Critical Reviews in Biochemistry and Molecular Biology, 37(4):199–258 (2002)

Transcription in the Archaea: Basal Factors, Regulation, and Stress Gene Expression

*Anthony J. Hickey, Everly Conway de Macario, and Alberto J. L. Macario**

Wadsworth Center, Division of Molecular Medicine, Room B-749, New York State Department of Health; and Department of Biomedical Sciences, School of Public Health, The University at Albany (SUNY); Empire State Plaza, P.O. Box 509, Albany, New York 12201-0509, USA

* **Corresponding author:** Alberto J. L. Macario, M.D., Wadsworth Center, Room B-749, Division of Molecular Medicine, New York State Department of Health, Empire State Plaza, P.O. Box 509, Albany, New York 12202-0509, USA. Telephone: (518) 474-2781; FAX: (518) 474-1213. e-mail: macario@wadsworth.org

ABSTRACT: A brief survey is presented of salient findings on transcription in the Archaea, focussing on stress genes of the *hsp70(dnaK) l*ocus, which code for the molecular chaperones Hsp70(DnaK), Hsp40(DnaJ), and GrpE. Archaeal basal factors and some recently characterized regulators pertinent to non-stress genes are presented first to show their similarities and differences with equivalents in organisms of the other two phylogenetic domains, Bacteria and Eucarya, and to reveal clues on how these or similar factors might transcribe and regulate the archaeal stress genes. The second part of the article deals with the *hsp70(dnaK)*-locus genes, particularly those from *Methanosarcina mazeii*, because they are virtually the only ones within the methanogenic Archaea whose patterns of constitutive and stress-induced expressions have been studied. Therefore, these genes, provide a standardized model system to elucidate transcription initiation and regulation at the molecular level in this phylogenetic group. Promoters, and other *cis*-acting sites that are, or might be, involved in stress-gene expression are described. Conformational changes of basal transcription factors after interaction with stress-gene promoters are discussed that suggest ways for generating a large diversity of initiation complexes using a few factors and DNA sites in different combinations. Likewise, the effects of stress on DNA topology and on TBP–TFB-promoter complex formation and tightness are described, which might also contribute to the generation of transcription-initiation complex diversity. This diversity would be key to differential gene expression, namely, which genes are transcribed, when (basal, steady expression vs. sporadic stress-induced expression), and to what level. Future research should investigate this diversity, and the mechanism of complex formation and action at the atomic, molecular, and supramolecular levels to elucidate the dynamics of transcription initiation in real time.

KEY WORDS: transcription initiation, promoter, TBP, TFB, TFE, regulatory factors, stress genes, *Methanosarcina mazeii*, *hsp70(dnaK)* locus, DNA binding, conformational changes.

1040-9238/02/$.50

Table of Contents

I. INTRODUCTION

A. Scope

1. Organisms

Living organisms, uni- and multicelled, rely on anti-stress mechanisms for survival. Stress genes and their products, stress proteins, many of which are molecular chaperones, are part of these mechanisms. It follows that understanding how these genes and proteins function is directly relevant to understanding how cells and organisms deal with stress, and to developing means for improving cell survival. The first step toward understanding how genes function is to elucidate their mechanism of transcription and its regulation.

Here a review of recent data on transcription is presented, focusing on basal mechanisms pertinent to a group of stress genes in organisms of the phylogenetic domain Archaea. Among these organisms, those belonging to the genus *Methanosarcina* are particularly interesting for various reasons, some of which are listed in Section I.B.2. Consequently, this review concentrates on *Methanosarcina* species, especially *M. mazeii*.

2. Genes

There are several stress genes, and families of stress genes, all of them involved in one way or another in the response to stress, and also in physiological functions in the absence of stress. The latter involvement in normal physiology adds to the importance of these genes and their protein products and confirms their worthiness as research targets.

Here, we focus on a group of genes whose products form the molecular chaperone machine. These genes are *hsp70(dnaK)*, *hsp40(dnaJ)*, and *grpE*, and their respective protein products are named Hsp70(DnaK), Hsp40(DnaJ), and GrpE. Because more often than not these genes are clustered in prokaryotes (archaea and bacteria) in a small chromosomal region with *hsp70(dnaK)* in the middle, they are known collectively as the *hsp70(dnaK)*-locus genes.

3. Transcription Factors

Archaeal organisms have a basal transcription machinery similar to that of eukaryotes, but much simpler. While in eukaryotes transcription initiation of protein-coding genes requires over 40 different molecules, including basal transcription factors, RNA polymerase (RNAP) subunits, TBP-associated factors (TAFs), and other accessory proteins (Hampsey, 1998; Albright and Tjian, 2000; Lee and Young, 2000; Lemon and Tjian, 2000; Woychik and Hampsey, 2002), only 16 or so are needed in archaea (see following sections).

To initiate transcription of protein-coding genes, eucaryal organisms use RNA polymerase II (RNAPII) consisting of several subunits, and various basal factors, such as the TATA-binding protein (TBP), a component of transcription factor TFIID, and transcription factors IIB (TFIIB), IIA (TFIIA), IIE (TFIIE), IIF (TFIIF), and IIH (TFIIH) (Lee and Young, 2000; Lemon and Tjian, 2000; Woychik and Hampsey, 2002). Archaea possess homologs of the eucaryal TBP, TFIIB, and TFIIE *alpha* subunit and an RNAP similar to RNAPII, named TBP, TFB, TFE, and RNAP, respectively (Hausner and Thomm, 1993; Marsh et al., 1994, Rowlands et al., 1994; Golh et al., 1995; Soppa, 1999; see also below). Although archaea have TBP, no TBP-associated factors (TAFs), which in eukaryotes are

components of TFIID (Albright and Tjian, 2000), have yet been found in these organisms.

In order to present a picture of transcription initiation of the *hsp70(dnaK)*-locus genes, it is necessary to first introduce the basal factors that are shared by archaea and eucarya and discuss in some detail when (e.g., under physiological vs. stressful conditions) and how (i.e., molecular mechanism) these factors work, what are their functions, and with what else (e.g., other factors basal or regulatory and DNA promoter or other sites) do they interact. These questions are dealt with in the first part of this article.

4. Regulatory Factors

Archaea must also have auxiliary factors to monitor and regulate the function of the basal factors, but, thus far, only a few of them have been well characterized. Interestingly, most of the few identified regulators resemble bacterial rather than eucaryal regulators (see Section VII). A minority seem to be closer in mechanism of action to eukaryotic activators, as discussed in Section VII.E. This is puzzling, and it is another reason why archaea are organisms of special interest in studying molecular biology and the genetics of transcription regulation, that is, they offer a tantalizing combination of factors and *cis*-acting elements of eukaryotic and bacterial types, seemingly challenging current evolutionary theories.

B. Model Systems

1. Archaea and Methanosarcina Species

The genus *Methanosarcina* includes several species that provide suitable model systems to investigate unresolved biological problems (e.g., the evolutionary origins and significance of multicellular structures), are ecologically important, and have promising biotechnological potential. They play a central role in the bioconversion of organic matter, including wastes from homes and cities, industries, and farms, and thereby generate methane, namely, *Methanosarcinae* are key components of renewable sources of energy (Zinder, 1993; Schmidt and Ahring, 1995; Soerensen et al., 1997). Furthermore, transformation systems for at least two species, *M. mazeii* (Conway de Macario et al., 1996) and *M. acetivorans* (Metcalf et al. 1997; Zhang et al., 2002), have been developed and standardized; consequently, genetic, molecular genetic, and molecular biologic analyses are possible with means matching those available for other, more ancient and better-known bacterial and eucaryal experimental models.

Most importantly, the genomes of at least four species (*M. acetivorans*, *M. barkeri*, *M. thermophila*, and *M. mazeii*) have been sequenced and are, or soon will be, available for study. There is no need to reemphasize how important it is to have access to all the genes of an organism, or group of related species, *Methanosarcinae* in this case, to understand their physiology, including their anti-stress mechanisms.

2. Methanosarcina mazeii

M. mazeii is a convenient model to study stress-gene regulation for the following reasons:

1. It is one of the archaeal species that possess the genes encoding the components of the molecular chaperone machine and that has been studied extensively (references in Macario et al., 1999). Other archaeal species either

do not have these genes, or, if they do [for example, *Haloarcula marismuorti (marismortui), Halobacterium cutirubrum*, *Thermoplasma acidophilum*, *Methanobacterium thermoautotrophicum*, and *Thermoplasma volcanium* (Gupta and Singh, 1992; 1994; Kawashima et al., 2000; Smith et al., 1997; Ruepp et al., 2000)], no data are available about them. Thus, no basic information exists pertinent to these latter organisms comparable to that available for *M. mazeii*.

