot		
Favoured (%)	94.6	94.3
Allowed (%)	4.7	5.0
Generously (%)	0.4	0.4
Disallowed (%)	0.4	0.4

[a] $R_{merge} = \mathbf{\Sigma_h \Sigma_l} |\boldsymbol{I}_{hl} - \langle \boldsymbol{I}_h \rangle| / \mathbf{\Sigma_h \Sigma_l} \langle \boldsymbol{I}_h \rangle$, where $\boldsymbol{I}_l$ is the lth observation of reflection **h** and $\langle \boldsymbol{I}_h \rangle$ is the weighted average intensity for all observations l of reflection **h**.

[b] $R_{work} = \Sigma(||F_p(obs)| - |F_p(calc)||)/ \Sigma|F_p(obs)|$; R_{free} = R factor for a selected subset (5%) of the reflections that was not included in prior refinement calculations.

[c.] Numbers in parentheses are corresponding values in the highest resolution shell.

Acknowledgements

The authors greatfully acknowledge Dr. Paola Londei (Rome) for valuable practical suggestions, and Dr. Zhiyong Lou (Beijeng) for the X-ray data collection of SsGBP-GDP.

References

Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. Acta Crystallographica Section D, 52, 30-42.

Brünger , A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr, 54 (Pt 5), 905-921.

Abrahams, J.P. and Leslie, A.G.W. (1996) Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallographica Section D*, **52**, 30-42.

Aravind, L. and Koonin, E.V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res*, **27**, 4658-4670.

Benko, A.L., Vaduva, G., Martin, N.C. and Hopper, A.K. (2000) Competition between a sterol biosynthetic enzyme and tRNA modification in addition to changes in the protein synthesis machinery causes altered nonsense suppression. *Proc Natl Acad Sci U S A*, **97**, 61-66.

Bos, J.L., Verlaan-de Vries, M., Jansen, A.M., Veeneman, G.H., van Boom, J.H. and van der Eb, A.J. (1984) Three different mutations in codon 61 of the human N-ras gene detected by synthetic oligonucleotide hybridization. *Nucleic Acids Res*, **12**, 9155-9163.

Bourne, H.R., Sanders, D.A. and McCormick, F. (1990) The GTPase superfamily: a conserved switch for diverse cell functions. *Nature*, **348**, 125-132.

Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117-127.

Brouns, S.J., Smits, N., Wu, H., Snijders, A.P., Wright, P.C., de Vos, W.M. and van der Oost, J. (2006) Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon Sulfolobus solfataricus. *J Bacteriol*, **188**, 2392-2399.

Brünger , A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., Read, R.J., Rice, L.M., Simonson, T. and Warren, G.L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*, **54 (Pt 5)**, 905-921.

Buglino, J., Shen, V., Hakimian, P. and Lima, C.D. (2002) Structural and biochemical analysis of the Obg GTP binding protein. *Structure*, **10**, 1581-1592.

Bustelo, X.R., Sauzeau, V. and Berenjeno, I.M. (2007) GTP-binding proteins of the Rho/Rac family: regulation, effectors and functions in vivo. *Bioessays*, **29**, 356-370.

Caldon, C.E. and March, P.E. (2003) Function of the universally conserved bacterial GTPases. *Curr Opin Microbiol*, **6**, 135-139.

Chen, X., Court, D.L. and Ji, X. (1999) Crystal structure of ERA: a GTPase-dependent cell cycle regulator containing an RNA binding motif. *Proc Natl Acad Sci U S A*, **96**, 8396-8401.

Delagoutte, B., Moras, D. and Cavarelli, J. (2000) tRNA aminoacylation by arginyl-tRNA synthetase: induced conformations during substrates binding. *Embo J*, **19**, 5599-5610.

Eathiraj, S., Pan, X., Ritacco, C. and Lambright, D.G. (2005) Structural basis of family-wide Rab GTPase recognition by rabenosyn-5. *Nature*, **436**, 415-419.

Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D*, **60**, 2126-2132.

Esue, O., Cordero, M., Wirtz, D. and Tseng, Y. (2005) The assembly of MreB, a prokaryotic homolog of actin. *J Biol Chem*, **280**, 2628-2635.

Ettema, T.J., de Vos, W.M. and van der Oost, J. (2005) Discovering novel biology by in silico archaeology. *Nat Rev Microbiol*, **3**, 859-869.

Grosshans, B.L., Ortiz, D. and Novick, P. (2006) Rabs and their effectors: achieving specificity in membrane traffic. *Proc Natl Acad Sci U S A*, **103**, 11821-11827.

Hasenohrl, D., Benelli, D., Barbazza, A., Londei, P. and Blasi, U. (2006) Sulfolobus solfataricus translation initiation factor 1 stimulates translation initiation complex formation. *Rna*, **12**, 674-682.

Hendrick, J.L., Wilson, P.G., Edelman, II, Sandbaken, M.G., Ursic, D. and Culbertson, M.R. (2001) Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis. *Genetics*, **157**, 1141-1158.

Karbstein, K. (2007) The role of GTPases in ribosome assembly. *Biopolymers*.

Krutchinsky, A.N., Chernushevich, I.V., Spicer, V.L., Ens, W. and Standing, K.G. (1998) Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry*, **9**, 569-579.

Kukimoto-Niino, M., Murayama, K., Inoue, M., Terada, T., Tame, J.R., Kuramitsu, S., Shirouzu, M. and Yokoyama, S. (2004) Crystal structure of the GTP-binding protein Obg from Thermus thermophilus HB8. *J Mol Biol*, **337**, 761-770.

La Fortelle, E.d. and Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology* **276**, 472-494.

Laskowski, R.A., MacArthur, M.W., Moss, D.S. and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, **26**, 283-291.

Lau, Y.F. and Zhang, J. (2000) Expression analysis of thirty one Y chromosome genes in human prostate cancer. *Mol Carcinog*, **27**, 308-321.

Leipe, D.D., Koonin, E.V. and Aravind, L. (2003) Evolution and classification of P-loop kinases and related proteins. *J Mol Biol*, **333**, 781-815.

Londei, P., Teixido, J., Acca, M., Cammarano, P. and Amils, R. (1986) Total reconstitution of active large ribosomal subunits of the thermoacidophilic archaebacterium Sulfolobus solfataricus. *Nucleic Acids Res*, **14**, 2269-2285.

Matthews, B.W. (1968) Solvent content of protein crystals. *Journal of Molecular Biology*, **33**, 491-497.

Milburn, M.V., Tong, L., deVos, A.M., Brunger, A., Yamaizumi, Z., Nishimura, S. and Kim, S.H. (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*, **247**, 939-945.

Moller, T., Franch, T., Hojrup, P., Keene, D.R., Bachinger, H.P., Brennan, R.G. and Valentin-Hansen, P. (2002) Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Mol Cell*, **9**, 23-30.

Muench, S.P., Xu, L., Sedelnikova, S.E. and Rice, D.W. (2006) The essential GTPase YphC displays a major domain rearrangement associated with nucleotide binding. *Proc Natl Acad Sci U S A*, **103**, 12359-12364.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method. *Acta Crystallographica Section D*, **53**, 240-255.

Nilsson, J. and Nissen, P. (2005) Elongation factors on the ribosome. *Curr Opin Struct Biol*, **15**, 349-354.

Noble, J.A., Innis, M.A., Koonin, E.V., Rudd, K.E., Banuett, F. and Herskowitz, I. (1993) The Escherichia coli hflA locus encodes a putative GTP-binding protein and two membrane proteins, one of which contains a protease-like domain. *Proc Natl Acad Sci U S A*, **90**, 10866-10870.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode. *Methods in Enzymology*, **276**, 307-326.

Pai, E.F., Krengel, U., Petsko, G.A., Goody, R.S., Kabsch, W. and Wittinghofer, A. (1990) Refined crystal structure of the triphosphate conformation of H-ras p21 at 1.35 A resolution: implications for the mechanism of GTP hydrolysis. *Embo J*, **9**, 2351-2359.

Pan, X., Eathiraj, S., Munson, M. and Lambright, D.G. (2006) TBC-domain GAPs for Rab GTPases accelerate GTP hydrolysis by a dual-finger mechanism. *Nature*, **442**, 303-306.

Perrakis, A., Morris, R. and Lamzin, V.S. (1999) Automated protein model building combined with iterative structure refinement. *Nat Struct Mol Biol*, **6**, 458-463.

Sabina, J. and Soll, D. (2006) The RNA-binding PUA domain of archaeal tRNA-guanine transglycosylase is not required for archaeosine formation. *J Biol Chem*, **281**, 6993-7001.

Scrima, A., Vetter, I.R., Armengod, M.E. and Wittinghofer, A. (2005) The structure of the TrmE GTP-binding protein and its implications for tRNA modification. *Embo J*, **24**, 23-33.

Tahallah, N., Pinkse, M., Maier, C.S. and Heck, A.J. (2001) The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun Mass Spectrom*, **15**, 596-601.

Takemaru, K., Li, F.Q., Ueda, H. and Hirose, S. (1997) Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. *Proc Natl Acad Sci U S A*, **94**, 7251-7256.

Urbonavicius, J., Qian, Q., Durand, J.M., Hagervall, T.G. and Bjork, G.R. (2001) Improvement of reading frame maintenance is a common function for several tRNA modifications. *Embo J*, **20**, 4863-4873.

Vetter, I.R. and Wittinghofer, A. (2001) The guanine nucleotide-binding switch in three dimensions. *Science*, **294**, 1299-1304.

Worthington, P., Blum, P., Perez-Pomares, F. and Elthon, T. (2003) Large-scale cultivation of acidophilic hyperthermophiles for recovery of secreted proteins. *Appl Environ Microbiol*, **69**, 252-257.

Wu, H., Sun, L., Brouns, S.J., Fu, S., Akerboom, J., Li, X. and van der Oost, J. (2007) Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon Sulfolobus solfataricus. *Acta Crystallograph Sect F Struct Biol Cryst Commun*, **63**, 239-241.

Yuan, Y.R., Pei, Y., Ma, J.B., Kuryavyi, V., Zhadina, M., Meister, G., Chen, H.Y., Dauter, Z., Tuschl, T. and Patel, D.J. (2005) Crystal structure of A. aeolicus argonaute, a site-specific DNA-guided endoribonuclease, provides insights into RISC-mediated mRNA cleavage. *Mol Cell*, **19**, 405-419.

Zhao, J., Leung, H.E. and Winkler, M.E. (2001) The miaA mutator phenotype of Escherichia coli K-12 requires recombination functions. *J Bacteriol*, **183**, 1796-1800.

Supplementary Tables and Figures

Table S1. Strains, plasmids and primers used in this study.

Strains	**Entry code / genotype**	**Reference**
S. solfataricus P2	DSM1617	Zillig et al. 1980
E. coli HB101	F_ *hsdS*20 (r_B^- m_B^-) *ara-14 galK2 lacY1 leuB6 mcrB mtl-1 proA2 recA13 rpsL20 supE44 thi-1 xyl-5* (Str^r)	Boyer et al. 1969
E. coli BL21(DE3)-pRIL	*hsdS gal* (λc*Its*857 *ind1Sam7 nin5 lacUV5*-T7 gene *1*)	Novagen
Plasmids	**description**	
pET24d/ pET26b	pET series T7 RNA polymerase expression system	Novagen
pWUR298	*ssmbf* (Sso0270) in pET26d (NdeI-XhoI) fusion His-tag	This study
pWUR299	*ssmbf* (Sso0270) in pET26d (NdeI-XhoI)	This study
pWUR300	*ssmbf-c* (ΔSso0270) in pET24d (NdeI-XhoI) C-term. His-tag	This study
pWUR301	*ssmbf-c* (ΔSso0270) in pET24d (NdeI-XhoI)	This study
pWUR335	*ssgbp* (Sso0269) in pET24d (BspHI-XhoI) C-term. His-tag	(Wu et al., 2007)
pWUR370	*ssmbf -ssgbp* in pET24d (BspHI-XhoI) C-term. His-tag	This study
pWUR371	*ssmbf -ssgbp* in pET24d (BspHI-XhoI)	This study
pWUR372	*ssgbp-n* (Sso0269) in pET24d (BspHI-XhoI) C-term His-tag	This study
pWUR373	*ssgbp-n* (Sso0269) in pET24d (BspHI-XhoI)	This study
pWUR374	*ssgbp-g* (Sso0269) in pET24d (BspHI-XhoI) C-term His-tag	This study
pWUR375	*ssgbp-g* (Sso0269) in pET24d (BspHI-XhoI)	This study
Primers	**sequence (5'→3')**	**Enzyme site**
ssmbf fw (Sso0270)	GCGCGCATATGCAAGCTAATAGTGAAGAATAC	NdeI
ssmbf rv (Sso0270)	GCGCGCTCGAGCTTCTTTCCCTCTTTAATATTTACC	XhoI
ssmbf rv (Sso0270)	GCGCGCTCGAGTCACTTCTTTCCCTCTTTAATATTTACC	XhoI
ssmbf-c fw (ΔSso0270)	GCGCGGCCCATATGCGTAAGAAAGCCACTCTTAAACCACC	NdeI
ssmbf-c rv (ΔSso0270)	GCGCGCTCGAGCTTCTTTCCCTCTTTAATATTTACC	XhoI
ssmbf-c rv (ΔSso0270)	GCGCGCTCGAGTCACTTCTTTCCCTCTTTAATATTTACC	XhoI
ssgbp fw (Sso0269)	GCGCGCTCATGAAAACAGCTGCTCTTTTTGTATC	BspHI
ssgbp rv (Sso0269)	CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG	XhoI
ssmbf-ssgbp fw (Sso0270-0269)	GCGCGTCATGATGCAAGCTAATAGTGAAGAATAC	BspHI
ssmbf–ssgbp rv(Sso0270-0269)	CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG	XhoI
ssmbf-ssgbp rv(Sso0270-0269)	GCGCGCTCGAGTTAACTCAACTGAGTTGCTAG	XhoI
ssgbp-n fw(sso0269)	GCGCGCTCATGAAAACAGCTGCTCTTTTTGTATC	BspHI
ssgbp-n rv(Sso0269)	GCGCGCTCGAGCTTATTAGATTCTATGGATTTTTC	XhoI
ssgbp-n rv(Sso0269)	GCGCGCTCGAGTTACTTATTAGATTCTATGGATTTTTC	XhoI
ssgbp-g fw(sso0269)	GCGCGCTCATGAGAAATAATATTCCTTCTATCGG	BspHI
ssgbp-g rv(Sso0269)	CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG	XhoI
ssgbp-g rv(Sso0269)	CGCGCCTCGAGTTAACTCAACTGAGTTGCTAGCTGG	XhoI