2. The molecular chaperone machine genes of *M. mazeii* were most likely received from bacteria by horizontal transfer (Gribaldo et al., 1999; Macario and Conway de Macario, 1999), and thus they provide a unique opportunity to study the expression of "foreign" bacterial genes "grafted" into a DNA with eukaryotic-like transcriptional signals within a cell with basal transcription factors also of eucaryal type.
3. The genes are clustered around *hsp70(dnaK)* and are flanked by non-heat shock-inducible genes (references in Macario et al., 1999), thus affording the opportunity of studying transcription of stress and non-stress genes side by side.
4. The expression patterns of the genes in the locus, and the promoters, transcription-initiation and termination sites, characteristics of the transcripts, and other basic features have been determined (see Sections XI and XII).
5. Transformation vectors and protocols and plasmids containing various wild-type and mutated promoter and promoter regions have been constructed and are available for testing (Conway de Macario et al., 1996; Zmijewski, EC de M, and AJLM, unpublished results).
6. The genome of *M. mazeii* has been sequenced and the data are available (Deppenmeier et al., 2002), which will help in the identification of genes pertinent to transcription regulation, and in their manipulation *in vivo* and *in vitro*.
7. The genome of *Methanosarcina acetivorans* (a closely related species) that has also been sequenced (Galagan et al., 2002) is currently under active investigation—its *hsp70(dnaK)* locus is very similar to that of *M. mazeii*, which will facilitate comparative analyses to elucidate the function, regulation, and mechanism of transcription of stress genes (unpublished results).
8. *M. mazeii* is unique among archaea, even among other species of the same genus, in that it grows in at least three different morphotypes (Mayerhofer et al., 1992; Yao et al., 1992), two of which are multicellular structures of a complexity that suggests they are the result of differentiation development (Macario and Conway de Macario, 2001); because of this, *M. mazeii* provides a model system to study the role of stress and anti-stress mechanisms in cell differentiation and primitive histogenesis (Shimkets and Brun, 2000).

II. TBP

A. Eucaryal vs. Archaeal TBPs

Eucaryal TBPs are basic proteins with isoelectric points (IPs) within the range 9.8 to 10.7, while the archaeal counterparts are acidic with IP ranging between 3.9 and 6.1, with two exceptions, whose IPs are 8.4 (Table 1). In addition, there are other differences between eucaryal and archaeal TBPs; for example, the latter are shorter on the average, mostly because they do not have the N-terminal domain (NTD) that most

TABLE 1
Properties of Archaeal and Eucaryal TBPs, TFBs, and TFIIBs

Organism	TBP				TFB (archaea); TFIIB (eucarya)			
	Acc. No.[a]	aa[b]	MM kDa[c]	IP[d]	Access No.	aa	MM kDa	IP
Methanosarcina mazeii	aj243798[e]	183	19.8	4.7	aj243755[e]	337	38.3	9.6
Methanosarcina mazeii	MM1027, #1[f]	183	19.8	4.8				
	MM1028, #2[f]	183	19.8	4.7				
	MM2187, #3[f]	185	20.2	4.7				
Methanococcus thermolithotrophicus	aj271331[e]	181	20.0	4.2	aj271467[e]	339	38.4	10.2
Methanosarcina barkeri[g]	None, #1	183	19.8	4.7	None	337	38.3	9.7
	None, #2	184	20.3	4.6	n.a.[h]			
Methanosarcina acetivorans	aam07674, #1	183	19.9	4.8	aam04054	337	38.3	9.6
	aam03731, #2	185	20.0	5.2				
	aam03632, #3	185	20.3	7.1				
Methanobacterium thermoautotrophicum	o27664	181	19.7	4.6	q26971	310	34.9	8.7
Methanococcus jannaschii	q57930	183	20.2	4.5	q58192[i]	673	77.0	9.5
Methanopyrus kandleri	None	185	20.7	4.5	aam02071, #1	307	34.3	9.6
					aam01834, #2	341	39.4	9.2
Thermoplasma acidophilum	q9hlm8	184	20.5	4.5	q9hjm7, #1	312	35.1	10.3
	n.a.				q9hjm2, #2	307	34.5	10.0
	n.a.				cac12149, #3[j]	228	25.9	10.5
Thermoplasma volcanium	bab60562	184	20.5	4.5	bab60249, #1	312	35.3	10.4
	n.a.				bab60255, #2	313	35.2	10.0
	n.a.				bab59711, #3[j]	224	25.1	10.7
Thermococcus celer	q56253	189	21.3	4.9	n.a.			
Ferroplasma acidarmanus[h]	None	185	20.5	4.6	None, #1	311	35.3	10.2
	n.a.				None, #2	305	34.2	9.9
Pyrococcus woesei (*furiosus*)	q57050	191	21.3	4.7	q51731	300	34.1	10.3
Pyrococcus (Thermococcus) kodakaraensis	q52366	190	21.5	4.6	p58109	306	34.5	10.8

TABLE 1 (continued)

Pyrococcus horikoshii	o58737	191	21.4	4.8	o59151, #1	300	34.1	10.3
	n.a.				o58594, #2[j]	208	23.9	10.9
Pyrococcus abyssi	q9v024	191	21.4	4.8	q9v0v5	300	34.1	10.3
Pyrobaculum aerophilum	None	199	22.1	4.5	None, #1	298	32.1	8.6
	n.a.				None, #2	333	37.3	10.3
	n.a.				None, #3[j]	159	17.9	10.6
Archaeoglobus fulgidus	o29874	183	20.1	8.4	o28970	326	36.9	9.5
Aeropyrum pernix	q9yat1	189	21.1	5.5	q9y942	325	35.7	9.0
Sulfolobus shibatae	q55031	198	22.3	6.1	p50387	309	34.8	10.1
Sulfolobus acidocaldarius	q53648	198	22.5	5.5	q9uwn6	309	34.9	9.5
Sulfolobus solfataricus	p58178	198	22.3	6.1	p58111, #1	309	34.7	10.0
	n.a				p58110, #2	293	33.3	10.0
Sulfolobus tokodaii	bab66317	166	18.9	8.4	bab65329, #1	308	34.7	9.7
	n.a.				bab66313, #2	295	33.3	10.0
Halobacterium sp. NRC-1	q48325	186	19.9	3.9	q9hsf7	323	36.6	5.7
Halobacterium (Haloferax) volcanii	n.a.				q9yga5	332	37.3	5.2
Drosophila melanogaster	p20227	353	38.5	10.1	p29052	315	34.4	8.3
Caenorhabditis elegans	p32085	340	36.6	10.4	aaa68790	759	84.0	6.7
Homo sapiens	p20226	339	37.7	10.6	q00403	316	34.8	8.3
Xenopus laevis	p27633	297	32.7	10.6	p29054	316	34.7	8.4
Saccharomyces cerevisiae	p13393	239	26.9	10.2	p29055	345	38.2	8.7
Arabidopsis thaliana	p28147	200	22.4	10.5	p48512	312	34.3	6.8
Glycine max	q42808	200	22.4	10.5	t06440	313	34.2	6.5
Mus musculus	p29037	316	34.7	10.6	n.a.			
Acanthamoeba castellanii	p26354	258	28.4	9.8	n.a.			
Tetrahymena thermophila	q27850	246	28.2	10.7	n.a.			
Pneumocystis carinii	q12652	229	25.2	10.5	n.a.			
Plasmodium falciparum	p32086	228	26.2	9.8	n.a.			
Dictyostelium discoideum	p26355	205	22.9	10.6	n.a.			
Triticum aestivum	q02879	201	22.4	10.4	n.a.			

TABLE 1 (continued)

Acetabularia cliftonii	p46272	191	21.7	10.3	n.a.			
Rattus norvegicus	n.a.				p29053	316	34.8	8.3

[a]Accession number. Archaeal and eucaryal molecules above and below the horizontal double line, respectively.
[b]Total amino acids.
[c]Molecular mass rounded off to the nearest hundredth of a kilodalton (kDa).
[d]Isoelectric point rounded to the nearest tenth.
[e]These are GenEMBL accession numbers and all the others belong to the protein databases SwissProt, PIR, and GenPept.
[f]Accession numbers are not available, these are the gene numbers assigned to these proteins for this organism; information was taken from http://www.ncbi.nlm.nih.gov
[g]Information obtained from http://www.jgi.doe.gov/tempweb/JGI_microbial/html/index.html
[h]Not available in the databases searched.
[i]The sequence used in these comparisons does not include the intein present in the original SwissProt entry.
[j]These molecules are shorter than the average archaeal TFB, lack the N-termimal domain (NTD) and, except for *P. horikoshii* #2, have no recognizable helix-turn-helix (HTH) motif. They were included here because they: a) are annotated in the databases as transcription initiation factor B (TFB) (*T. volcanium*), transcription initiation factor IIB-related protein (*T. acidophilum*), and conjectural or hypothetical transcription initiation factor IIB (*P. horikoshii*, and *P. aerophilum*); and b) aligned (PILEUP) with the C-terminal portion of the other archaeal TFBs. (See also Plate 3.)

eucaryal molecules possess. In eukaryotes and in archaea, TBP interacts with the minor groove of the DNA: it binds the promoter, distorting and unwinding the double helix via the insertion of phenylalanines between the first and last two base pairs of the TATA box (Kim et al., 1993b; DeDecker et al., 1996; Guzikevich-Guerstein and Shakked, 1996; Kosa et al., 1997; Soppa, 1999; Lee and Young, 2000).

B. Structure

1. N-terminal Domain (NTD)

Archaeal TBPs lack the NTD present in their eukaryotic counterparts. The function of the NTD is not fully understood, or is the NTD well conserved among eukaryotes—a relatively better degree of conservation occurs in the TBPs from vertebrates (Soppa, 1999) in which it is involved in the regulation of transcription carried out by RNA polymerase (RNAP) III (references in Hampsey, 1998). The fusion of the *Drosophila* TBP NTD with a GAL4 DNA-binding domain revealed that NTD by itself is a strong activator of transcription, but in the intact TBP molecule it is masked (Um and Manley, 2000). The low degree of conservation of NTD indicates that the activator function ascribed to the NTD of *Drosophila* TBP is not a property shared by all eucaryal TBPs, and it does not give clues on what could be its equivalent in archaeal TBPs if such equivalent exists at all.

2. Repeats

Archaeal and eucaryal TBP molecules are similar in that both have two copies of an imperfect, direct, repeat, and adopt a symmetrical saddle-shaped form (Nikolov

et al., 1992; Marsh et al., 1994; DeDecker et al., 1996; Kosa et al., 1997), Figure 1A. The degree of similarity between the two repeats is greater in archaea than in eukaryotes (Soppa, 1999; Thomsen et al., 2001). These repeats divide TBP—excluding the N-terminal extension present in the eucaryal molecules—into two halves or domains.

Sequence alignment comparisons revealed that the bulk of the amino acid (aa) sequence similarity between the two repeat-containing halves of the molecule lies within the C-terminal portions of each repeat, whereas the N-terminal portions share considerable less similarity (data not shown). These regions or segments of higher similarity are relatively short, comprising slightly more than half of each repeat (Table 2). These segments are separated by another of approximately 50 aa in length that contains a small portion of the first repeat's C-terminus, the interrepeat region, and the N-terminal portion of the second repeat (Figure 1B). The three segments are also discernable in the eukaryotic TBPs, although not quite as clearly as in the archaeal molecules. Interestingly, the three segments in archael TBPs show disctinctive isoelectric points (IP) (see below, Section II.C), suggesting that they might differ, at least to some extent, in function and/or mechanism of action.

These observations suggest that the core of the archaeal TBP, and possibly that of the eukaryotic counterpart, contains three segments (N-terminal, intermediate, and C-terminal) that are discernable over the two-repeat region (Figure 1B). Interestingly, the shortest archaeal TBP (166 aa; Table 1) with identifiable repeats, found in *Sulfolobus tokodaii* (Kawarabayasi et al., 2001), has the three segments complete, despite the fact that it lacks a portion of the first repeat. This would suggest that the three segments described here are essential for function. However, it is intriguing that TBPa, one of the six TBPs found in *Halobacterium* sp. NRC-1, has only 100 aa and lacks the N-terminal portion containing the first repeat that is present in all other known TBPs, archaeal and eukaryotic (Ng et al., 1998). Consequently, TBPa lacks the N-terminal segment and has only a fraction of the intermediate segment, while the C-terminal segment is conserved. The function of this very short TBP has not yet been elucidated, but, together with that from *S. tokodai*, it offers an opportunity for segmental dissection of TBP evolution and function by comparison with the other TBPs that possess the complete repeats.

C. Physicochemical Properties

The segments described in the preceding section have distinctive physicochemical properties, as demonstrated, for example, by their different isoelectic points (IPs) (Table 3). The C-terminal segment is acidic, except for those in the TBPs from *S. tokodaii* (Kawarabayasi et al., 2001) and *P. aerophilum* (Fitz-Gibbon et al., 2002), which are close to neutrality with IPs of 6.7 and 6.5, respectively. In contrast, the IPs of the N-terminal and intermediate segments vary among species. Nevertheless, some patterns are discernable. At least, either the N- or C-termimal segment has an IP that is different from that of the intermediate segment, excepting the TBPs from *M. thermolithotrophicus*, *T. acidophilum*, *T. volcanium*, and *Halobacterium* sp. NRC-1. In 9 (31%) of the 29 archaeal proteins examined, the intermediate segment is more basic than the other two segments, which results in the acidic/basic/acidic pattern. In *M. mazeii* S-6 (Thomsen et al., 2001), *M. barkeri Mb*TBP1, and *A. fulgidus* TBPs the pattern is basic/near neutral/basic. The TBP from *P. aerophilum* is unique in that it displays a basic/neutral/near neutral (but on the acidic side) pattern. In contrast, the corresponding segments in the eucaryal TBPs have basic IPs ranging between 8.4 and 11.4. The very short TBPa

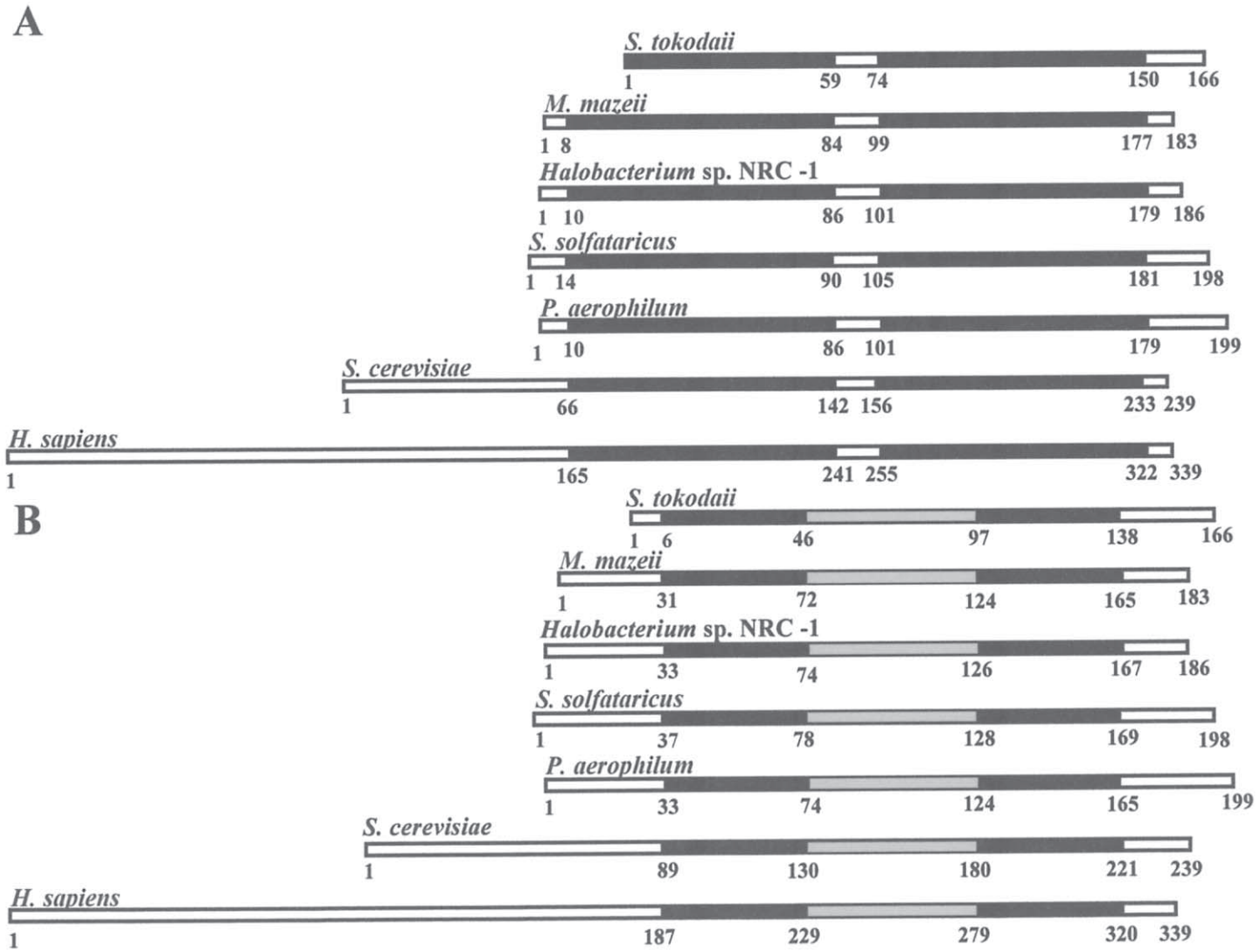

FIGURE 1. Schematic of TBPs of five representative archaeal and two eucaryal molecules. Each horizontal bar represents a TBP molecule, with the name of the respective organism on top. Black sections show in **A** the repeats and in **B** the N- and C-terminal segments of highest similarity separated by a less conserved region shown in gray. The numbers underneath the bars indicate the position number at which the protein**,** repeats, and segments begin and end (N-terminus to the left). The archaeal TBPs displayed include the shortest and the longest of those listed in Table 1 from *S. tokodaii* and *P. aerophilum*, respectively.