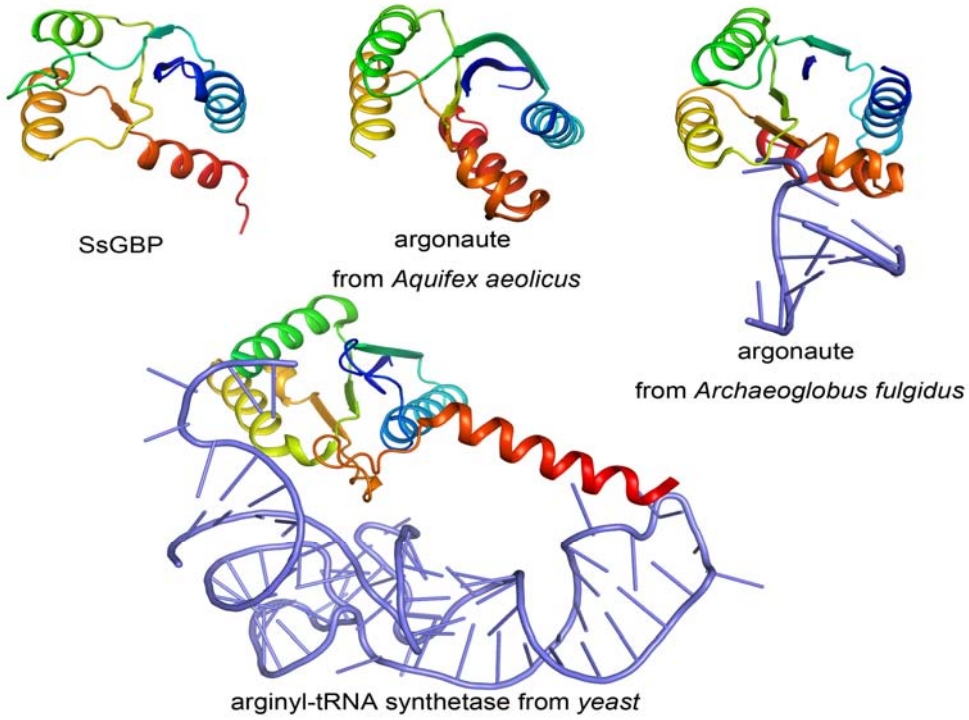


Figure S1. Structural comparison of the SsGBP-N subdomain I. Dali search revealed distant similarity to α/β domains of distinct proteins. The best hits includes the RNA-binding domain of Argonaute from *Archaeoglobus fulgidus* (PDB ID: 2BGG; Z-score: 5.5; RMSD: 2.8) and, Aquifex aeolicus (PDB ID: 1YVU; Z-score: 6.1; RMSD:3.3), and yeast arginyl-tRNA synthetase (PDB ID: 1F7U. Z-scroe: 6.0; RMSD: 3.1).

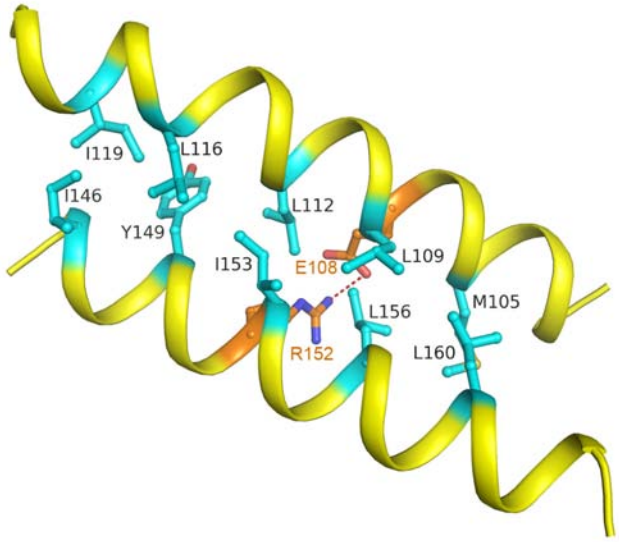


Figure S2. Structural details of the SsGBP-N subdomain II. The two helices form a coiled–coil structure, with a large number of hydrophobic residues between the two helical coils. The hydrophobic residues (cyan) and polar residues (orange) involved in the interaction are showed in a ball-and-stick representation. The salt bridge between E108 and R152 is represented as dashed line.

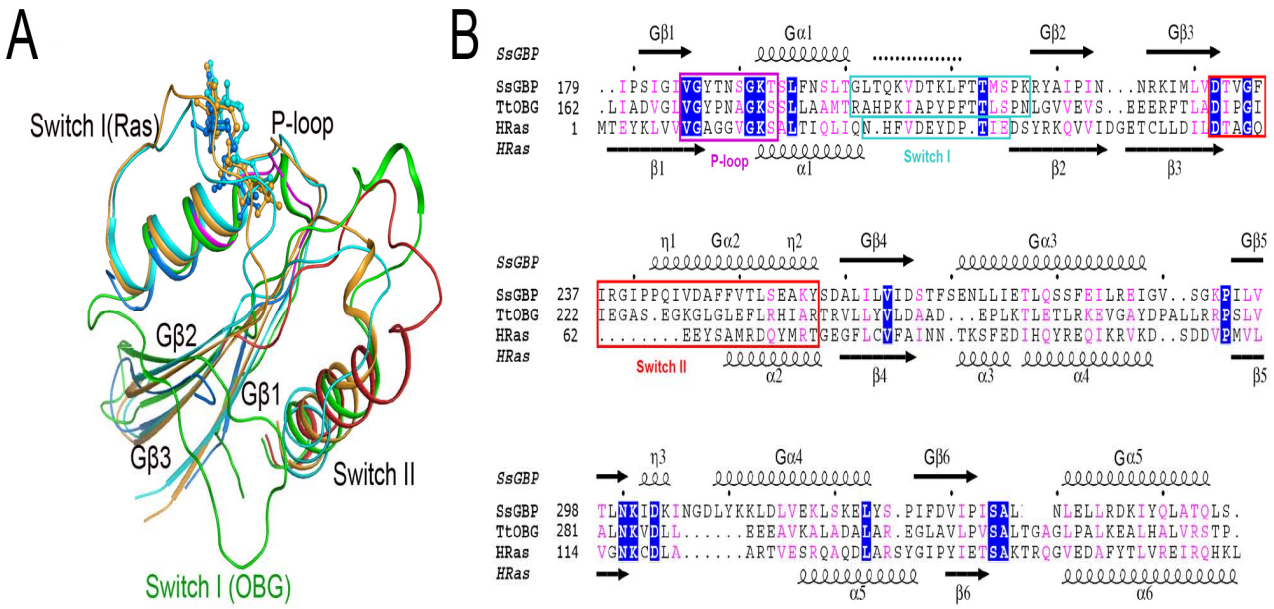


Figure S3. Comparison of the G domains from SsGBP, Obg and Ras. (A) Superposition of the G domain of SsGBP (blue) with *Thermus thermophilus* HB8 Obg (green; PDB ID: 1UDX), GDP-bound (cyan; PDB ID: 1QRA) and GTP-bound human Ras (orange; PDB ID: 4Q21). Comparison of the P-loop and switch regions. In the G domain of SsGBP, the P-loop region is shown in magenta and switch II region is red. (B) Structure-based sequence alignment of the G domain of GBP, Obg and Ras. Secondary-structure elements are shown for SsGBP and Ras structures, as assigned by the program DSSP. The alignment was generated with ClustalW. Identical and conserved residues are highlighted in blue and colored in purple, respectively. The missing region is indicated as dashed line. The P-loop and switch regions are highlight by box and the conserved motifs (G1-G5) are indicated by green dots.

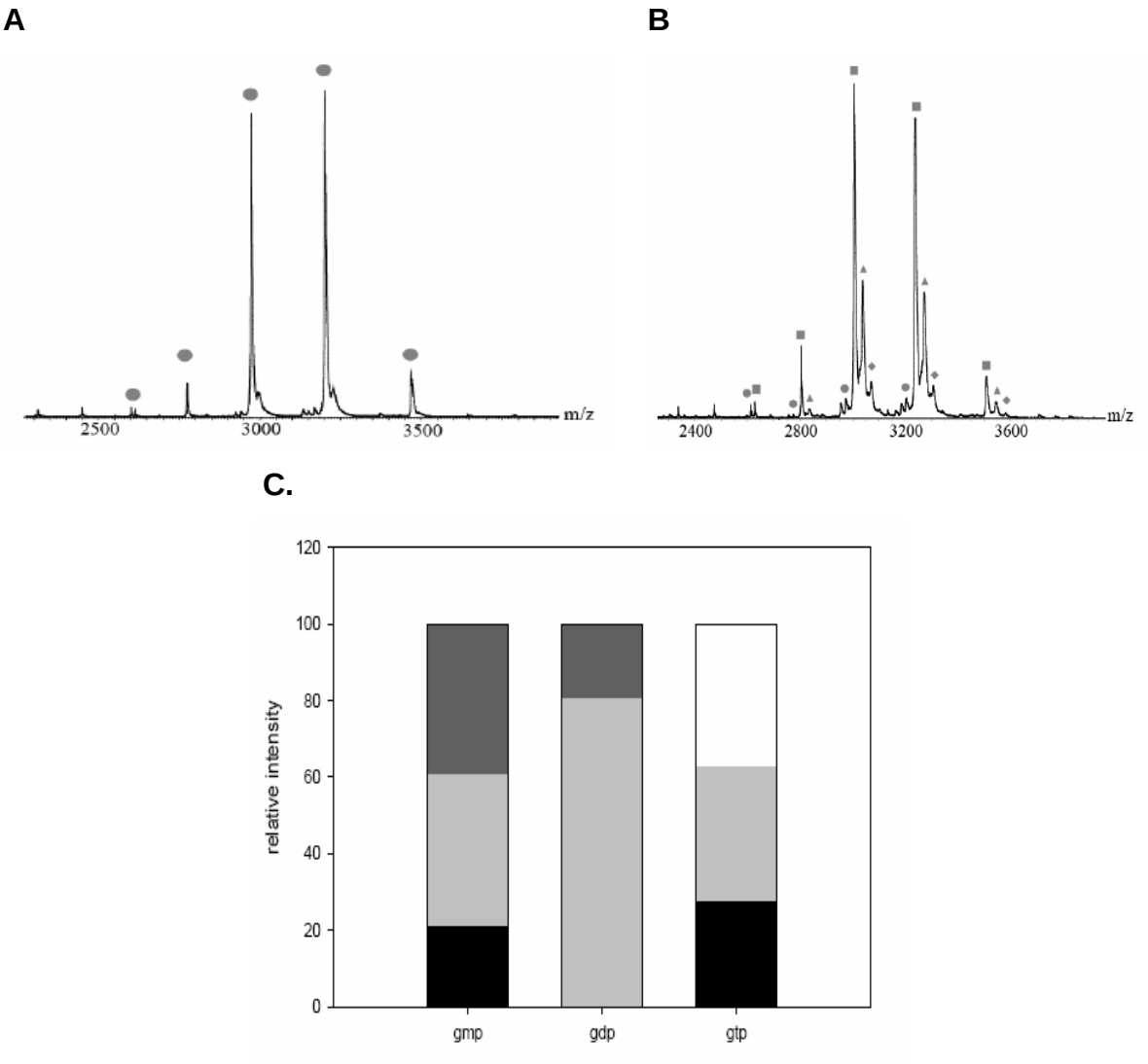
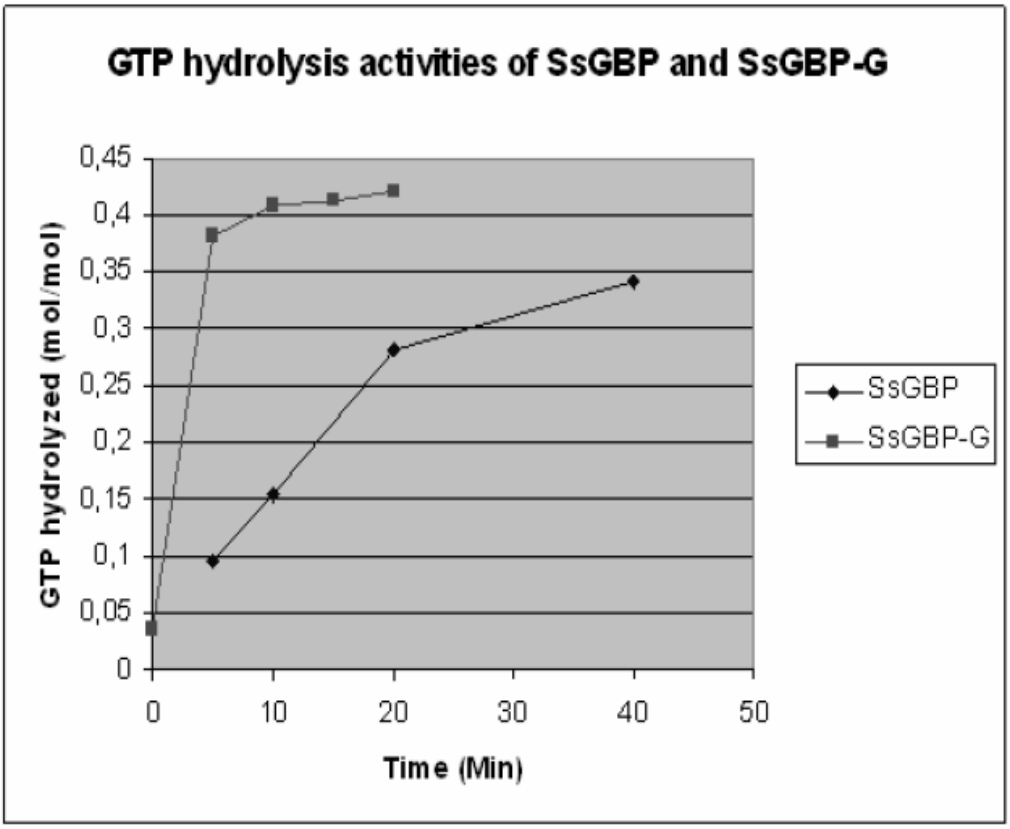


Figure S4. Esi-MS analysis. (A) Spectrum of apo-SsGBP. (B) Spectrum of SsGBP with GDP in the presence of Magnesium (circle – SsGBP; square - SsGBP plus one GDP; triangle - SsGBP two GDP; diamond - SsGBP three GDP). (C) Bar diagram of the binding behavior of the different nucleotides GMP GDP and GTP to SsGBP in the presence of Magnesium. The black fields represent the apo SsGBP, the light gray fields the holo SsGBP with one nucleotide bound and the dark grey fields the holo SsGBP with two nucleotides bound. The white field in the bar of the GTP represents the relative amount of GDP bound to SsGBP. For all measurements there was a low amount of SsGBP present where three nucleotides where bound to the protein. These were below 5% and did not change in relation to the other forms present with addition of Magnesium (data not shown). They were therefore regarded as unspecific and are not considered in this calculation.

A.



B.

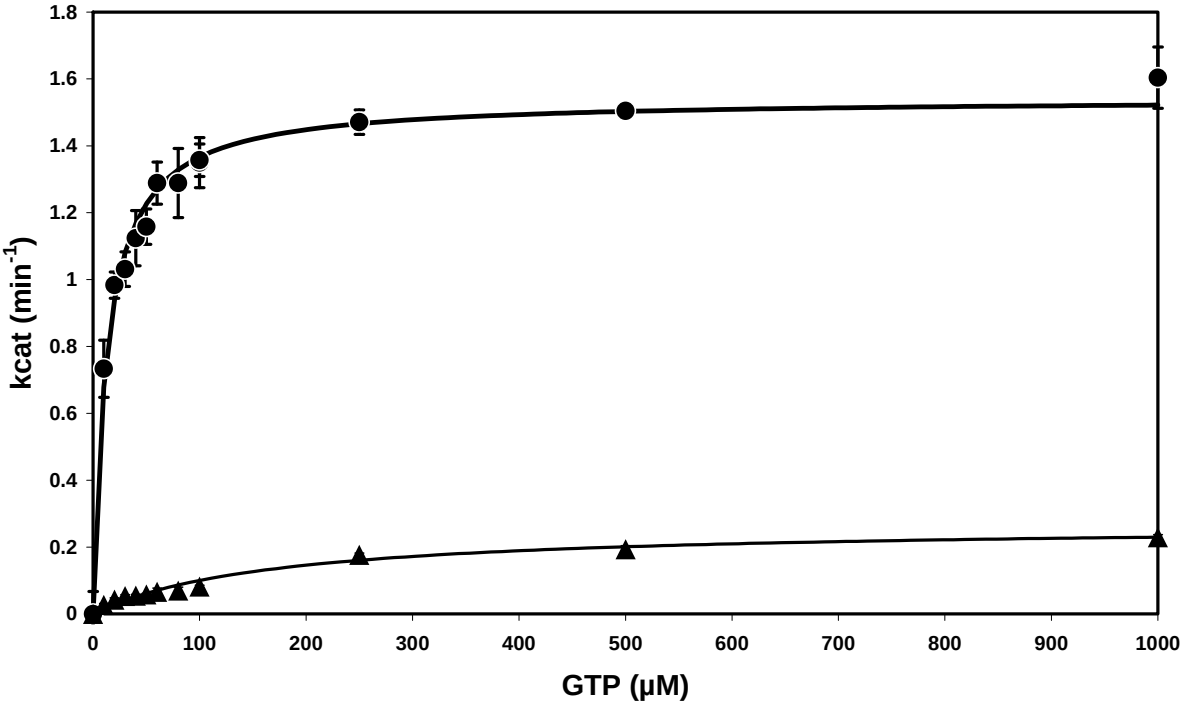


Figure S5. GTPase activity of SsGBP and SsGBP-G. A. The time-course of GTP hydrolysis of SsGBP and SsGBP-G based on TLC analysis of SsGBP and SsGBP-G. Negative control was SsGBP-N (data not shown). **B.** Substrate concentration dependent activity at 50^0C. Closed circle: SsGBP-G; Closed triangle: full-length SsGBP.

Chapter 6

Molecular characterization of the eukaryotic-like Multi-protein bridging factor (MBF) from thermophilic archaeon *Sulfolobus solfataricus*

Hao Wu, Fabian Blombach, Bart de Koning, Lei Sun , Ambrosius P. L. Snijders , Thijs J. G. Ettema, Stan J.J. Brouns, Willem M. de Vos , and John van der Oost

Manuscript in preparation

ABSTRACT

Multi-protein bridging factor (MBF) is highly conserved in archaea and eukarya, but absent in bacteria. In archaea, MBF shows a conserved genomic context with genes encoding a proteasome-activating nucleotidase (PAN), a GTP-binding protein (GBP), a basic transcription factor (TFE), and proteins involved in RNA binding/modification. In the hyperthermophilic archaeon *Sulfolobus solfataricus,* an overlap is observed for the stop-codon of the *mbf* gene and the start-codon of the *gbp* gene that encodes an HflX-like GTPase; the anticipated operon structure has been confirmed by Northern blot analysis. We have cloned and overexpressed the gene encoding the full-length Mbf protein and a C-terminal Mbf-fragment (Mbf-C) from *S. solfataricus.* Both Mbf and Mbf-C could be produced in soluble form. Analytical ultracentrifugation analysis and native mass spectrometry revealed that Mbf-C is a monomer in solution. Furthermore, we demonstrate that the heterologously produced Mbf-C is co-purified with host DNA. We cloned the *mbf-gbp* operon from *S. solfataricus* overexpressed in *E.coli*. Interestingly, co-expression of the two genes in *E.coli* resulted in an apparent stable complex. Molecular analyses indicate that MBF might interact with GBP. A possible role of both proteins in transcription and/or translation is discussed.

INTRODUCTION

Archaea and bacteria share many so-called operational genes, the products of which are involved in cellular metabolic processes (Rivera *et al.*, 1998; Rivera and Lake, 2004). However, many proteins involved in information-processing pathways of archaea (e.g., replication, transcription, and translation) most closely resemble their counterparts in eukaryotes. (Bell and Jackson, 1998; Karlin *et al.*, 2005; Keeling and Doolittle, 1995; White, 2003). One of these proteins is multi-protein bridging factor 1 (MBF), which is highly conserved in archaea and eukarya, but absent in bacteria (Makarova *et al.*, 1999; Makarova and Koonin, 2003).

Several *in vitro* studies have led to the conclusion that the eukaryal MBF functions as a co-activator that mediates transcriptional activation by bridging TATA-box binding

protein (TBP) and gene-specific activators (Jindra *et al.*, 2004; Li *et al.*, 1994; Liu *et al.*, 2003; Liu *et al.*, 2007; Takemaru *et al.*, 1998; Takemaru *et al.*, 1997). These gene-specific activators include the *Drosophila* nuclear receptor FTZ-F1(Li *et al.*, 1994; Liu *et al.*, 2003) and the yeast basic leucine zipper protein GCN4 (Takemaru *et al.*, 1998). In human, MBF is thought to function as a cofactor for endothelial differentiation that might regulate lipid metabolism (Brendel *et al.*, 2002; Kabe *et al.*, 1999). Both eukaryal and archaeal MBF comprise a C-terminal helix-turn-helix motif (Aravind *et al.*, 2005; Aravind and Koonin, 1999). Recently, an *in vivo* study has indicated that MBF in yeast contributes to translation fidelity (Hendrick *et al.,* 2001).

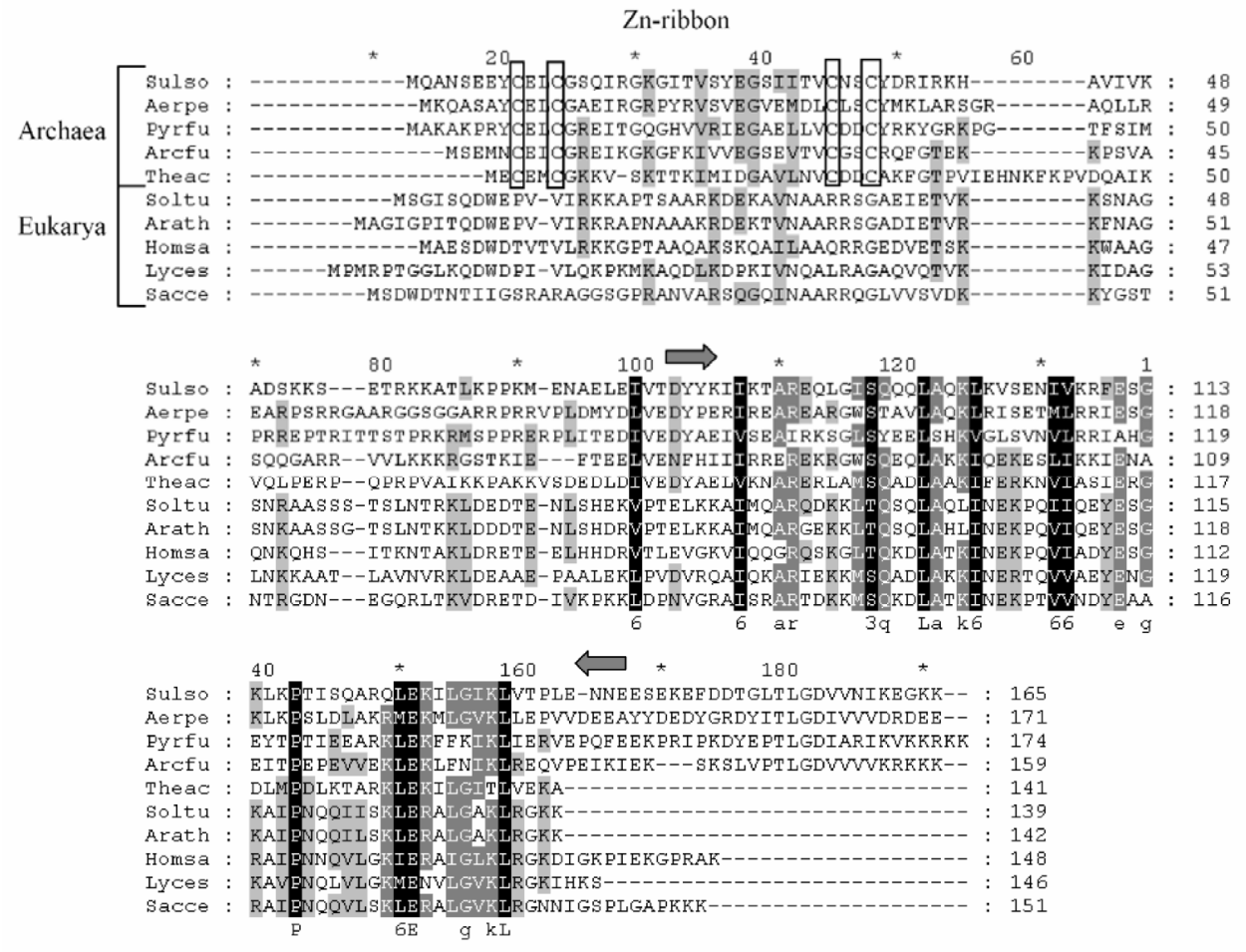


Figure 1. Alignment of MBF from Archaea and Eukarya. Arrows indicate boundaries of MBF-C. Boxes indicate conserved cysteines that constitute a potential Zn-binding site in the archaeal orthologs, but not in the eukaryotic ones. Sulso*: Sulfolobus solfarticus*; Aerpe*: Aeropyrum pernix;* Pyrfu*: Pyrococcus furiosus;* Theac*: Thermoplasma acidophilum;* Arath: *Archaeloglobus fulgidus*; Lyces*: Lycopersion esculentum;* Soltu: *Solanum tuberosum;* Homsa: *Homo sapiens;* Arath*: Arabidopsis thaliana;* Sacce: *Saccharomyces cerevisiae.*

The N-terminal domain of archaeal MBF contains a Zn-ribbon, whereas a distinct N-terminal domain is present in the eukaryal MBF (Fig. 1). In Crenarchaea, such as in the hyperthermophilic *Sulfolobus solfataricus*, *mbf* is located upstream of the *gbp* gene that encodes for a HlfX-like GTPase (Leipe, 2002; Wu *et al.* 2007). To reveal the function of the archaeal MBF and study the mechanism of its interaction with GBP, we cloned and expressed in *Escherichia coli* the *S.solfataricus* full-length *mbf* , the 3' fragment of *mbf* (MBF-C, residues R57–K165) and the *mbf-gbp* operon, after which biochemical and genetic analyses were performed. The involvement of MBF and GBP in transcription and/or translation is discussed.

MATERIALS & METHODS

Chemicals and enzymes

Most chemicals used in this study were obtained from Sigma and were of analytical grade. Oligonucleotide primers were obtained from MWG Biotech AG (Ebersberg, Germany). Restriction enzymes were from New England Biolabs. Polymerase chain reactions were performed with *Pfu* TURBO (Stratagene). The usual DNA manipulations were done according to standard procedures. PCR fragments were extracted from agarose gel by means of QIAquick gel extraction Kit (Qiagen) and plasmid DNA was purified with QIAprep Miniprep Kit (Qiagen).

Strains, plasmids, and DNA manipulation

The *E. coli* strain HB101 was used for cloning and *E. coli* BL21(DE3) (Novagen) containing tRNA accessory plasmid pRIL (Stratagene) was used for protein overproduction. Vectors of pET24d and pET26b (Novagen) containing a T7 promoter and $(His)_6$ tag were used for cloning and expression.