TABLE 2
Percent Identity and Similarity of Repeats and Most Conserved Segments and Intervening Regions in Archaeal and Eucaryal TBPs

Organism	Repeats		Segments	
	Identity	Similarity	Identity	Similarity
S. solfataricus	51.9	63.6	61.9	78.6
S. shibatae	51.9	63.6	61.9	78.6
S. tokodaii	61.0	78.0	54.2	67.8
S. acidocaldarius	49.4	62.3	59.5	78.6
P. aerophilum	59.5	66.7	54.5	59.7
P. abyssi	57.1	71.4	46.8	61.0
A. pernix	48.0	59.7	61.9	76.2
P. woesei (furiosus)	44.2	58.4	54.8	69.0
P. horikoshii	44.2	61.0	54.7	71.4
P. (T.) kodakaraensis	46.8	62.3	52.4	69.0
T. celer	46.7	61.0	54.8	69.0
M. thermolithotrophicus	51.3	63.2	69.0	73.8
M. jannaschii	51.9	63.6	69.0	71.4
T. volcanium	57.1	73.8	49.3	65.3
T. acidophilum	48.1	62.3	59.5	73.8
F. acidarmanus	49.4	59.7	61.9	69.0
M. mazeii S-6	39.0	48.1	50.0	59.5
M. mazeii Goe1, 1	39.0	48.1	50.0	59.5
M. barkeri 1	40.3	49.4	54.8	61.9
M. acetivorans 1	39.0	49.4	52.4	61.9
M. mazeii Goe1, 2	43.3	53.7	52.4	61.9
M. acetivorans 2	38.5	48.7	47.6	57.1
M. barkeri 2	36.4	46.8	52.4	57.1
M. acetivorans 3	33.8	45.5	40.5	50.0
M. mazeii Goe1, 3	39.6	49.1	47.6	54.7
A. fulgidus	48.1	59.7	54.8	59.5

TABLE 2 (continued)

Halobacterium sp. NRC1	41.6	53.2	50.0	57.1
M. thermoautotrophicum	57.1	70.1	66.7	78.6
M. kandleri	52.6	63.2	61.9	71.4
H. sapiens	31.2	37.7	42.9	45.2
M. musculus	31.2	37.7	42.9	45.2
X. laevis	31.2	37.7	42.9	45.2
G. max	32.5	41.6	40.6	43.8
T. aestivum	33.8	44.2	45.2	52.4
A. thaliana	35.1	41.6	45.2	47.6
A. castellanii	35.1	42.9	45.2	47.6
A. cliftonii	31.2	40.3	42.9	45.2
P. carinii	32.5	41.6	42.9	47.6
S. cerevisiae	33.6	41.6	42.9	47.6
D. discoideum	29.9	40.3	40.5	47.6
D. melanogaster	35.1	41.6	42.9	45.2
C. elegans	32.5	36.4	42.9	42.9
T. thermophila	35.1	45.5	45.2	50.0
P. falciparum	32.9	42.1	31.7	41.4

[a] The complete names and accession numbers of the organisms, archaea above and eucaryal below the horizontal double line, respectively, are given in Table 1.

TABLE 3
Isoelectric Points of Most Conserved Segments and Intervening Region in Archaeal and Eucaryal TBPs

Organism[a]	Segment			Whole protein
	N-terminal	Intermediate	C-terminal	
S. solfataricus	4.88	11.39	4.38	6.12
S. shibatae	4.88	11.39	4.55	6.12
S. tokodaii	4.46	11.26	6.68	8.38
S. acidocaldarius	4.83	11.34	4.42	5.52
P. aerophilum	8.42	4.86	6.51	4.52
A. pernix	9.32	10.05	4.37	5.50
P. abyssi	7.29	10.19	4.88	4.78
P. woesei (furiosus)	5.67	10.60	4.93	4.86
P. horikoshii	6.46	10.60	4.93	4.77
P. (T.) kodakaraensis	4.77	9.26	4.93	4.63
T. celer	5.75	9.35	5.77	4.87
M. thermolithotrophicus	4.93	3.92	4.93	4.18
M. jannaschii	6.47	4.12	4.08	4.48
T. acidophilum	4.59	4.16	4.19	4.51
T. volcanium	4.55	4.16	4.19	4.50
F. acidarmanus	6.56	4.33	4.19	4.63
M. mazeii S-6	9.35	7.69	4.21	4.66
M. mazeii Goe1, 1	9.35	9.42	4.21	4.75
M. barkeri 1	9.35	7.69	4.17	4.83
M. acetivorans 1	9.35	9.42	4.21	4.75
M. mazeii Goe1, 2	9.35	9.42	4.21	4.72
M. acetivorans 2	10.07	7.46	4.25	5.16
M. barkeri 2	10.01	4.76	4.21	4.62
M. acetivorans 3	10.61	10.39	4.59	7.11
M. mazeii Goe1, 3	10.44	5.67	4.37	4.67
A. fulgidus	10.33	6.73	4.32	8.36

TABLE 3 (continued)

Halobacterium sp. NRC-1	4.83	3.92	3.66	3.90
M. thermoautotrophicum	6.53	4.93	4.17	4.63
M. kandleri	4.93	4.26	4.50	4.54
H. sapiens	10.79	11.07	10.66	10.61
M. musculus	10.79	11.07	10.66	10.61
X. laevis	10.79	11.07	10.66	10.64
G. max	10.80	10.64	10.15	10.47
T. aestivum	10.96	10.91	10.25	10.42
A. thaliana	10.63	10.91	10.15	10.45
A. castellanii	10.63	11.29	8.41	9.82
A. cliftonii	10.63	10.82	8.41	10.30
P. carinii	10.63	10.82	10.15	10.50
S. cerevisiae	10.63	10.92	10.15	10.16
D. discoideum	10.80	10.80	10.09	10.56
D. melanogaster	10.34	10.80	10.84	10.08
C. elegans	10.79	10.35	10.83	10.38
T. thermophila	10.22	10.58	10.43	10.66
P. falciparum	11.38	10.26	9.37	9.84

[a] The complete names and accession numbers of the organisms, archaea above and eucarya below the horizontal double line, respectively, are given in Table 1.

from *Halobacterium* sp. NRC-1, referred to above, has an overall IP of 3.99, and the C-terminal segment, and the portion of the intermediate segment that are present in this molecule have IPs of 3.87 and 3.70, respectively. Thus, this molecule is exceptional because of its very short length by comparison with all the others known, and because it does not possess the complete copies of the typical TBP repeat or the three segments described in this work.

D. DNA Binding

Several parameters of the interaction between eukaryotic TBP and DNA, including specific aa-base contacts, and interactions between the protein and the phosphate backbone have been studied (Kim et al., 1993a; Pastor and Weinstein, 1995; Nikolov et al., 1996). In addition, certain character-

istics of TBP-DNA interactions have been identified recently that seem unique to archaea. The binding of *P. woesei* TBP (*Pw*TBP) to DNA was tested at various salt concentrations, and the results showed that binding was enhanced at the higher concentrations tested (DeDecker et al., 1996; O'Brien et al., 1998; Bergqvist et al., 2001). It was hypothesized that the binding enhancenment results from integration of cations by *Pw*TBP into the protein-DNA interface and the simultaneous release of water molecules into the solvent (Bergqvist et al., 2001). Mutational analyses were performed to determine the aa that participate in cation uptake, and it was established that a glutamic acid (E) at position 12 (E12) is a major protagonist (Bergqvist et al., 2001). Mutants of *Pw*TBP were generated in which E12 was replaced by the basic aa lysine (E12K mutant) or by the neutral aa alanine (E12A mutant). The DNA binding constants of the mutant and wild-type proteins were measured at a range of salt concentrations. The binding constant of the E12K mutant was lower, and that of E12A was higher than that of the wild type over the entire range of salt concentrations.

The buffer needed for the binding of *M. mazeii* TBP (Thomsen et al., 2001) to DNA must contain Mg^{2+} within a restricted range of concentrations (Macario and Conway de Macario, unpublished results), demonstrating that cations at particular concentrations are required for TBP-DNA binding also in a mesophile like *M. mazeii*. It is not known whether the mechanism by which cations influence TPB-DNA binding is the same in the mesophile *M. mazeii* as that operating in the hyperthermophile *Pw*TBP.

Multiple sequence alignments of archeael TBPs were generated, and it was found that the TBPs of hyperthermophiles contain an E that matches the *Pw*TBP E12 (Bergqvist et al., 2001). Recently, we found that TBPs from mesophilic archaea also contain an E12 equivalent. Multiple sequence alignments of a large sample of known archaeal TBPs and several eucaryal TBPs were generated using the GCG program PILEUP and manually (Plate 1*). The results showed that the majority of the archaeal TBPs have an E that aligns with the E12 of *Pw*TBP; exceptions are the TBPs from *M. jannaschii* and *M. thermolithotrophicus* that have valine (V) at this position; the TBP from *Methanopyrus kandleri* (Slesarev et al., 2002) that has a glutamine (Q) at this position; and the shortest TBP from *S. tokodaii* that lacks the portion where the equivalent position would have been. Glutamic acid is a negatively charged acidic aa, while V is neutral, indicating that they cannot substitute for one another in such a key position of the protein without altering its properties. A careful examination of the multiple sequence alignment revealed that the *M. jannaschii* and *M thermolithotrophicus* proteins have an E only four positions upstream of the V placed by the alignment to match the *Pw*E12 position (Plate 1), and the short TBP from *S. tokodaii* has E close to its N terminus (position two). Although PILEUP did not align these Es with the key E present in the other archaeal proteins, it is possible that they carry out the function that E12 performs in *P. woesei*. Another possibility is that the molecules lacking E12, or a functionally equivalent aa in the appropriate position, are not as sensitive to salt concentration as those that have it; if they were, it would be via a different mechanism not involving E12. In any case, experimental data show that archaeal TBPs, such as *Pw*TBP, require high cation concentrations to bind to DNA (Bergqvist et al., 2001). This requirement may consti-

* Plates appear following page 226.

tute a mechanism of transcription regulation in which cations play a central role, a possibility that ought to be explored.