Table 1. Strains, plasmids, and primers used in this study

Strains	**Entry code / genotype**	**Reference**
S. solfataricus P2	DSM1617	Zillig et al. 1980
E. coli HB101	F- *hsdS*20 (r_B^- m_B^-) *ara-14 galK2 lacY1 leuB6 mcrB mtl-1 proA2 recA13 rpsL20 supE44 thi-1 xyl-5* (Str^r)	Boyer et al. 1969
E. coli BL21(DE3)-pRIL	*hsdS gal* (λcI*ts*857 *ind1Sam7 nin5 lacUV5*-T7 gene *1*)	Novagen
Plasmids	**description**	
pET24d/ pET26b	pET series T7 RNA polymerase expression system	Novagen
pWUR298	*mbf* (Sso0270) in pET26d (NdeI-XhoI) fusion His-tag	This study
pWUR299	*mbf* (Sso0270) in pET26d (NdeI-XhoI)	This study
pWUR300	*mbf-c* (ΔSso0270) in pET24d (NdeI-XhoI) C-term. His-tag	This study
pWUR301	*mbf-c* (ΔSso0270) in pET24d (NdeI-XhoI)	This study
pWUR335	*gbp* (Sso0269) in pET24d (BspHI-XhoI) C-term. His-tag	(Wu et al., 2007)
pWUR370	*mbf-gbp* in pET24d (BspHI-XhoI) C-term. His-tag	This study
pWUR371	*mbf-gbp* in pET24d (BspHI-XhoI)	This study
Primers	**sequence (5'→3')**	**Enzyme site**
mbf fw (Sso0270)	GCGCGCATATGCAAGCTAATAGTGAAGAATAC	NdeI
mbf rv (Sso0270)	GCGCGCTCGAGCTTCTTTCCCTCTTTAATATTTACC	XhoI
mbf rv (Sso0270)	GCGCGCTCGAGTCACTTCTTTCCCTCTTTAATATTTACC	XhoI
mbf-c fw (ΔSso0270)	GCGCGGCCCATATGCGTAAGAAAGCCACTCTTAAACCACC	NdeI
mbf-c rv (ΔSso0270)	GCGCGCTCGAGCTTCTTTCCCTCTTTAATATTTACC	XhoI
mbf-c rv (ΔSso0270)	GCGCGCTCGAGTCACTTCTTTCCCTCTTTAATATTTACC	XhoI
gbp fw (Sso0269)	GCGCGCTCATGAAAACAGCTGCTCTTTTTGTATC	BspHI
gbp rv (Sso0269)	CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG	XhoI
mbf -gbp fw (Sso0270-0269)	GCGCGTCATGATGCAAGCTAATAGTGAAGAATAC	BspHI
mbf -gbp rv (Sso0270-0269)	CGCGCCTCGAGACTCAACTGAGTTGCTAGCTGG	XhoI
mbf -gbp rv (Sso0270-0269)	GCGCGCTCGAGTTAACTCAACTGAGTTGCTAG	XhoI

Cloning and overexpression

The synthetic oligonucleotide primers containing specific restriction enzyme sites (*e.g.* NdeI, BspHI, and XhoI) are listed in Table 1. They were used for PCR amplification of the genes encoding the full-length MBF (Sso0270, gene ID 15897213), the C-terminal fragment of MBF (MBF-C), and MBF-gbp (Sso0270 and Sso0269, (Wu et al., 2007)) were PCR-amplified from *S. solfataricus* genomic DNA (Table 1). The PCR products were electrophoretically separated on a 1.0% agarose gel, Qiaex purified, digested with the appropriate restriction enzymes, and ligated in either pET24d or pET26b (Table1). All generated constructs were verified by DNA sequencing to confirm the correct nucleotide sequences. Overexpression of the recombination constructs was achieved by using *E. coli* BL21(DE3)-pRIL as host, freshly transformed with either pWUR298, pWUR299, pWUR300, pWUR301, pWUR335, pWUR370 or pWUR371 (Table 1). Transformed colonies were grown in a rotary shaker at 37^0C, in LB medium in baffled flasks. They were induced with 1mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) when A_{595nm} reached 0.5 (flasks). After 5h of induction, cells were harvested by centrifugation (5000 rpm, 30 min, 4 ^{0}C). Cell pellets were frozen in liquid nitrogen and kept at –80 ^{0}C for further purification.

Purification of the recombination protein

The purification of recombinant MBF, MBF-C, GBP, and MBF-GBP were performed from cell pellets equivalent to 1 liter culture. Lysis buffer consisted of 50 mM NaH_2PO_4 buffer (pH 8.0). The pellets were thawed on ice and resuspended in 50 ml of lysis buffer containing 300 mM NaCl and 5% glycerol (v/v)(for $(His)_6$ tag fusion proteins, the buffer includes 10 mM imidazol). The cells were disrupted by sonication, and the obtained extract was centrifuged at 13,200 rpm for 30 min at 4 ^{0}C to remove cellular debris and aggregated proteins. The supernatant was subjected to 20 min heat treatment at 70 ^{0}C, and immediately centrifuged a second time at 13,200 rpm for 30 min at 4 ^{0}C. The following steps were performed at room temperature.

A heparin-sepharose column was applied for MBF and MBF-C purification since

these proteins include a predicted HTH DNA-binding motif. After a heat incubation step, samples of MBF (or MBF-C) were loaded onto the HiTrap™ Heparin column connected to a fast protein liquid chromatography system (Amersham Biosciences) and eluted using a 1 M NaCl gradient. Fractions containing MBF (or MBF-C) were combined and concentrated to a volume of 1.0 ml using an Amicon Ultra-15 centrifugal filter with a 5 kDa molecular weight cut-off (Millipore). The concentrated samples were applied onto a 10/300 GL Superdex 200 column (Amersham Biosciences) equilibrated with either buffer A (20 mM HEPES, pH 7.0 and 150 mM NaCl) or buffer B (20 mM HEPES pH 7.0 ,150 mM NaCl and 1mM DTT).

The Ni-NTA affinity column purification was also used for MBF, MBF-C and MBF-GBP $(His)_6$-tag fusion proteins after the heat incubation. The purification steps were the same as described previously (Wu et al., 2007).

Ultracentrifugation analysis

A Beckman analytical ultracentrifuge (Proteome Lab XL-1) was applied for the analysis of the conformation of MBF-C in solution. The reference wavelength of MBF-C is 276nm, as determined by Proteome Lab XL-I-[cell2-Abs] (Colfen *et al.*, 1998). Then, it was centrifuged at 60,000 rpm (31,4000 x g**)** at 16^0C for 30 h to obtain the sedimentation coefficient of MBF-C. Finally, transformed the sedimentation coefficient distribution to a molar mass distribution of MBF-C by using SEDFIT.

DNase and RNase treatment

To check the purified MBF-C for the presence of nucleic acid, protein samples were analyzed by 1% agarose gel and Ethidium Bromide (EtBr) staining. Protein samples were treated DNaseI or RNase A/T1. The reaction mixtures of DNase treatment contained: 12 µl sample, 1.5 µl DNase I, 0.5 µl DNase buffer 10×, and 1 µl Mili-Q water); that of the RNase treatment contained: 12 µl nucleic acids sample, 1.5 µl RNase A/T1, 0.5 µl RNase buffer 10×, and 1 µl Mili-Q water). Negative control did not contain enzyme or buffer (12 µl nucleic acids sample, 3 µl Mili-Q water). Samples were incubated at 37^0 C for 15 min, and subsequently analyzed on 1% agarose gel/EtBr as described above.

ESI mass spectrometry

Tandem mass spectrometry was carried out to confirm the identity of the recombinant proteins (Snijders et al., 2006). Protein samples were manually excised from the Coomassie-stained 2-DE gels, destained with 200 mM ammonium bicarbonate with 40% Acetonitrile (ACN). The gel slices were incubated overnight in 50 mL of 40 mM ammonium bicarbonate in 9% ACN, with 0.4 mg trypsin (Sigma). After that, peptides were extracted in three subsequent extraction steps using 5 mL of 25 mM NH_4HCO_3 (10 min, room temperature), 30 mL (15 min, 37^0C), 50 mL of 5% formic acid (15 min, 37^0C) and finally with 30 mL ACN (15 min, 37^0C). All extracts were pooled and dried in a vacuum centrifuge, then stored at -20^0C. The lyophilized peptide mixture was resuspended in 0.1% formic acid in 3% ACN. This mixture was separated on a PepMap C-18 RP capillary column (LC Packings, Amsterdam, The Netherlands) and eluted in a 30-min gradient *via* an LC Packings Ultimate nanoLC directly onto the mass spectrometer. Peptides were analysed using an Applied Biosystems QStarXL® ESI quadrupole TOF tandem mass spectrometer (ESI qQ-TOF). The data acquisition on the MS was performed in the positive ion mode using information dependent acquisition (IDA). Peptides with charge states 2 and 3 were selected for MS/MS. IDA data were submitted to MASCOT for database searching in a MS/MS ion search type of search (www.matrixscience.com). The peptide tolerance was set to 1.2 Da and the MS/MS tolerance was set to 0.6 Da. Methionine oxidation was set as a variable modification. Up to one missed cleavage site by trypsin was allowed. The search was performed against the Swissprot database. His-tagged N- and C- terminal sequences were added manually to the database.

Western blotting

Western blotting was applied to detect MBF and GBP. Since the $(His)_6$ tag was fused to the C-terminus of GBP, the anti-(His)6 antibody was used to detect the bands which include the $(His)_6$ tag. SDS-PAGE gels were blotted on a nylon filter, using a minigel Apparatus (Mini-gel blot). The anti-MBF antibody was used to identify the peptides included MBF. Buffers include: TBS (20 mM Tris-HCl (pH 8.0), 150 mM NaCl)*;* TTBS

(0.1% Tween20 in TBS); Blocking buffer (3% BSA in TBS)*;* AP buffer (100 mM Tris/HCl, pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$); BCIP solution (50 mg/ml in dimethylformamide); NBT solution (50 mg/ml in 70% dimethylformamide).

Blotted filters were incubated in 20 ml blocking buffer at room temperature for 30 min while gently shaking. Primary antibodies were diluted (1:3000; rabbit anti-MBF); 1:1000; mouse anti-$(His)_6$ (Promega, Benelux BV) in 20 ml blocking buffer, and incubated for 30 min. Then filters were washed three times for 15 min in 100 ml TTBS. After that, blots were washed in 100 ml TBS for 5 min. In case of the anti-MBF antibody, secondary antibody (anti-rabbit IgG-alkalic phosphatase conjugate (Promega)) was diluted 1:6000 in blocking buffer; in case of the anti-$(His)_6$, the secondary antibody (anti-mouse AP conjugate) was diluted 1: 10000. Blots were incubated in 20 ml diluted secondary antibody for 30 min, washed three times for 15 min in 100 ml TTBS, and washed once in 100 ml TBS for 5 min. diluted. Then, blot was transferred to a Petri dish and added AP solution (16.7 µl BCIP and 33.3 µl NBT solution in 5 ml AP buffer) was applied on the blot without shaking in the dark. When signal intensity was sufficient, reaction was stopped by washing two times with MilliQ. Blots were scanned immediately with GS800 densitometer.

Native ESI MS

Native ESI-MS was used to detect the possible interaction of MBF and GBP. The sample buffer was exchanged sequentially to 50 mM aqueous ammonium acetate pH 6.8 using centrifugal filter units with a cut-off of 5 kDa (Millipore, England). The concentration used for the mass spectrometry measurements was 10 µM. The samples were measured on a LCT electronspray time-of-flight instrument (Waters, Manchester, UK). Nanospray glass capillaries were used to introduce the samples into the Z-spray source. The source pressure was increased to 10 mbar to create increased collisional cooling (Krutchinsky et al., 1998; Tahallah et al., 2001). Source temperature was set to 80°C and sample cone voltage varied from 80 to 125 V. Needle voltage was around 1300 V.

Crystallization trials

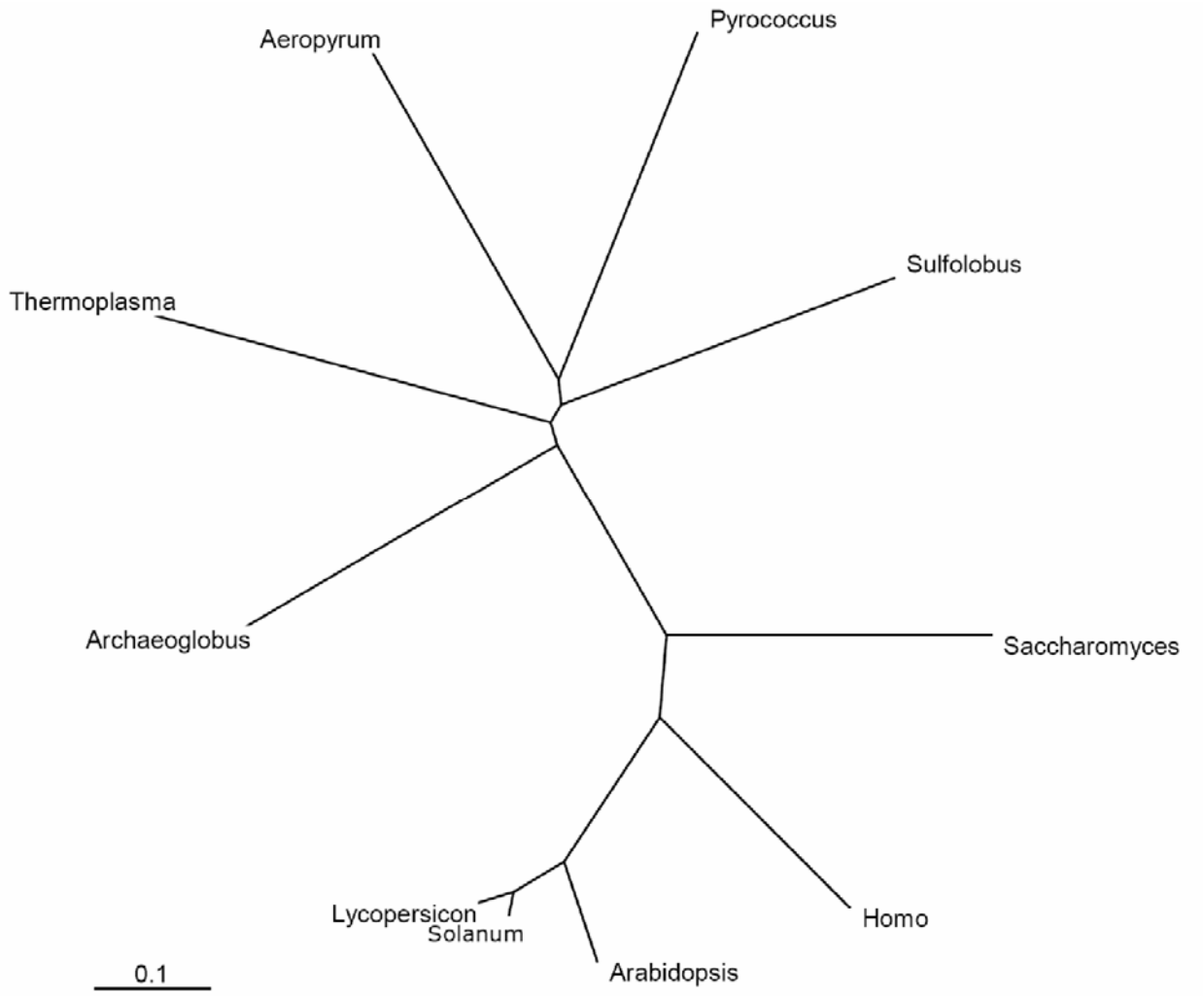
A complete reagent kit (Hampton Research) has been designed to provide a highly effective and rapid screening method for the crystallization of protein macromolecules. In total five series of crystal screens were performed for the crystallization for MBF-C as well as for the complex of MBF-C and GBP. They were Crystal screen 1 (50 conditions), Crystal screen 2 (48 conditions), Index & Index HT (96 conditions), PEG/Ion (48 conditions) and MembFac & MembFac (Membrane protein sparse matrix screen, 48 conditions). After preliminary screens, the optimization conditions were selected to focus on for further refining experiment. In order to refine the crystallization conditions, a gradient of protein concentrations was applied first. Secondly, crystallization conditions were refined by variation gradient of precipitants (e.g. ethanol, PEG series etc.). Thirdly, we changed the pH and temperature. Only a single parameter was changed at a time.

RESULTS & DISCUSSION

Phylogenetic tree of MBF

The eukaryotic MBF1 has been characterized as a transcriptional co-activator that plays a crucial role in the expression of certain genes by communicating between specific regulatory factors and the basal transcription machinery (Liu *et al.*, 2007; Takemaru et al., 1997). In agreement with this, a recent analysis has reported that MBF co-evolved with TBP among all organisms in which TBP is used as the general transcription factor (archaea and eukaryotes) (Liu *et al.*, 2007). On the other hand, an apparent distinct function of MBF is its proposed involvement in translation fidelity (Hendrick et al., 2001). Analysis of multi-alignment and phylogeny of MBF (Fig. 1, 2), clearly indicates that archaeal MBF are distinct from the eukaryotic MBF1 orthologs, including the counterparts from *S. cerevisiae* and *H. sapiens*. A striking difference concerns the fact that the archaeal MBF has a typical 4-cysteine motif that most likely forms a zinc ribbon structural motif; the eukaryotic orthologs have a distinct N-terminal domain, lacking an obvious metal binding site. (Fig.1)

Figure 2. Phylogenetic tree of MBF from Archaea and Eukarya.



Genomic context of mbf

An *in silico* analysis was preformed of the conserved genomic neighborhood of archaeal *mbf* genes. In a total of 11 crenarchaeal genomes, the *mbf* is always clustered with the *gbp* gene (the only exception is the sponge symbiont *Cenarchaeum symbiosum*), which implies that these two genes might functionally interact with each other (Fig. 3). In addition, context analysis at ERGO (http://ergo.integratedgenomics.com/ERGO/) and STRING (http://string.embl.de/), revealed that the archaeal *mbf* gene is located adjacent to the gene that encodes the proteasome regulatory subunit (PAN) in 11 of the available 19 archaeal genomes, from distinct phylogenetic lineages, including the euryarchaeal *Methanopyrus kandleri*. In the three available *Sulfolobus* genomes, MBF (Sso0270;COG1813 - MBF-homolog; Multiprotein Bridging Factor) appears to be part of

a 20 kb super-operon, including the genes encoding the following proteins: (1) a small nuclear RNP (RNA-binding protein), (2) a tRNA ribosyl transferase, (3) ATP-GRASP ligase, (4) an RNA-binding PUA domain, (5) an AAA+type ATPase, (6) an HflX-like GTPase (GTP-binding protein, GBP), (7) a hypothetical protein, (8) transcription factor E

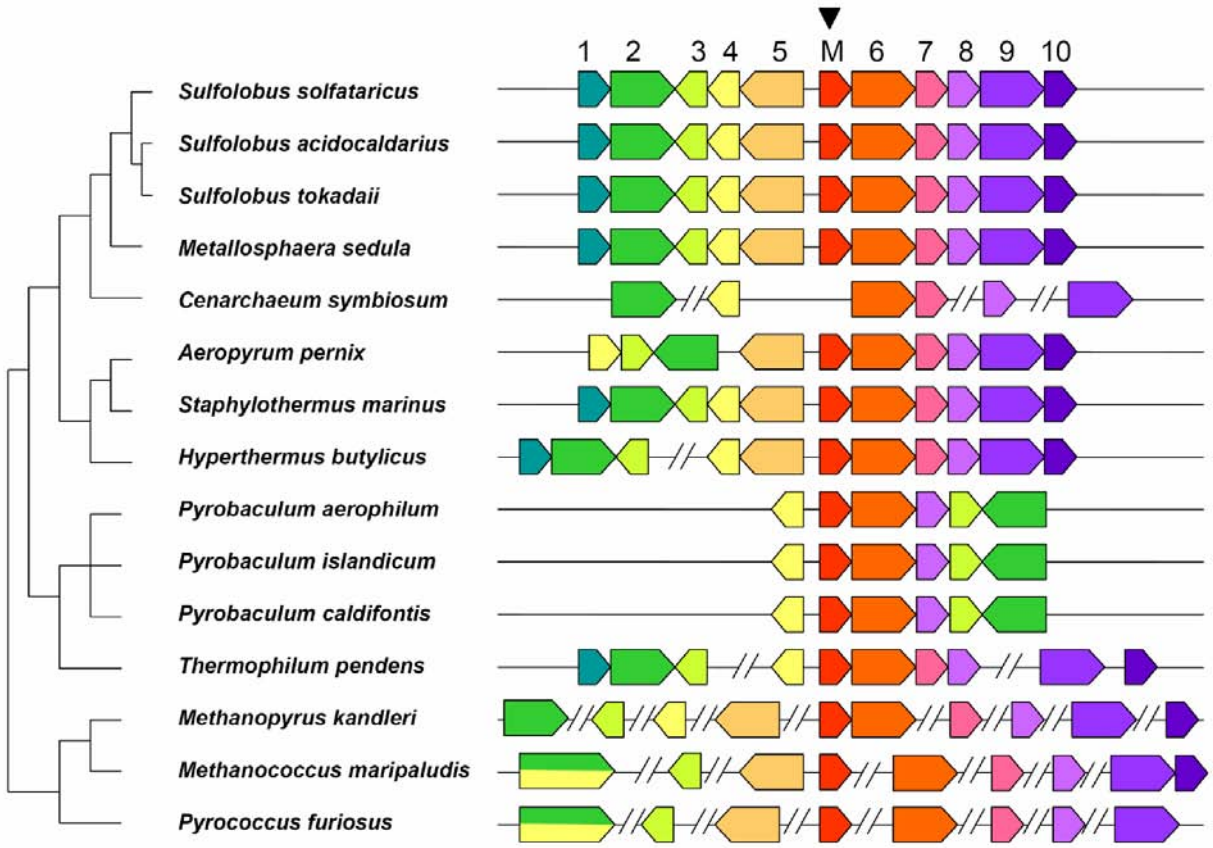


Figure 3. Genomic context of MBF from archaea. MBF (M; Sso0270; COG1813) is indicated with an arrow. Other gene products are (1) Sso0276 (COG1958 - small nuclear RNP (RNA-binding protein)); (2) Sso0274 (COG0343 - tRNA ribosyl transferase (TGT)); (3) Sso0273 (COG1938 - DUF75; ATP-GRASP superfamily; ligase (C/N, C/S-CoA)); (4) Sso5542 (COG1370 - PUA domain, RNA binding; in many Euryarchaea fused to TGT (COG0343)); (5) Sso0271 (COG1222 - AAA+type ATPase; Proteasome Activating Nucleotidase); (6) Sso0269(COG2262 - GTP-binding protein; HflX-like GTPase); (7) Sso0267 (COG1303 - DUF12); (8) Sso0266 (COG1675 - TFE; transcription factor; stabilizes open complex); (9) Sso0265 (COG0270 - Site-specific DNA methylase); (10) Sso0264 (COG4080 – RecB-family nuclease).

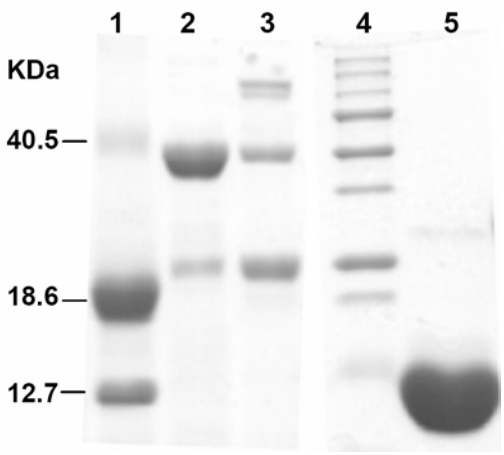


Figure 4. Production and purification on Ni-column of MBF and GBP. Lane 1, MBF-His (and degradation product MBF-C), Lane 2, GBP-His, Lane 3, MBF & GBP-His, Lane 4 MW Marker, Lane 5 MBF-C-His. His-tags include 6 His residues at the C-terminus.

(TFE), (9) a DNA methylase, and (10) a RecB-family nuclease (Fig. 3). Interestingly, in the recently released genome sequence of the Korarchaeum (James Elkins, personal communication), the *mbf* gene is clustered, amongst others, with the genes encoding two transcription factors: BTF3 (Basic Transcription Factor 3) and TFE. Both BTF3 and TFE are, like MBF, highly conserved in archaea and eukaryotes, and absent in bacteria, and both have been reported to be basal transcription factors. However, like MBF, a dual function has been described for BTF3: it is also referred to as the NAC-α subunit, the nascent polypeptide associated complex. This protein appears to contribute to protein synthesis quality control. Hence, based on conserved genomic context, it is tempting to speculate on an important function for MBF, possibly related to transcription (as originally proposed), but most likely also related to RNA processing/maturation, and/or translation quality control.

Production of MBF, and the co-expression of MBF-SsGBP

The expression of full length *S. solfataricus mbf* gene results in a truncated MBF-C, as been demonstrated by SDS-PAGE (Fig. 4, lane 1) and mass spectroscopy (not shown). Apparently, the full-length MBF is frequently truncated in the poorly conserved

interdomain sequence, which probably is a flexible, exposed linker (Fig. 1). With that knowledge, the co-expression of MBF and GBP has been attempted because it was anticipated that this might stabilize the full-length MBF, and form a complex *in* vivo. The MBF and MBF-SsGBP were purified from 1 L cell culture. After heat treatment at 70^0 C for 20 min, most *E. coli* proteins were removed. The subsequent chromatographic step was Ni-NTA affinity column purification (the GBP had a C-terminal $(His)_6$ tag), and the proteins were at least 95% pure as judged from 15% SDS- PAGE (Fig. 4). Interestingly, the analysis of the co-expression revealed the presence of 2 high molecular weight products (Fig. 4A, lane 3). To identify the products, all major protein bands were excised and treated with trypsin, after which the peptides were eluted from the gel slice, desalted, concentrated, and analyzed by ESI Q-TOF tandem mass spectrometry. Twenty different peptides (80% sequence coverage, and quite high MOWSE scores) were found to match the top two bands with 59.6-KDa and 53.4-KDa respectively from the co-expression of MBF-GBP (Fig. 4A, lane 3). It should be mentioned that only peptides corresponding to the C-terminal domain of MBF were detected; hence, the co-expression does not lead to stabilization of the full-length MBF. MS analysis of the two lower bands revealed peptides from SsGBP and a few from MBF. To further characterize the obtained polypeptide products, Western blotting and native ESI mass spectroscopy were performed. The result of the Western blotting of anti-MBF shows that the top two bands include MBF (not shown). However, the other two bands do not include MBF. The anti-$(His)_6$ Western blotting detected four bands, corresponding to GBP with its C-terminal $(His)_6$. The Native MS did not reveal a mass that would correspond to a stable MBF-SsGBP complex.

The nature of the obtained MBF-SsGBP complex is not clear at present. In a recent study (Wu et al. 2007), *in vitro* pull down analysis suggested a weak interaction between the two proteins. However, analysis of *Sulfolobus* lysates has not revealed a similar complex (Blombach, pers. comm.). Hence, at present there is no evidence that the observed complex reflects a physiological state; it rather may correspond to an artifact of the heterologous expression, or the subsequent processing.

MBF-C binds to nucleic acid

MBF-C has been purified to homogeneity (Fig. 5). The initial steps of the MBF-C purification were a heat treatment and a separation on a HiTrap Heparin column. When the MBF fractions were concentrated and loaded on a 10/300 GL Superdex 200 column, different elution profiles for MBF-C were observed with and without DTT (1 mM) present. Without 1mM DTT two peaks are obtained, one in the void volume at 7 ml volume, and one that elutes at 16 ml volume. The former peak fractions contain DNA, the latter MBF-C (estimated 26 kDa, corresponding to dimer); with DTT in the running buffer only the latter peak is obtained. The association of MBF-C with DNA could be demonstrated by analysis on ethidium-bromide stained agarose gel, after DNase treatment (Fig. 5). Apart from DNA, RNA has recently been demonstrated to be associated with MBF (MBF-C) that was purified from its *E. coli* host (Blombach, personal communication). The physiological significance of the interaction with nucleic acid remains to be solved. The gel filtration elution profile suggests that, at least under these conditions, MBF-C appears to migrate as a dimer. On the other hand, analysis of MBF-C with His-tag by analytical

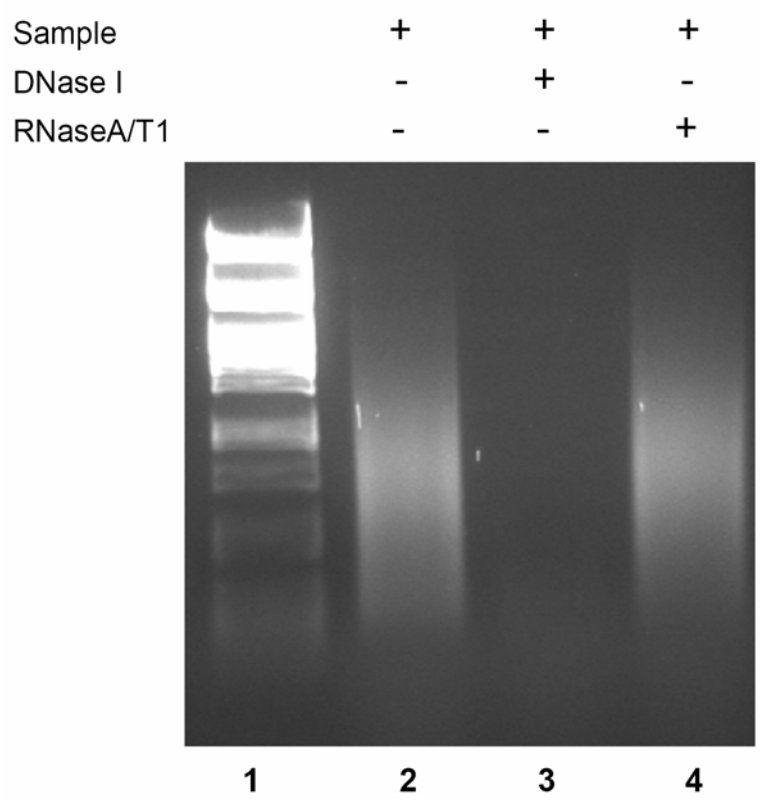


Figure 5. MBF-C binds to DNA verified by DNaseI and RNase A/T1 digestion. Lane 1: DNA marker; Lane 2: control (only the sample from peak I, Fig. 7C); Lane 3: The sample treated by DNaseI; Lane 4. The Sample treated by RNaseA/T1. See text for further explanation.

ultracentrifugation has revealed that MBF-C is a monomer in solution (data not shown). This discrepancy may reflect different purification procedures, and as such different amounts of contaminating nucleic acid; this matter requires more careful analysis in the future.