Eukaryotic TBPs as well as *M. kandleri* TBP have Q at the position aligned with the *Pw*E12 residue (Plate 1). Interestingly, another aa demonstrated to influence DNA binding within *Pw*TBP is the Q at position 103 (Q103; Plate 1). Replacement of this aa with an alanine (A) resulted in an increase of the molecule's binding affinity for DNA. Q is a polar aa and cannot capture cations; therefore, it was suggested that the higher DNA-binding affinity of the mutant Q103A was due to steric changes in the mutant TBP (Bergqvist et al., 2001). Multiple sequence alignments revealed that Q103 is conserved throughout species of the eucaryal and archaeal phylogenetic domains, with the exception of *P. falciparum* that contains E instead of Q at the matching position (Plate 1).

The E12 aa was the first of four Es of an archaeal TBP that was demonstrated to influence DNA binding through the uptake of cations. Subsequently, the other three aa, E41, E42, and E128, were identified in *Pw*TBP, which also influences this transcription factor's ability to bind DNA in a cation dependent manner (Bergqvist et al., 2002). All four of these Es have a combined effect on *Pw*TBP's ability to bind DNA. E12 is the most conserved of these four aa throughout archaea, being present in all but four of all the molecules examined (Plate 1). E128 is conserved in all the archaeal TBPs examined except those belonging to the genus *Sulfulobus* and to *Pyrobaculum aerophilum*, which have M instead of E at this position, and to *Aeropyrum pernix* TBP, which has an L. In contrast, this position is not conserved in eukaryotic TBPs (Plate 1).

E41 and E42 are the least conserved, with both of them found together only in the molecules belonging to *Thermococcus celer* and the *Pyrococcus* genus (Plate 2*). The aa that align with these two Es in the other TBPs varies from genus to genus. The *Sulfolobus* TBPs as well as the TBPs from *A. pernix* and *P. aerophilum* have a D and a Q that aligns with E41 and E42, respectively. The *Methanosarcina* TBPs all have a basic K that aligns with one of these Es. The TBP examined from *Halobacterium* sp. NRC-1 has an E that aligns with E41 and a D that aligns with E42. The remaining TBPs all have one E and one Q that align with *PwTBP* E41 and E42, respectively. Therefore, it is likely that different combinations of presence and absence of these Es might influence the DNA-binding capabilities of these archaeal TPBs. One must remember, however, that an alignment reflects neither structure nor function, so it is possible that DNA binding is influenced by other aa in TBPs from species outside of the *Pyrococcus* genus. None of the eucaryal TBPs examined have the two aa corresponding to E41 and E42, but instead they often have the basic aa K and R, respectively, at these positions.

E. Stability of TBP from Thermophilic Organisms

P. woesei is a thermophilic archaeon with an optimal temperature for growth (OTG) of 95 to 100°C. The stability of its TBP (*Pw*TBP) at such a high temperature might be due, at least in part, to the formation of a disulfide bond between two cysteine (C) residues located at positions 33 (C33) and 48 (C48) (DeDecker et al., 1996), although the reasons why molecules from hyperthermophiles can survive and work at temperatures that are considered extreme compared to what is optimal for human cells are not yet fully understood

* Plates appear following page 226.

(Scandurra et al., 2000). All of the archaeal and eucaryal TBP sequences listed in Table 1 were examined by comparing them using PILEUP and manual alignments to see if they contained these two Cs. Four archaeal sequences have Cs corresponding to C33 and C48 like in *Pw*TBP (Plate 2); these TBPs are from the hyperthermophiles *P. abyssi* (http://www.genoscope.cns.fr/Pab/), *P. horikoshii* (Kawarabayasi et al., 1998), *P. kodakaraensis* (Rashid et al., 1995), and *T. celer* (Marsh et al., 1994) OTGs of 103, 95-100, 95 – 100, and 85°C, respectively. In contrast, *A. pernix*, *S. tokodaii*, and *P. aerophilum* TBPs do not contain these two Cs despite the fact that these species are also hyperthermophilic (OTGs 90, 80, and 100°C, respectively) (Plate 2). The TBPs from *M. thermolithotrophicus* and *M. jannaschii* with OTGs of 65 and 85°C, respectively, have a single C corresponding to *Pw*C48. These data suggest that the presence of C33 and C48 is more a characteristic of *Pyrococcus* and *Thermococcus* species than of organisms with high (60 to 80°C) or very high (85°C or higher) OTG.

The isoelectric points of the segments identified within the two repeats of TBP show that the acidic/basic/acidic pattern is a common feature of the molecules belonging to organisms of the *Pyroccocus* and S*ulfulobus* genera. Perhaps this distribution of charges contributes to the thermostability of these proteins. This pattern, however, is not apparent in the TBPs from *A. pernix* or *P. aerophilum*, which also have high OTGs. Therefore, thermostability of extremophilic TBPs may be due to a factor yet to be identified, or to a combination of factors, such as the presence of disulfide bonds and charge distribution on the molecule. These factors individually may not play a significant role in thrermostability, but together they might act synergistically to endow TBP with the capacity to endure and function at high temperature (Scandurra et al., 2000; DeDecker et al., 1996).

F. Number of *tbp* Genes per Genome

The genome of *Halobacterium* sp. NRC-1 has five *tbp* genes (Ng et al., 2000). *Methanosarcinae* contain multiple TBPs also. *M. mazeii* and *M. acetivorans* both have three *tbp* genes (Galagan et al., 2002; Deppenmeier et al., 2002). For clarity these proteins will be referred to by their assigned gene numbers: *M. mazeii* Goe1: MM1027, MM1028, and MM2184; and *M. acetivorans*: MA4331, MA0278, and MA0179 (accession numbers: aam07674, aam03731, and aam03632, respectively; see Table 1). BLAST searches of the *M. barkeri* genome database (http://www.jgi.doe.gov/tempweb/JGI_microbial/html/) indicated that this organism has at least two *tbp* genes—the data are preliminary and the information must be considered with caution. The two BLAST hits suggesting two *tbp* genes were available only as nucleotide sequences and were retrieved from the database as such to be translated into aa sequences using the GCG program TRANSLATE. They are referred to as *Mb*TBP1 and *Mb*TBP2.

The TBP from *M. mazeii* S-6 shown in Tables 1 to 3 differs by one aa from the *M. mazeii* Goe1 MM1027; for comparative analyses the S-6 sequence was used together with TBP 2 (MM1028) and TBP 3 (MM2187) from *M. mazeii* Goe1 (Plates 1 and 2), or alone (Figure 1).

III. TFB

A. DNA Contacts

Archaeal TFB makes sequence-specific contacts with the DNA upstream of the

TATA box. Mutational and deletional analyses of the T6 viral promoter from the archaeon *S. shibatae* revealed that promoter strength was influenced in part by the interaction of TFB with a stretch of about six base pairs (bp) upstream of the TATA box (Qureshi and Jackson, 1998; Bell et al., 1999b). This region is the TFB-responsive element or TFB-recognition element (BRE). DNA binding experiments using the electrophoretic mobility-shift assay (EMSA) showed that TFB by itself did not bind to the T6 promoter, and that TBP by itself bound only weakly to the TATA box. The complex formed by these two transcription factors, however, did associate with the T6 promoter with high efficiency, suggesting that these factors must work together to bind the promoter in a functionally significant manner. Other experiments revealed that TBP alone did not select for specific base sequences other than the TATA box, although TBP complexed with TFB did. Results obtained with *M. mazeii* cell lysates indicate that cytosolic TBP-TFB complex binds to stress gene promoters, and that this binding is enhanced by heat stress (De Biase et al., 2002; see also Section XII.E.2).

B. Role in TBP-TFB Complex Orientation

Archaeal TFB, like eucaryal TFIIB, plays a role in the orientation of the transcription initiation complex (Bell et al., 1999b). The results of experiments involving the manipulation of DNA sequences flanking the TATA box of the *efl-alpha* promoter from *P. woesii* demonstrated that the orientation of transcription is strongly influenced by the binding of TFB to the BRE—alluded to in section III.A and explained in more detail in section VI.E. Both eucaryal TFIIB and archaeal TFB interact with the BRE through helix-turn-helix (HTH) motifs that are present within the C-terminal portions of these proteins, and mutations of key motif aa disrupt this binding (Lagrange et al., 1998; Littlefield et al., 1999). The motif is more conserved among archaeal TFBs than among the eucaryal TFIIBs (Thomsen et al., 2001) (Plate 3*). Several archaeal species have more than one TFB (Table 1). Among these species, *T. acidophilum*, *T. volcanium*, and *P. aerophilum* have one of their TFBs without a detectable HTH motif (unpublished searches using the GCG program MOTIFS) that also lacks the NTD. Because the HTH motif is necessary for interaction with the BRE, and the NTD is required for RNAP recruitment (see following section), one wonders what are the functions of these atypical TFBs (see box SHORT in Plate 3).