A. MBF-C **B. Complex of MBF-SsGBP**

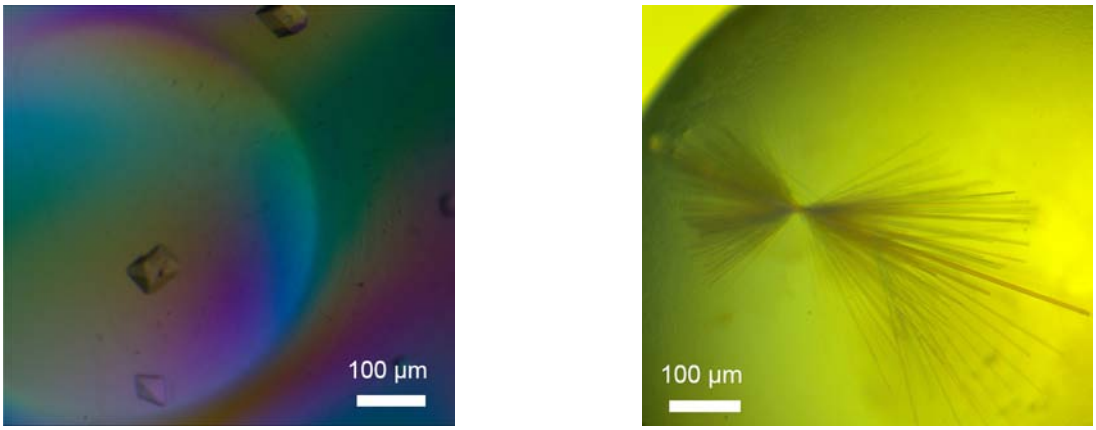


Figure 6. Crystallization trials of MBF and GBP

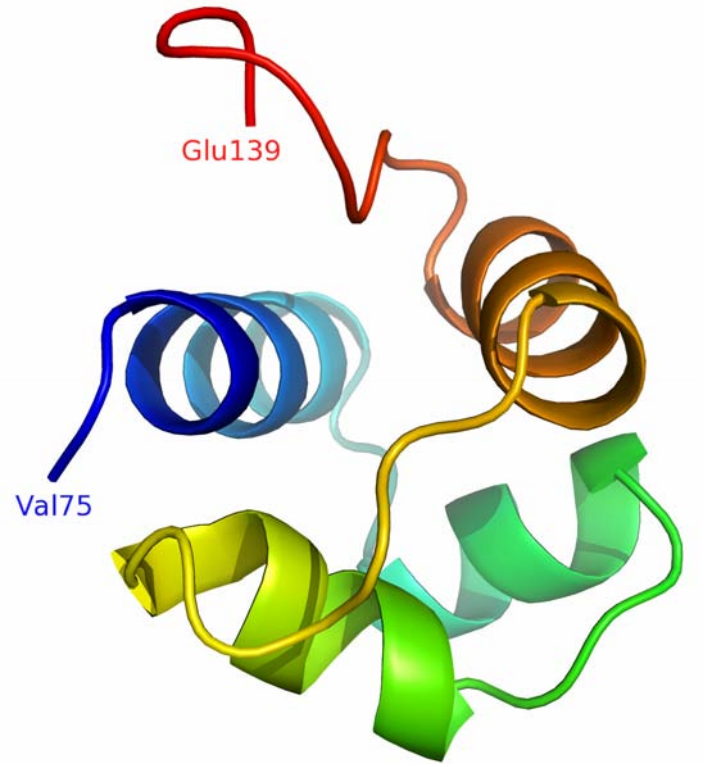


Figure 7. Structural model of *S. solfataricus* MBF-C comprising residues Val75-Glu139 based on NMR structure of human MBF (PDB ID: 1X57; Nameki, N. *et.al.*, unpublished).

Crystallization trials

So far, there is no 3D structure of MBF available from archaea yet. The only available structural information of MBF/MBF1 concerns the C terminus of human MBF. For that reason, it is still worth the effort to obtain the 3D structure of MBF in hyperthermophilic archaea. Unfortunately, the full-length MBF is not stable and always truncated at the linker region; the co-expression of MBF-SsGBP also does not help to obtain the stable MBF. For that reason, the MBF-C was used for the first crystallization trial. A second crystallization trail was to co-crystallize MBF-C and SsGBP in an attempt to obtain a complex. It was found that conditions of No. 4 of Screen 1 (0.1 M Tris hydrochloride pH 8.5, 2.0 M Ammonium sulfate), No. 35 of Screen II (0.1 M HEPES pH 7.5, 70% v/v (+/-)-2-Methyl-2,4-pentanediol) and No. 75 of Index I/II (0.2 M Lithium sulfate monohydrate, 0.2 M Lithium sulfate monohydrate,25% w/v Polyethylene glycol 3,350) are optimization conditions for the crystallization of MBF-C. For the complex of MBF and SsGBP, the optimized condition was No. 44 of MembFac (Fig. 6). The available crystals have been tested, but unfortunately did not diffract X-rays. Attempts to obtain better crystals are ongoing.

Structural model of MBF

Both eukaryal and archaeal MBF have a significant HTH-like motif at its C terminus domain (Aravind *et al.*, 2005; Aravind and Koonin, 1999). The 3D structure of the C-terminal domain of human MBF (residues D71 – K148) was recently resolved by NMR (PDB ID: 1X57; Nameki, N. *et.al.*, unpublished). It is a HTH motif, the sequence of which closely resembles that of the archaeal MBF (Fig. 1). Based on the 3D structure, a 3D model of the C terminus of MBF from *S. solfataricus* was built by SWISS-MODEL (http://swissmodel.expasy.org/) (Fig. 7). Since there are total 15 kinds of possibilities of conformations at the N- and C- terminal flexible regions in the human MBF NMR structure, it was necessary to delete these flexible regions during the homology-modeling of *S. solfataricus* was created (residues Val75 – Glu139).

Conclusions

MBF is present in all organisms in which TBP is used as the general transcription factor (Liu et al., 2007). Both in archaea and eukarya, it may play a role during transcription regulation (Kabe *et al.*, 1999; Tsuda *et al.*, 2004) and/or a translation-related process (Hendrick et al., 2001). Our data suggest that at least under certain conditions MBF from *Sulfolobus solfataricus* can interact with DNA or RNA. A recent immunological analysis by sucrose gradient fractions from *S. solfataricus* shows that MBF interacts with the 30S ribosomal subunit (Blombach, pers. comm.), whereas SsGBP fractionates with the 50S subunit. In archaea, transcription and translation are coupled (French *et al.*, 2007). So, it could be that MBF in archaea can interact with (stable) RNAs via DNA or other components during the transcription/translation regulation processing. To further find if *MBF* and *gbp* play a role in this coupling by ensuring ribosome fidelity, it is need to find whether the presence of a translating ribosome has any influence on the transcription termination by the advancing archaeal RNA polymerase (Liu *et al.*, 2007).

ACKNOWLEDGEMENTS

The authors acknowledge Dr. Kristina Lorenzen (Utrecht University) for help with native MS analysis; and Ms. Xiaoxia Yu, Mr. Yi Han and Dr. Zhiyong Lou at Tsinghua University (Beijing) for assistance with analytical ultracentrifugation and protein crystallization experiments.

REFERENCES

Aravind, L., Anantharaman, V., Balaji, S., Babu, M.M. and Iyer, L.M. (2005) The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev*, **29**, 231-262.

Aravind, L. and Koonin, E.V. (1999) DNA-binding proteins and evolution of transcription regulation in the archaea. *Nucleic Acids Res*, **27**, 4658-4670.

Bell, S.D. and Jackson, S.P. (1998) Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features. *Trends Microbiol*, **6**, 222-228.

Brendel, C., Gelman, L. and Auwerx, J. (2002) Multiprotein bridging factor-1 (MBF-1) is a cofactor for nuclear receptors that regulate lipid metabolism. *Mol Endocrinol*, **16**, 1367-1377.

Colfen, H., Boulter, J.M., Harding, S.E. and Watts, A. (1998) Ultracentrifugation studies on the transmembrane domain of the human erythrocyte anion transporter band 3 in the detergent C12E8. *Eur Biophys J*, **27**, 651-655.

French, S.L., Santangelo, T.J., Beyer, A.L. and Reeve, J.N. (2007) Transcription and translation are coupled in Archaea. *Mol Biol Evol*, **24**, 893-895.

Hendrick, J.L., Wilson, P.G., Edelman, II, Sandbaken, M.G., Ursic, D. and Culbertson, M.R. (2001) Yeast frameshift suppressor mutations in the genes coding for transcription factor Mbf1p and ribosomal protein S3: evidence for autoregulation of S3 synthesis. *Genetics*, **157**, 1141-1158.

Jindra, M., Gaziova, I., Uhlirova, M., Okabe, M., Hiromi, Y. and Hirose, S. (2004) Coactivator MBF1 preserves the redox-dependent AP-1 activity during oxidative stress in Drosophila. *Embo J*, **23**, 3538-3547.

Kabe, Y., Goto, M., Shima, D., Imai, T., Wada, T., Morohashi, K., Shirakawa, M., Hirose, S. and Handa, H. (1999) The role of human MBF1 as a transcriptional coactivator. *J Biol Chem*, **274**, 34196-34202.

Karlin, S., Mrazek, J., Ma, J. and Brocchieri, L. (2005) Predicted highly expressed genes in archaeal genomes. *Proc Natl Acad Sci U S A*, **102**, 7303-7308.

Keeling, P.J. and Doolittle, W.F. (1995) Archaea: narrowing the gap between prokaryotes and eukaryotes. *Proc Natl Acad Sci U S A*, **92**, 5761-5764.

Krutchinsky, A.N., Chernushevich, I.V., Spicer, V.L., Ens, W. and Standing, K.G. (1998) Collisional damping interface for an electrospray ionization time-of-flight mass spectrometer. *Journal of the American Society for Mass Spectrometry*, **9**, 569-579.

Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol*, **317**, 41-72.

Li, F.Q., Ueda, H. and Hirose, S. (1994) Mediators of activation of fushi tarazu gene transcription by BmFTZ-F1. *Mol Cell Biol*, **14**, 3013-3021.

Liu, Q.X., Jindra, M., Ueda, H., Hiromi, Y. and Hirose, S. (2003) Drosophila MBF1 is a co-activator for Tracheae Defective and contributes to the formation of tracheal and nervous systems. *Development*, **130**, 719-728.

Liu, Q.X., Nakashima-Kamimura, N., Ikeo, K., Hirose, S. and Gojobori, T. (2007) Compensatory change of interacting amino acids in the coevolution of transcriptional coactivator MBF1 and TATA-box-binding protein. *Mol Biol Evol*, **24**, 1458-1463.

Makarova, K.S., Aravind, L., Galperin, M.Y., Grishin, N.V., Tatusov, R.L., Wolf, Y.I. and Koonin, E.V. (1999) Comparative genomics of the Archaea (Euryarchaeota): evolution of conserved protein families, the stable core, and the variable shell. *Genome Res*, **9**, 608-628.

Makarova, K.S. and Koonin, E.V. (2003) Comparative genomics of Archaea: how much have we learned in

six years, and what's next? *Genome Biol*, **4**, 115.

Rivera, M.C., Jain, R., Moore, J.E. and Lake, J.A. (1998) Genomic evidence for two functionally distinct gene classes. *Proc Natl Acad Sci U S A*, **95**, 6239-6244.

Rivera, M.C. and Lake, J.A. (2004) The ring of life provides evidence for a genome fusion origin of eukaryotes. *Nature*, **431**, 152-155.

Snijders, A.P., Walther, J., Peter, S., Kinnman, I., de Vos, M.G., van de Werken, H.J., Brouns, S.J., van der Oost, J. and Wright, P.C. (2006) Reconstruction of central carbon metabolism in Sulfolobus solfataricus using a two-dimensional gel electrophoresis map, stable isotope labelling and DNA microarray analysis. *Proteomics*, **6**, 1518-1529.

Tahallah, N., Pinkse, M., Maier, C.S. and Heck, A.J. (2001) The effect of the source pressure on the abundance of ions of noncovalent protein assemblies in an electrospray ionization orthogonal time-of-flight instrument. *Rapid Commun Mass Spectrom*, **15**, 596-601.

Takemaru, K., Harashima, S., Ueda, H. and Hirose, S. (1998) Yeast coactivator MBF1 mediates GCN4-dependent transcriptional activation. *Mol Cell Biol*, **18**, 4971-4976.

Takemaru, K., Li, F.Q., Ueda, H. and Hirose, S. (1997) Multiprotein bridging factor 1 (MBF1) is an evolutionarily conserved transcriptional coactivator that connects a regulatory factor and TATA element-binding protein. *Proc Natl Acad Sci U S A*, **94**, 7251-7256.

Tsuda, K., Tsuji, T., Hirose, S. and Yamazaki, K. (2004) Three Arabidopsis MBF1 homologs with distinct expression profiles play roles as transcriptional co-activators. *Plant Cell Physiol*, **45**, 225-231.

White, M.F. (2003) Archaeal DNA repair: paradigms and puzzles. *Biochem Soc Trans*, **31**, 690-693.

Wu, H., Sun, L., Brouns, S.J., Fu, S., Akerboom, J., Li, X. and van der Oost, J. (2007) Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon Sulfolobus solfataricus. *Acta Crystallograph Sect F Struct Biol Cryst Commun*, **63**, 239-241.