C. Role in RNAP Recruitment

The N-terminal domain (NTD) of archaeal TFB is, like that of the eucaryal TFIIB, essential for the recruitment of RNAP (Bell and Jackson, 2000a). This was demonstrated using a mutant *S. acidocaldarius* TFB, *Sa*TFBc, lacking the NTD. *Sa*TFBc formed a more stable ternary complex with TBP and DNA than did its wild-type counterpart; however, the complex with the mutant failed to recruit RNAP *in vitro*.

Recently, it was found that the TFB from *S. solfataricus* specifically interacts with subunit K (RpoK) of RNAP and, to a much lesser extent, with subunit E (Magill et al., 2001). The separation of the N-terminal and C-terminal domains of *S. solfataricus* TFB not only demonstrated that NTD is the portion of the molecule that interacts with the K subunit, but also that the NTD by itself associates with the RNAP subunit more strongly than does the full-length transcription factor.

* Plates appear following page 226.

D. Zinc Ribbon

The NTDs of TFB and TFIIB contain a zinc ribbon that functions in the interaction of the transcription factor with archaeal RNAP and eucaryal RNAPII, respectively (references in Zhu et al., 1996; Chen et al., 2000). The substitution of residues C31 and L38 in the zinc ribbon of *S. solfataricus* TFB for an alanine and a proline, respectively, hindered the binding of this transcription factor to RNAP, demonstrating the functional importance of this zinc-containing structure in archaea (Magill et al., 2001). It has been suggested that *beta* sheets in TFIIB zinc ribbon interact with RNAPII, while the binding of the active aa to zinc stabilizes the sheets (Chen et al., 2000).

The four metal binding aa in the zinc ribbon of eucaryal TFIIBs are three cysteines (C) and one histidine (H), or four Cs (Chen et al., 2000). As mentioned previously, archaeal TFB also has a zinc ribbon, but it is less conserved than that of eukaryotic TFIIB (Plate 4). A detailed examination of the putative metal-binding structure in the TFB from S. *shibatae* revealed that the first and last Cs are replaced by a serine (S) and a threonine (T), respectively (Qureshi et al., 1995). This observation prompted the suggestion that S and T might also support the metal-binding capacity of the structure because the oxygen atoms in these aa can interact with metal ions.

An updated multiple sequence alignment of TFBs and TFIIBs generated by the GCG program PILEUP showed that the zinc-binding structures of TFBs from *S. solfataricus* (TFB1), *S. acidocaldarius*, *S. tokodaii* (TFB1), and *A. pernix*, in addition to that from *S. shibatate*, all have either S or T instead of the first and the last Cs typical of canonical zinc fingers (Plate 4*). Interestingly, the genomes of *S. solfataricus* (She et al., 2001) and *S. tokodaii* (Kawarabayasi et al., 2001) have a second *tfb* gene encoding a second TFB (TFB2) that contains all four metal-binding Cs in its zinc ribbon. Perhaps the two *S. solfataricus* TFBs perform different functions or recognize different promoters. It is also possible that only one of these *tfb* genes is functional, while the other is a pseudogene.

The sequences of the *S. shibatae* and *S. acidocaldarius* genomes are not yet available, and one can only hypothesize, but it seems reasonable to predict that these two organisms have a second TFB with four Cs in their zinc ribbons, just like the second TFB from *S. solfataricus*, because all three organisms belong to the same genus.

The genome sequence of *A. pernix* (Kawarabayasi et al., 1999) shows only one *tfb* gene encoding the TFB molecule mentioned above, in which the second C of the zinc ribbon does not align with the corresponding Cs of other archaeal TFBs (Plate 4). The genome of *P. horikoshii* has two *tfb* genes and those of *T. acidophilum*, *T. volcanium*, and *P. aerophilum* each have three. The zinc ribbon is present in the encoded TFBs except for one in each organism. These exceptions lack the NTD and, consequently, do not have a zinc ribbon, implying that the molecules are unable to bind RNAP (see box SHORT in Plate 3).

E. Promoter Cearance and aa E46

In addition to the *S. acidocaldarius* TFB mutant (*Sa*TFBc) described in Section III.C, a second mutant, *Sa*TFB-E46K, was generated in which E46 (the E residue at the 46th position of the protein), conserved in both TFBs and TFIIBs, was replaced with a lysine (K) (Bell and Jackson, 2000a). This mutant was able to mediate transcription from some promoters but not from others, the

difference being attributed, at least in part, to differences in the sequences near the respective transcription-initiation sites. The *Sa*TFB-E46K mutant, despite the aa substitution, formed a ternary complex with TBP and DNA and recruited RNAP as efficiently as the wild-type molecule, regardless of whether it was able to mediate transcription. Because the E46K mutation did not hinder either preinitiation or open complex formation, it was hypothesized that it somehow inhibits promoter clearance. Experiments were performed in which the availability of NTPs for polymerization in the presence of the mutant protein was limited. At high NTP concentrations the mutant and wild-type TFBs mediated equal levels of transcription. When the NTP concentration was lowered, less transcript was produced by reaction mixtures containing the mutant transcription factor than by those containing the wild type. This suggested that without sufficient amounts of NTP to propel polymerization forward, the stability of the preinitiation complex (PIC), including the mutant TFB, cannot be overcome and transcription elongation does not occur. Perhaps the mutation strengthens the interaction between TBP and RNAP making it more difficult for RNAP to dissociate from the transcription factor and proceed with elongation. It is also possible that E46 is involved in triggering the dissociation of RNAP from the PIC, and mutation of this residue interferes with this process. It was proposed that different nucleotide sequences near the promoter may form stable interactions with RNAP, also interfering with promoter clearance, which may be exacerbated by the E46K mutation and thus explain the sensitivity of some promoters but not others to the mutation (Bell and Jackson, 2000a).

The examination of TFIIB/TFB multiple sequence alignments revealed that one of the two TFBs from *S. solfataricus* and the TFB from *A. pernix* contain aspartic acid (D) in place of the key glutamic acid (E46). Interestingly, the TFB from *S. solfataricus* (*Ss*TBP 2) with this D46 is the same molecule that has a canonical zinc ribbon with four Cs (Plate 4). It is likely that the D46 residues in the *S. solfataricus* and *A. pernix* TFBs are able to perform the same function as the E in the other TFBs and in the TFIIBs. The two TFBs from *S. tokodaii*, and those from *P. aerophilum* and *T. volcanium* with the NTD, have an E that corresponds to E46.

F. Regulation of NTD

The presence of the CTD in the *S. solfataricus* TBP hinders the association of the NTD with the subunit K of RNAP (Magill et al., 2001). In solution, the NTD of eucaryal TFIIB interacts with CTD, which results in a "closed" structure that inhibits the transcription factor's ability to mediate transcription (Roberts and Green, 1994). This "closed" conformation can be reopened via interaction with some eucaryal transcriptional activators. It was proposed that a similar mechanism of transcription regulation might occur in archaea (Magill et al., 2001). Deletion of the NTD of *S. acidocaldarius* TFB resulted in that the remaining portion of the factor formed a more stable ternary complex than the wild-type molecule (Bell and Jackson, 2000a). If the NTD and CTD in the same molecule interact with one another, it is possible that NTD inhibits the association of CTD with TBP and DNA, just as CTD inhibits the association of NTD with RNAP.

G. Intein-Containing TFB

The TFB from *M. jannaschii* (Table 1) is unique in that it contains an intein (Bult

et al., 1996; Thomsen et al., 2001). Inteins are portions of a peptide that are not integral parts of the final molecule, but rather are excised posttranslationally (references in Perler et al., 1994). The remaining portions of the peptide, which are referred to as exteins, are spliced to form the mature protein. The primary, unprocessed TFB peptide from *M. jannaschii* contains 673 aa, of which the first 99 at the N-terminus include the zinc ribbon and the conserved E46. The intein begins at position 100 and continues up to position 434. The TFB's C-terminal portion begins at the 435th aa and contains the HTH (Thomsen et al., 2001). The TFB sequence used for comparative alignments in this work did not include the intein.

IV. TFE

A. Archaeal vs. Eucaryal TFE

Archaea have a transcription factor that is the homolog of the *alpha* subunit of eucaryal transcription factor TFIIE (references in Soppa, 1999). Eukaryotic TFIIE is composed of two 56-kDa and two 34-kDa subunits (Inostroza et al., 1991). It is known that the 56-kDa or *alpha* subunits interact with TBP (Maxon et al., 1994) and enhance basal transcription (Yokomori et al., 1998), but the pertinent mechanisms are not yet clear. TFE, the archaeal homolog of the eucaryal TFIIE *alpha* subunit, might also interact with TBP and RNAP to aid in transcription. For instance, in *S. solfataricus*, TFE binds to both TBP and RNAP in the absence of DNA, but not to TFB (Bell et al., 2001).