Chapter 7

Summary & Conclusions

SUMMARY AND CONCLUSIONS

The work presented in this thesis has provided an integrated molecular approach to gain insight in eukaryal-like enzymes and regulatory proteins in archaea. Several eukaryal-like proteins from *Sulfolobus solfataricus*, such as α-galactosidase, Multi-protein bridging factor (MBF), and a HflX-like GTPase (SsGBP), were selected in this research. *S. solfataricus* is a hyperthermophilic archaeon that serves as a primary model system for several practical reasons: (i) *S. solfataricus* not only can thrive in extremely thermophilic environments, but also can grow aerobically on a wide variety of substrates in liquid media as well as plates (Zillig et al., 1998); (ii) its complete genome sequence (She et al., 2001), genetic systems (overproduction, knockout mutagenesis), in vitro transcription system (Bell and Jackson, 2001), and functional genomics technologies (transcriptomics, proteomics) (Brouns et al., 2006) have recently been established; (iii) general tools are available for the functional production of *S. solfataricus* proteins in a series of convenient heterologous expression systems of *Escherichia coli* that exploit a heat incubation as a very simple and efficient first purification step; (iv) the archaeal factors and complexes involved in gene expression differ substantially from bacterial systems; instead, they closely resemble counterparts in eukarya (reviewed by Thomm,1996; Forterre, 1997, 2002; Myllykallio et al.; Woese et al., 2000; Myllykallio and Forterre 2000; Bell and Jackson 2001; White and Bell, 2002; Reeve, 2003). Although some of these eukaryal systems are well characterized, there are still several proteins that have not been studied in any detail. In the latter case, insight in the homologous archaeal systems may elucidate relevant features of their eukaryal counterparts.

Following a brief outline of this thesis (**Chapter 1**), **Chapter 2** describes the construction of a novel thermostable marker for use in thermophilic archaea. Several double mutants of a bleomycin-binding protein were isolated with enhanced performance at high temperature both *in vivo* and *in vitro*. Structural analysis showed that the mutations gave rise to different means of stabilization in four parts of the protein. A mutated gene with a low GC content was created and made by DNA synthesis that can serve as an antibiotic resistance marker for aerobic and microaerophilic mesophiles, thermophiles, and hyperthermophiles. This

may allow for the development of efficient shuttle vectors and knock-out strategies for hyperthermophilic archaea and bacteria based on positive selection schemes. Moreover the mutated genes with an optimized high GC content may now permit multigene knock-out strategies in thermophiles, such as *Thermus thermophilus*, allowing further exploration and exploitation of thermophilic microbial sources.

In **Chapter 3**, we describe the purification and characterization of a thermostable α-galactosidase from cell extracts of *S. solfataricus* P2 grown on the trisaccharide raffinose. The enzyme, designated GalS, turned out to be highly specific for α-linked galactosides, which are optimally hydrolyzed at pH 5 and 90°C. The protein consists of 74.7-kDa subunits and has been identified as the gene product of open reading frame Sso3127. Its primary sequence is most related to plant enzymes of glycoside hydrolase family 36, which are involved in the synthesis and degradation of raffinose and stachyose. Both the *galS* gene from *S. solfataricus* P2 and an orthologous gene from *S. tokodaii* have been cloned and functionally expressed in *E. coli*, and their activity was confirmed. At present, these *Sulfolobus* enzymes not only constitute a distinct type of thermostable α-galactosidases within glycoside hydrolase clan D, but they also represent the first members from the Archaea.

In **Chapter 4**, the gene for a predicted GTP-binding protein from the hyperthermophilic archaeon *S. solfataricus*, termed SsGBP, has been cloned and overproduced in *E. coli*. The purified protein was crystallized using the hanging-drop method. A single-wavelength anomalous dispersion data set was collected to a maximum resolution of 2.0 Å using a single cadmium-incorporated crystal. The work presented here is the basis for the structural analysis described in the next chapter.

In **Chapter 5**, the first structural and functional information is presented on a member of the ubiquitous HflX family of GTPases. Two crystal structures of the SsGBP have been solved, in the nucleotide-free and the GDP-bound form. The structures reveal a two-domain arrangement with an unprecedented N-terminal domain and a classical GTPase (G) domain at the C-terminus. Subdomain I of SsGBP-N distantly resembles an Argonaute

domain, which is functionally related to target mRNA recognition and cleavage. Subdomain II has a Helix-Break-Helix fold, which is similar to the human Ras (or Rab) effector-like domain. Although it is tempting to speculate on functional similarities with these proteins, the significance of these observations remains to be established. Biochemical analyses indicate GTP hydrolysis by a truncated G domain, and nucleic acid binding by a truncated N-terminal domain. In *S. solfataricus* co-transcription (most likely) occurs of the genes encoding SsGBP and Multiprotein Bridging Factor (SsMBF), a protein reported to be involved in translation fidelity. Evidence is provided that both proteins interact *in vitro*, and that they co-purify with *S. solfataricus* ribosomal subunits. Based on these experimental results and analysis of the conserved context of HflX-encoding genes in bacterial and archaeal genomes, we propose a role of the HflX-like GBPs in RNA maturation.

In **Chapter 6**, the Multi-protein bridging factor (MBF) from hyperthermophilic archaeon *S. solfataricus* is characterized. MBF is highly conserved in archaea and eukarya, but absent in bacteria . In archaea *mbf* shows a conserved genomic context with genes encoding a proteasome-activating nucleotidase (PAN), a GTP-binding protein (GBP), a basic transcription factor (TFE), and proteins involved in RNA binding/modification. In the hyperthermophilic archaeon *S. solfataricus, mbf*-stop and *gbp*-start codons overlap (ATGA), and hence the genes are most likely organized as an operon. We have cloned and over-expressed the gene for the full-length MBF and a C-terminal MBF-fragment (MBF-C) from *S. solfataricus.* Both MBF and MBF-C could be produced as soluble protein, although the full length MBF generally appeared to be truncated in *E. coli*. Ultracentrifugation analysis and native mass spectrometry revealed that MBF-C is a monomer in solution. Furthermore, we demonstrate that the heterologously expressed MBF-C is co-purified with host DNA. We cloned the predicted operon *mbf-gbp* from *S. solfataricus* in *E.coli*. Interestingly, co-expression of the two genes as operon in *E. coli* resulted in an apparent stable complex. *In vitro* analyses suggests that MBF might interact with SsGBP. A possible role of both proteins in transcription/translation regulation in archaea is discussed.

REFERENCES

Bell, S.D. and Jackson, S.P. (2001) Mechanism and regulation of transcription in archaea. *Curr Opin Microbiol*, **4**, 208-213.

Brouns, S.J., Walther, J., Snijders, A.P., van de Werken, H.J., Willemen, H.L., Worm, P., de Vos, M.G., Andersson, A., Lundgren, M., Mazon, H.F., van den Heuvel, R.H., Nilsson, P., Salmon, L., de Vos, W.M., Wright, P.C., Bernander, R. and van der Oost, J. (2006) Identification of the missing links in prokaryotic pentose oxidation pathways: evidence for enzyme recruitment. *J Biol Chem*, **281**, 27378-27388.

Forterre, P. (1997) Archaea: what can we learn from their sequences? *Curr Opin Genet Dev*, **7**, 764-770.

Myllykallio, H., Lopez, P., Lopez-Garcia, P., Heilig, R., Saurin, W., Zivanovic, Y., Philippe, H. and Forterre, P. (2000) Bacterial mode of replication with eukaryotic-like machinery in a hyperthermophilic archaeon. *Science*, **288**, 2212-2215.

Reeve, J.N. (2003) Archaeal chromatin and transcription. *Mol Microbiol*, **48**, 587-598.

She, Q., Singh, R.K., Confalonieri, F., Zivanovic, Y., Allard, G., Awayez, M.J., Chan-Weiher, C.C., Clausen, I.G., Curtis, B.A., De Moors, A., Erauso, G., Fletcher, C., Gordon, P.M., Heikamp-de Jong, I., Jeffries, A.C., Kozera, C.J., Medina, N., Peng, X., Thi-Ngoc, H.P., Redder, P., Schenk, M.E., Theriault, C., Tolstrup, N., Charlebois, R.L., Doolittle, W.F., Duguet, M., Gaasterland, T., Garrett, R.A., Ragan, M.A., Sensen, C.W. and Van der Oost, J. (2001) The complete genome of the crenarchaeon Sulfolobus solfataricus P2. *Proc Natl Acad Sci U S A*, **98**, 7835-7840.

Thomm, M. (1996) Archaeal transcription factors and their role in transcription initiation. *Fems Microbiology Reviews*, **18**, 159-171.

White, M.F. and Bell, S.D. (2002) Holding it together: chromatin in the Archaea. *Trends Genet*, **18**, 621-626.

Woese, C.R., Olsen, G.J., Ibba, M. and Soll, D. (2000) Aminoacyl-tRNA synthetases, the genetic code, and the evolutionary process. *Microbiol Mol Biol Rev*, **64**, 202-236.

Zillig, W., Arnold, H.P., Holz, I., Prangishvili, D., Schweier, A., Stedman, K., She, Q., Phan, H., Garrett, R. and Kristjansson, J.K. (1998) Genetic elements in the extremely thermophilic archaeon Sulfolobus. *Extremophiles*, **2**, 131-140.

Hoofdstuk 8

Nederlandse Samenvatting & Conclusies

SAMENVATTING EN CONCLUSIES

Het werk dat in deze thesis wordt gepresenteerd heeft middels een geïntegreerde moleculaire benadering het inzicht in eukaryote enzymen en regulerende eiwitten in archaea versterkt. Verschillende, aan eukaryote verwante, eiwitten van *Sulfolobus solfataricus*, het α-galactosidase, Multi-protein bridging factor (MBF), en een HflX-like GTPase (SsGBP), zijn onderzocht. *S. solfataricus* is een hyperthermophiele archaeon, dat om verscheidene praktische redenen als modelorganisme wordt gebruikt: (i) *S. solfataricus* kan in uiterst thermofiele milieus aeroob op een grote verscheidenheid van substraten in zowel vloeibare als vaste media groeien (Zillig et al, 1998); (ii) de volledige genoomsequentie van *S. solfataricus* is bepaald (She et al, 2001) en genetische systemen (overproductie, knock-out mutagenese), *in vitro* transcriptie (Bell en Jackson 2001), en functionele genomics technologieën (transcriptomics, proteomics; Brouns et al 2006) zijn onlangs ontwikkeld; (iii) de algemene moleculaire technieken zijn bruikbaar voor de functionele productie van eiwitten uit *S. solfataricus* in een reeks van heterologe expressiesystemen in *Escherichia coli*, waarbij als eerste efficiënte en eenvoudige zuiveringsstap een hitte incubatie gebruikt kan worden; (iv) de archaeële factoren en de complexen betrokken bij genexpressie verschillen wezenlijk van bacteriële systemen; in plaats daarvan lijken zij meer op hun tegenhangers in eukarya (samengevat door Thomm, 1996; Forterre, 1997..2002; Myllykallio et al, 2000; Garrett 1999; Woese et al, 2000; Myllykallio en Forterre 2000; Bell en Jackson 2001; White en Bell, 2002; Reeve, 2003). Hoewel sommige van deze eukaryote systemen goed zijn gekarakteriseerd, zijn er nog talloze eiwitten die nog in geen enkel detail zijn bestudeerd. In het laatstgenoemde geval zou inzicht in de homologe archaeële systemen relevante eigenschappen van hun eukaryote tegenhangers kunnen ophelderen.

Na een kort overzicht van deze thesis (**Hoofdstuk 1**), wordt in **Hoofdstuk 2** beschreven hoe de bouw van een nieuwe thermostabiele marker voor gebruik in thermophiele archaea in zijn werk is gegaan. Verschillende dubbelmutanten van een bleomycine-bindend eiwit werden geïsoleerd met verbeterde prestaties, zowel *in vitro* als *in vivo,* bij hoge temperatuur. Structurele analyse toonde aan dat de

mutaties leidden tot verschillende manieren van stabilisatie in vier delen van het eitwit.
Een gemuteerd gen met een laag GC gehalte was ontworpen en gemaakt door middel van DNA synthese. Dit gen kan gebruikt worden als een marker voor antibioticum resistentie in aerobe en microaerofiele mesophielen, thermophielen en hyperthermophielen en zou gebruikt kunnen worden voor de ontwikkeling van efficiënte shuttlevectoren en knock-outmethoden voor hyperthermophiele archaea en bacteriën gebaseerd op positieve selectie. Daarnaast zouden de gemuteerde genen met een verhoogd GC gehalte gebruikt kunnen worden voor multigene knock-outmethoden in thermofielen, zoals *Thermus thermophilus*, wat verder onderzoek en exploitatie van microbiële thermophiele bronnen mogelijk maakt.

In **Hoofdstuk 3** beschrijven wij de purificatie en karakterisatie van een thermostabiel α-galactosidase uit celextracten van *S. solfataricus* P2, die gegroeid is op trisaccharide raffinose. Het enzym, benoemd tot GalS, bleek zeer specifiek te zijn voor α verbonden galactosiden, welke optimaal gehydroliseerd worden bij pH 5 en 90°C. Het eiwit bestaat uit subunits van 74,7 kDa groot en is geïdentificeerd als het genproduct van het open reading frame Sso3127. Zijn primaire sequentie is het meest gerelateerd aan dat van plantenenzymen uit de glycosidehydrolase familie 36, welke een rol spelen in de synthese en afbraak van raffinose en stachyose. Zowel het *galS* gen uit *S. solfataricus* P2 als een ortholoog gen uit *S. tokodaii* zijn gekloneerd en functioneel tot expressie gebracht in *E. coli*. De activiteit hiervan is bevestigd. Op dit moment vormen deze *Sulfolobus* enzymen niet alleen een apart type thermostabiele α-galactosidases binnen de glycosidehydrolase clan D, maar vertegenwoordigen zij ook de eerste archaeële leden hierbinnen.

In **Hoofdstuk 4** wordt beschreven hoe een gen voor een voorspeld GTP-bindend eiwit uit het hyperthermophiele archaeon *S. solfataricus,* genaamd SsGBP, is gekloneerd en tot expressie is gebracht in *E. coli*. Het gezuiverde eiwit is gekristalliseerd via de “hanging drop”-methode. Een enkel kristal, waarin cadmium was geïncorporeerd, leverde een dataset op met single-wavelength anomalous dispersion met een maximale resolutie van 2,0 Å. Het werk beschreven in dit

hoofdstuk levert de basis voor de structurele analyse beschreven in het volgende hoofdstuk.