Potential TFEs were identified in the sequenced archaeal genomes except those from *T. acidophilum* and *T. volcanium*, in which our BLAST and string searches failed to produce hits with the TFIIE *alpha* homologs used as queries. Several of the TFEs we detected in the other species were not annotated as such, but rather as hypothetical or conserved proteins. Because archaeal TFE is not as conserved in sequence as TBP and TFB, its identification is less certain: it may very well be that all archaea have TFE, but it escapes detection by current methods.

TFE is considerably smaller than the TFIIE *alpha* subunit and aligns only with the N-terminal portion of it (Bell et al., 2001). The N-terminal but not the C-terminal portion of TFIIE *alpha* subunit is required for growth at the physiological temperature in yeast, although the latter is also required at low temperatures (Kuldell and Buratowski, 1997); it would be interesting to determine whether TFE is essential for growth in archaea, especially under stressful conditions.

B. Function

TFE is able to increase the amount of transcript produced from certain promoters. The promoters that were most sensitive to this transcription factor were those that, in the presence of TBP, TFB, and RNAP, showed less than optimal interaction with TBP (Bell et al., 2001; Hanzelka et al., 2001). This became apparent when the effects of TFE on six different *S. solfataricus* promoters were examined *in vitro* (Bell et al., 2001). The three promoters that were the most responsive to the presence of TFE contained TATA regions that deviated from the canonical consensus (TTTATATA), specifically when the first or last base was replaced by a C or a G, respectively. Furthermore, mutations within TFE-insensitive promoters, particularly the substitution of the fifth and seventh nucleotides in the TATA region, both Ts, with G, lowered

basal levels of transcription *in vitro*, while at the same time increased the promoter's sensitivity to TFE.

TBP concentration also influenced the sensitivity of certain promoters to TFE. Promoters that were not normally TFE sensitive became so when the concentration of TBP was low, indicating that TFE somehow aided TBP in DNA binding (Bell et al., 2001). DNase footprinting experiments carried out to examine the binding of TFE to DNA, produced results similar to those performed with TFIIE *alpha* subunit (Yokomori et al., 1998): the footprints produced by TBP alone or in the presence of TFE were virtually the same (Bell et al., 2001).

M. thermoautotrophicum TFE stimulated transcription from some but not all methane-gene promoters *in vitro* (Hanzelka et al., 2001). Similar to what was observed with the TFE-sensitive promoters in *S. solfataricus* as discussed above, very little transcript was produced from the *M. thermoautotrophicum* methane-gene promoters in the presence of TBP, TFB, and RNAP, without TFE. In the presence of these same factors and TFE, transcription from these promoters increased. Unlike the TFE-sensitive promoters found in *S. solfataricus*, however, no discernable bases or other DNA elements that would determine TFE sensitivity or insensitivity could be found near to or within these promoters (Hanzelka et al., 2001).

C. Zinc Finger

1. Structure vs. Function

The TFIIE *alpha* subunit contains a zinc finger that is critical to stimulate transcription (Maxon and Tjian, 1994; Yokomori et al., 1998). A zinc finger was identified in the archaeal TFE (Bell et al., 2001; Hanzelka et al., 2001). Of the four cysteines (C) in the finger, three are conserved across archaea and eucarya.

The second C is not conserved in archaea, which prompted the suggestion that the TFEs without this aa do not bind zinc (Bell et al., 2001). It is unclear whether the lack of the second C affects the ability of TFE to interact with TBP or RNAP. The replacement of the third C with alanine (A) abolished the zinc-binding and transcription-enhancing capacities of the *alpha* subunit of *Homo sapiens* TFIIE (Maxon and Tjian, 1994; Yokomori et al., 1998). It was not tested whether these same effects would also be caused by the replacement of the second C rather than the third. Perhaps the second C is necessary for zinc binding but not for stimulating transcription. It is possible that the third C, aside from participating in zinc binding, plays an additional role in recruiting TBP and, thus, in stimulating transcription. It is also possible that the TFEs without the four canonical Cs are able to bind metal ions in as yet unknown ways, or that in archaea there is no need for binding zinc, or other metal ions, for the transcription factor to carry out its functions. These are areas for future research (see Section IV.C.2).

In the zinc fingers of human and yeast TFIIE *alpha* subunits, the first two and the last two Cs are separated by two aa, while the second and third Cs are separated by 17 aa (Maxon and Tjian, 1994; Bell et al., 2001). The zinc fingers of the archaeal TFEs do not match this pattern: the second and third Cs are separated by either 10 or 11 aa rather than 17 (Plate 5). Early multiple sequence alignments of human and yeast TFIIE *alpha* subunits and archaeal TFEs indicated that the TFEs from *A. fulgidus* (Klenk et al., 1997), *S. solfataricus*, *M. thermoautotrophicum*, and *Pyrococcus* species were abnormal in that the second Cs were replaced by other aa. It

was hypothesized that the TFEs from *S. solfataricus* and *A. fulgidus* could bind zinc, because the aa replacing the second Cs, aspartic acid (D) or methionine (M), are potentially capable of interacting with metal ions (Bell et al., 2001). The TFEs from *M. thermoautotrophicum* and *Pyrococci*, on the other hand, have glycine (G) or proline (P) instead of the second C. Both, G and P, are non-polar and hydrophobic, and, consequently, it was suggested that the TFEs with these aa in the region where a zinc finger ought to be might not be able to interact with metal ions (Bell et al., 2001). It has been demonstrated that the TFE from *M. thermoautotrophicum*, despite the composition of its zinc finger, does enhance transcription of certain genes as well as other archaeal TFEs and eucaryal TFIIE *alpha* subunits with canonical zinc fingers do (Hanzelka et al., 2001). Furthermore, it was demonstrated that the substitution of the third C (C155) in the *M. thermoautotrophicum* TFE for an A abrogates the protein's ability to initiate transcription.

2. Conservation

As mentioned above, the critical Cs in the zinc finger are conserved in eucaryal TFIIE *alpha* subunits but not as much in the archaeal homologs, TFEs. New multiple aa sequence alignments, constructed using the GCG program PILEUP and encompassing currently available archaeal TFEs and yeast and human TFIIE *alpha* subunits, provided novel insights on their zinc fingers (Plate 5*). Conclusions about the nonconserved nature of the zinc finger in TFEs were drawn from a multiple sequence alignment (Bell et al., 2001). Reexamination of the alignment revealed that it was driven by eukaryotic sequences, with the second C always separated from the first by exactly two aa. Another multiple sequence alignment showed that *A. fulgidus* TFE, for example, does contain a second C, but it is separated from the first by three aa not two (Hanzelka et al., 2001). The same was observed in the TFEs from *M. barkeri* and *M. jannsachii* when we generated alignments with updated sequence information. These new alignments also showed that the TFEs from *P. horikoshii*, *P. abyssi*, and *M thermoautotrophicum* do not have Cs but histidines (H) at positions that could function as the second C in the zinc finger—and it is known that H can substitute for C in functional zinc fingers (Miller et al., 1995). In the molecules with H in place of the second C, the separation between the first C and H is, like in those with Cs at both positions, variable, ranging from 3 to 4 aa (Plate 5).

The TFEs from *S. solfataricus* and *S. tokodaii* do not have Cs or Hs where they could form a zinc finger, alignable with the archaeal homologs, but instead have aspartic acids (D) separated by two aa from the first C (Plate 5). Aspartic acid might interact with metal ions via its charged oxygen atom and thus substitute for C (Bell et al., 2001). In summary, the data indicate that the zinc finger structure is conserved, and so would the function of TFE, but its key aa and their positions with regard to one another are somewhat variable.

D. TFE and Promoter Structure-Function Relationships

TFE does not seem to be as highly conserved as TBP and TFB; for example, human and yeast TIIFE *alpha* subunits share only 26% identity. One may conclude that each organism has its own, unique TFE to meet its specific needs. TFE from *A. fulgidus* did not stimulate transcription from *M. thermoautotrophicum* promoters, implying that these two archaeal species possess

* Plates appear following page 226.

distinct TFEs that are not interchangeable (Hanzelka et al., 2001). This TFE uniqueness might be due, at least in part, to the composition and architecture of the zinc fingers. Zinc fingers mediate protein-protein and protein-DNA interactions (references in Berg, 1990). If the zinc finger in TFE does mediate this protein's interaction with DNA, then it is possible that the different aa that substitute for the second C, and also other variations within the finger, may influence the bases or pattern of bases that the TFE will recognize. It is also possible that the zinc finger participates in the interaction of TFE with TBP, as well as DNA as mentioned above. A substitution of the third C, for example, may abrogate the TFE's ability to bind TBP, while a substitution of the second C may alter the factor's interaction with DNA.