In **Hoofdstuk 5** wordt de eerste structurele en functionele informatie over een lid van de alomtegenwoordige HflX-familie van GTPases gepresenteerd. Twee kristal structuren van SsGBP zijn opgehelderd: de nucleotidevrije vorm en in de GDP gebonden vorm. De structuren tonen een ordening in twee domeinen, namelijk een nooit eerder gevonden N-terminaal domein (N) en een klassiek GTPase domein (G) aan de C-terminus. Subdomein I van SsGBP-N lijkt op een Argonaut domein, welke functioneel in verband wordt gebracht met mRNA herkenning en splitsing. Subdomein II heeft een Helix-Break-Helix motief gelijk aan dat van het humane Ras-(of Rab-)effectorachtige domein. Toch moet de betrouwbaarheid van deze observaties in twijfel worden getrokken, ook al is het verleidelijk om te speculeren over de functionele overeenkomsten met deze eiwitten. Biochemische analyses wijzen op GTP hydrolyse door het ingekorte G domein en nucleïnezuur binding door het ingekorte N-terminale domein. In *S. solfataricus* vindt naar alle waarschijnlijkheid cotranscriptie tussen de genen die coderen voor het SsGBP en het Multi-protein bridging factor (SsMBF), een eiwit waarvan het vermoeden bestaat dat het betrokken is bij translatieprecisie. Bewijs is geleverd dat deze beide eiwitten *in vitro* interacteren, en dat ze meegezuiverd worden met *S. solfataricus* ribosomale subeenheden. Uitgaande van deze experimentele resultaten en de analyse van de geconserveerde context van Hflx-achtige genen in bacteriële en archaeële genomen, voorspellen wij een rol voor HflX-achtige GBP's in RNA maturatie.

In **Hoofdstuk 6** is de Multi-protein bridging factor (MBF) uit het hyperthermofiele archaeon *S. solfataricus* gekarakteriseerd. MBF is zeer geconserveerd in archaea en eukarya, maar afwezig in bacteria. In archaea vertoont *mbf* een geconserveerd genomische context met daarin genen die coderen voor het proteasome-activating nucleotidase (PAN), het GTP-binding protein (GBP) en basal transcription factor E (TFE), en eiwitten die betrokken zijn bij RNA modificatie of binding. In het hyperthermofiele archaeon *S. solfataricus* overlappen het stopcodon van *mbf* en het startcodon van *gbp* (ATGA), en daardoor zijn de genen waarschijnlijk georganiseerd in een operon. We hebben het complete *mbf* gen en het c-terminale

mbf-fragment (MBF-C) uit *S. solfataricus* gekloneerd en tot overexpressie gebracht. Zowel MBF als MBF-C konden als oplosbaar eiwit worden geproduceerd, alhoewel het complete MBF eiwit vaak ingekort in *E. coli* verscheen. Ultracentrifugeanalyse en native mass spectrometry hebben aangetoond dat MBF-C een monomeer is in oplossing. Verder demonstreren wij dat DNA wordt meegezuiverd met heteroloog tot expressie gebracht MBF-C. We hebben het voorspelde operon *mbf-gbp* uit *S. solfataricus* in *E. coli* gekloneerd. Deze coexpressie van beide genen als een operon leidt interessant genoeg tot een duidelijk stabiel complex in *E. coli*. *In vitro* analyses tonen dus aan dat MBF zou kunnen interacteren met GBP. Een vermoedelijke rol van beide eiwitten in transcriptie/translatie regulatie in archaea wordt bediscussieerd.

Acknowledgements

It has been a full four-year journey to complete my dissertation and subsequent Ph.D. When you are completing your dissertation, you find it's true that "Life is what happens". Life doesn't stand still, nor wait until you are finished and have time to mange it. Much has happened and changed during the time of my Ph.D. research. This thesis would not be here without the help of many people, who made their contribution in one way or another.

First of all, I would like to thank my promotors. John, you are a constant source of inspiration and motivation for my research. Your patience, flexibility, genuine caring and concern, and faith in me during the dissertation process enabled me to attend to life while also earning my Ph.D. When I go back to look at my project now, it seems that you set up a milestone road map for my Ph.D. program four years before. You always keep the research group staying on the frontier of scientific research by actively open-minded thinking, creating research environments with a high degree of freedom and so many innovative strategies. I am so glad that you accepted me and provided me such a great opportunity to do the Ph.D. study in your research group. When we sit together on the Great Wall, I thought you are the great man who knows how to make correct decisions at the correct time in the correct place. I also learn from you how to behave like a scientific researcher, which will benefit me for my entire lifetime.

Willem, thank you so much for your valuable support and input, especially at late stage your great comments and suggestions for the manuscript, which speeded me up to finish this dissertation in time. It was a fantastic event to join the biweekly Ph.D. & Post-doc meeting and enjoy your great comments for the presentations. It is amazing that you could often use very simple and fast words to get me back in motion when I was fogged in some talks. Before I went to Beijing to continue my research in 2005, you told me that I would miss the lab. Indeed, I will miss the Laboratory of Microbiology for ever.

Also, my most earnest acknowledgment must go to my co-promotor, Prof. Zihe Rao. Nearly four years ago, a telephone conversation between you and John started me on the path I traveled between Wageningen and Beijing. As a national strategic scientist, you have to deal with lots of multiple important tasks everyday. However,

you always tried to find time to enter the lab, to organize the group meeting and to guide us effectively. Your valuable support and input make the collaboration fruitful and enjoyable, and make this thesis as it is now. I am very glad that I had the great chance to work in your lab. Thank you!

I'd like to special thanks to Dr. Cai Zhang from Oklahoma Medicinal Research Foundation. As a famous crystallographer and GTPase expert, you helped push me through the chapter 5. With gentle encouragement, you guided me, and pointed me in the direction of professionalism and outstanding research in this area. I am truly thankful for what you have done for me.

I'd also like to give a heartfelt, special thanks to my paranymph and routine supervisor. Stan, you was one of the first friendly faces to greet me when I began this doctoral program and has always been a tremendous help no matter the task or circumstance. We had daily discussions that greatly inspired me. It was impressive that you have known the answer to the questions I've ever asked. You have taught me so much about science, and proved yourself a great friend. Thank you!

I also owe a huge debt of gratitude to Lei. In Autumn, 2005, while I worked in Prof. Rao's Lab, I got the chance to meet Lei for the first time. Lei, your excellent knowledge and experience in the field of crystallography was very impressive to me. You shall always be remembered as a patient smiling face, a warm and friendly heart and one of the few who assisted me to completing my doctoral program. We have had a fruitful collaboration, I hope we will continue our collaboration in the future.

Bart and Fabian, I was glad to work with you in one Vici-project. It was a good experience for me not only in scientific research, but also in the capacity to work as a team member within a multi-disciplinary group. Bart, thank you for your great arrangement so that I could join the Ph.D. Study Trip California, 2006. Also, thanks for becoming my paranymph and translating the Dutch summary of my thesis. Fabian, your technique experience in the lab was impressive. Thank you for your input, especially during I wrote the chapter 5 and 6 of my thesis. Although the MBF project is a bit tricky, I hope both of you will keep on rolling on the road of this project.

Matthijs, you are a great mate. After "MBF" project, you are becoming the new hot topic core together with Stan. I hope I will read your CRISPRs story in *CNS* soon.

You are indeed a good brother. Thank you for your invitation to watch such a nice movie, and I really enjoyed the Italian style cake you made.

I wish to convey my sincere gratitude to Ans and Servé. Ans, your great arrangement always kept the lab perfectly order and clean. Servé, I was happy to sit side by side with you once in the office and enjoyed surfing with you in San Diego. Far too many people to mention individually have assisted in many ways during my work at Wageningen. They all have my sincere gratitude. In particular, I would like to thank Thijs, Johan, Krisztina, Jasper (Jaape), Ronnie, Harmen, Suzanne, Jasper W., Marco, Mark, John R., Marcel, Odette, Petra, Pawel, Corné, Marke, Colin, Katrin, Ratnesh, Pierpaolo, Magnus. Special thanks to my roommates in MolEco Thomas, Sahar, Farai, Rozelin, Hermien. Hauke, I acknowledge you that you let me sit in your group temporally when I returned back from Beijing. Fos, I thank you for your kindly invitation for the Christmas dinner. I would like to thank Francis, Wim, Nees and Jannie for the great help when I worked in the lab.

My gratitude is also extended to my colleagues from Prof. Rao's group in Beijing. They were always enthusiastic and provided a positive environment, especially Wei Peng, Sheng Fu, Xiaohang Tong, Shuang Li, Yanling Ma, Zhiyong Lou. I am very grateful to Dr. Mark Bartlam, Dr. Xuemei Li, Dr. Neil Shaw, Ms. Xiaoxia Yu, Mr. Yi Han and Ms. Ping Shan. Their academic support and input and personal cheering are greatly appreciated.

I thank Bram from Sheffield, Kristina and Robert from Utrecht for Mass Spectrometry analysis of my enzymes. Aalt-Jan for Docking experiment.

In addition, these acknowledgements would not be complete if I did not mention my lovely son, Juncheng Wu. One week later after the big day of my defense, he will be one year old…the "little" support system…There is nothing like the innocent face of a sleeping child to bring out the strongest determination to succeed. It was his magic that pulled me out of many intervals of depression.

Of course no acknowledgements would be complete without giving thanks to my parents. They taught me about hard work and self-respect, about persistence and

how to be independent. Mom, especially, was a great role model of resilience, strength and character. I am too proud of them and love them very much.

Last, but certainly not least, I must acknowledge with tremendous and deep thanks my wife, Lei Wang. Through her love, patience, supporting, I've been able to complete this long dissertation journey. She has patiently endured many, many long hours alone while I worked on my dissertation. I love you and am forever indebted to you for taking care of our son.

Hao

December 2007

ABOUT THE AUTHOR

Hao Wu was born in Anhui, China on November 02, 1970. He became interested in Biology, which might be affected by his old brother who is a biology teacher. He chose to study Applied Fungi in Hua Zhong Agricultural University, China after he finished the high school. From 1996 to 2000, he worked as Assistant Researcher & Lecturer in the Department of Biotechnology at Anhui Agricultural University, China.

After he visited the University of Georgia, USA in November 2000, he made the decision to move to Wageningen University, the Netherlands to study his M.Sc. degree with specialization in Bioinformatics in May 2001. In March 2003, he obtained his M.Sc. degree of Biotechnology after he finished the M.Sc. thesis, "*Information Research for a SNP Database for GPCRs*", at the division of Medicinal Chemistry of the Leiden/Amsterdam Center for Drug Research in Leiden University, the Netherlands.

On the basis of his experience and his enthusiasm, he got a Research Fellowship (European Union project) in April 2003, which allowed him to work in Prof. John van der Oost's research group, Bacterial Genetics of Laboratory of Microbiology. He worked in the field of Molecular Microbiology and Crystallography. This has led him to continue his Ph.D. scientific career officially in August 2003. Since then, his research interests have always been linked to characterization of eukaryal-like proteins in archaea by Biochemistry and Crystallography. For the goal of crystallography research in his Ph.D. project, he went to the Institute of Biophysics of Chinese Academy of Sciences in Beijing and joined Prof. Zihe Rao's research group from August 2005 to August 2006. After one-year fruitful collaboration research, he returned to Wageningen University and continued his Ph.D. project and wrote this thesis.

LIST OF PUBLICATIONS

Brouns, S.J., **Wu, H.**, Akerboom, J., Turnbull, A.P., de Vos, W.M. and van der Oost, J. (2005) Engineering a selectable marker for hyperthermophiles. *J Biol Chem*, **280**, 11422-11431.

Brouns, S.J., Smits, N., **Wu, H.**, Snijders, A.P., Wright, P.C., de Vos, W.M. and van der Oost, J. (2006) Identification of a novel alpha-galactosidase from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *J Bacteriol*, **188**, 2392-2399.

Wu, H., Sun, L., Brouns, S.J., Fu, S., Akerboom, J., Li, X. and van der Oost, J. (2007) Purification, crystallization and preliminary crystallographic analysis of a GTP-binding protein from the hyperthermophilic archaeon *Sulfolobus solfataricus*. *Acta Crystallograph Sect F Struct Biol Cryst Commun*, **63**, 239-241.

Wu, H., Sun, L., Blombach, F., Brouns, S.J. , Snijders, A.P.L., Lorenzen, K., van den Heuvel, R.H.H., Heck, A.J.R., Fu, S., Li, X., Zhang, X. C., Rao, Z., and van der Oost, J. (2007) Structural and functional analysis of a ubiquitous HflX-like GTPase. *(Submitted to EMBO J.)*

Wu, H., Blombach, F., de Koning, B., Sun L. , Snijders, A.P.L. , Thijs J. G. Ettema, T.J.G., Brouns, S.J., de Vos, W.M., and van der Oost, J. Molecular characterization of the eukaryotic-like Multi-protein bridging factor (MBF) from thermophilic archaeon *Sulfolobus solfataricus*. (*Manuscript in preparation*)

EDUCATION STATEMENT OF GRADUATE SCHOOL VLAG

The graduate school hereby declares that the PhD candidate has compiled with the educational requirements of the graduate school VLAG by following the discipline specific activities listed below.

	ECTS credits
Discipline specific activities	
Courses	
Protein Engineering, Wageningen, 2004	1.0
Bio-Nanotechnology, Wageningen, 2005	1.5
Radiation expert 5B, Larenstein, Velp, 2004	1.4
Crystallography Summer School, Beijing, 2005	3.0
Get-Phase Beijing, 2005	2.4
Meetings	
Annual Dutch Molecular Genetics meeting, Lunteren	1.7
Annual Dutch Protein Study Group, Lunteren	3.0
Annual CNLN forum, the Netherlands	1.7
The 5th "TICPS" International Conference of Protein Science, Beijing, 2005	3.0
International High-Level Forum on Bioeconomy, Beijing, 2005	1.4
Biophysics Society Meeting, Qingdao, China, 2006	1.9
General courses	
English Intermediate, 2005	1.7
Optionals	
Preparation PhD research proposal	6.0
VLAG PhD trip California, 2006	2.8
Bacterial Genetics weekly group meeting	3.0
Microbiology biweekly PhD & Postdoc meeting	3.0
Total	***38.5***

I ECTS (European Credit Transfer System) represents the normative study load of 28 hours

Following the unanimous approval of the examination committee and after careful consideration and consultation with the promotores, the Academic Board of Wageningen University has assessed the total contribution of the candidate to the work described in this thesis as worthy for conferring the degree of doctor. The latter assessment is primarily based on the volume and quality of the contribution of the candidate to the body of work described in the chapters 4, 5 and 6.

The research described in this thesis has been carried out with the financial support of the European Union (SCREEN Project; QLK3-CT-2000-00649), KNAW (NL-China Exchange Programme; 04CDP006), and NWO (ALW-Vici project; 865.05.001).

Cover SsGBP and Chang'e – I

Printing: Drukkerij Propress, Wageningen, The Netherlands