S. solfataricus TFE stimulates transcription from certain promoters that have specific TATA box sequence variations (Bell et al., 2001). It is likely that other archaeal TFEs perform a similar function, but because of the poor conservation of TFE and considering the noninterchangeability of this factor across species (Hanzelka et al., 2001), it seems likely that different archaeal TFEs recognize different discrepancies in the TATA box sequence. The fact that the promoters of different groups of archaea, that is, methanogenic and halophilic, are different (Brown et al., 1989) supports this hypothesis. If one considers these observations, it then makes sense that the TFEs from *M. thermoautotrophicum* and *A. fulgidus* are not interchangeable (Hanzelka et al., 2001) considering that *M. thermo-autotrophicum* is a methanogen, while *A. fulgidus* is not. The combination of TFE architecture-composition on the one side, with promoter structure on the other, would be unique for each species, and even for each gene in any given species, which might generate a great diversity of transcription-initiation mechanisms differing only subtly, yet enough to achieve differential gene expression (see also Section XII.E).

The above data and considerations suggest that TFE stimulates transcription using a variety of mechanisms. It has been demonstrated that TFE aids TBP in binding weak promoters, which diverge from the consensus sequence, and when TBP is at low concentrations. It was proposed that TFE interacts with dimerized TBP that would be present in the cytosol (Qureshi et al., 1997) and induce dimer disassembly with generation of free, monomeric TBP (Bell et al., 2000). In eukaryotes, TBP dimerization is considered one mechanism for regulation of gene expression (Jackson-Fisher et al., 1999; Pugh, 2000).

It has also been suggested that TFE, like eucaryal TFIIE, might play a role in promoter melting and/or clearance and/or that TFE, again like TFIIE, might interact with activator factors to enhance transcription (Hanzelka et al., 2001). However, relatively few transcriptional activators have been found in archaea (Krueger et al., 1998; Gregor and Pfeifer, 2001; Pfeifer et al., 2001; see Section VII.E), whereas transcriptional repressors seem to be more common (Cohen-Kupiec et al., 1997; 1999; see Section VII.A).

V. RNA POLYMERASE (RNAP)

A. A Subunits

Archaeal RNAP is similar to eucaryal RNAP II and comprises a variable number of subunits, depending on the species. The archaeal homolog of the largest subunit of yeast RNAP, RBP1 (Lee and Young, 2000), is composed of two smaller subunits, A′ and A″ (Langer et al., 1995; Best and Olsen,

2001). These archaeal subunits are also homologous to the *beta* prime subunit of bacterial RNAP (Puhler et al., 1989). *M. thermoautotrophicum*, unlike any of the other known archaea, has a gene named *rpoA1b* encoding a putative second A′ subunit (Smith et al., 1997). The gene contains three Orfs possibly as a result of frameshift mutations. The combined lengths and molecular masses of the peptides encoded in the three Orfs add up to approximately those of the known A′ subunit, suggesting that they might form an alternative A′ subunit.

B. B Subunits

The archaeal homolog of the second largest yeast RNAP subunit, RBP2 (Lee and Young, 2000), exists as a single subunit, B, or as a complex of two smaller subunits, B′ and B″, depending on the archaeal species (Langer et al., 1995; Best and Olsen, 2001) these subunits are also homologous to the *beta* subunit of bacterial RNAP (Puhler et al., 1989). Early information had suggested that only methanogens and extreme halophiles possess the two B subunits (Langer et al., 1995), but further examination of genome sequences revealed that *A. fulgidus* and *S. solfataricus* also have both B′ and B″ subunits, rather than a single B subunit (Klenk et al., 1997; She et al., 2001).

C. Subunits E and F

1. Eucaryal RBP4 and RBP7

Subunits RBP4 and RPB7 of yeast and human RNAPII are examples of the eucaryal counterparts of the archaeal subunits E and F, respectively (Langer et al., 1995; Darcy et al., 1999; Werner et al., 2000; Best and Olsen, 2001). These subunits work together as a complex and are believed to be involved in the stress response. The RBP4/7 complex is apparently interchangeable in that it is able to dissociate from one core enzyme complex and reassociate with another (references in Hampsey, 1998). In yeast, these two subunits exist in only trace amounts during normal growth, but they are found in more substantial quantities associated with the core enzyme during growth under suboptimal conditions. It is believed that the RBP4/7 complex is involved in the interaction of the active site of RNAPII with the DNA (references in Lee and Young, 2000).

2. Subunit E

As in the case of the archaeal homolog of RPB2 (see above), the archaeal counterpart of RBP7 exists either as a single subunit, E, or as two, E′ and E″ (Langer et al., 1995; Best and Olsen, 2001). Subunit E′ is homologous to RBP2, while nothing homologous to subunit E″ has yet been identified in eucarya (Best and Olsen, 2001). Subunits E′ and E″ did not interact with one another, which is surprising considering that the single E subunit present in *S. acidocaldarius* contains domains that are similar to both subunits (Langer et al., 1995; Best and Olsen, 2001). Furthermore, immunoprecipitation experiments did not detect E″ among the separated RNAP subunits from *M. thermoautotrophicum* (Darcy et al., 1999), and mass spectroscopy analysis of *M. jannaschii* RNAP produced similar results (Best and Olsen, 2001), despite the fact that the genomes of both organisms were predicted to contain genes encoding this polypeptide (Bult et al., 1996; Smith et al., 1997). Hence, it was concluded that subunit

E″ might not be a structurally critical component of the archaeal RNAP holoenzyme (Best and Olsen, 2001). Therefore, the function of this subunit is still a mystery. It has to be borne in mind that the reason why subunit E″ is considered to be an archaeal RNAP subunit is that it shares similarity with the single subunit E from *S. acidocaldarius.*

Subunit E′ from *M. jannaschii* contains two potential RNA-binding domains, suggesting that it interacts with nascent RNA transcript and subunit F (see below), which stabilizes the complex (Todone et al., 2001).

3. Subunit F

It was believed that subunit F, that is, the archaeal counterpart of the eukaryotic RBP4, is not present in all archaea (Smith et al., 1997), but it was later identified as a previously unknown component of *M. thermoautotrophicum* RNAP (Darcy et al., 1999). Based on its predicted functional role and its limited sequence similarity to RBP4, this protein was identified as the equivalent of subunit F from *S. acidocaldarius* as described by Langer et al. (1995). Other candidate archaeal homologs have been identified subsequently in several other species (Darcy et al., 1999). Recent work showed that subunits E′ (or E) and F interact, like RBP4 and RBP7, with one another and form a complex (Werner et al., 2000). Also, it was shown that subunit F from *M. jannaschii* can form a complex with human RPB7 (the eucaryal counterpart of the archaeal subunit E), which testifies to the high degree of conservation of these RNAP subunits.

D. Subunits D, L, and N

Eucaryal RNAPII subunits RBP3, RBP11, and RBP10 build a complex, which in turn associates with subunit RPB2, the second largest subunit of the eukaryotic enzyme (Cramer et al., 2000). Archaeal homologs of RPB3, RBP11, and RBP10 have been identified and are referred to as subunits D, L, and N, respectively (Langer et al., 1995; Best and Olsen, 2001). Subunits D and L have been found to be also homologous to the *alpha* subunit of bacterial RNAP (Best and Olsen, 2001; Langer et al., 1995). These archaeal subunits form a subcomplex similar to that built by their eucaryal counterparts, which in turn interacts with the larger subunits A′, A″, B′, and B″ (or B, depending on the archaeon) (Cramer et al., 2000; Best and Olsen, 2001).

E. Subunit P

The eucaryal RBP12 is a structurally critical subunit of RNAPII as it associates with RBP3 and anchors the RPB3/11/10 subcomplex to RPB2 (Cramer et al., 2000). The archaeal counterpart of RBP12, subunit P, like subunit F, was originally believed to be absent from most archaea, but it was later shown to exist when it was purified with *M. thermoautotrophicum* RNAP and homologs were identified in archaeal genomes (Darcy et al., 1999; Werner et al., 2000). After its discovery, it was assumed that subunit P functions in connecting the D/L/N subcomplex with the larger subunits. Pairwise affinity-pulldown assays with *M. jannaschii* RNAP, however, could not detect any interactions between subunits D and P (Best and Olsen, 2001). Hence, it was hypothesized that subunit P interacts more with subunit B′ than with subunit D, and that if any interactions between subunits D and P were to occur, subunit B′ needed to be present (Werner et al., 2000; Best and Olsen, 2001).

F. Subunit K

Experiments involving yeast two-hybrid screenings and *in vitro* biochemical analyses have demonstrated that subunit K (RpoK, the archaeal counterpart of RBP6 from *Saccharomyces cerevisiae*; Best and Olsen, 2001) from *S. solfataricus* RNAP interacts with the NTD of TFB and is ultimately responsible for the association of RNAP with basal transcription factors and DNA (Magill et al., 2001). Homologs of subunit K are also present in bacterial RNAP, and in all three eucaryal RNAPs despite the fact that RNAPI and RNAPIII do not associate with TFIIB. Furthermore, TFIIB often does not associate directly but indirectly with RNAP through TFIIF, which appears to have no homolog in archaea (references in Magill et al., 2001). Therefore, it has been hypothesized that the three eucaryal RNAPs share as a common ancestor an archaeal-like RNAP, and that over time they have developed alternative or additional subunits for the recognition of different promoters (Magill et al., 2001).

G. Transcription Factor TFIIS or Subunit M

Open reading frames in the genomes of *T. celer* and *S. acidcaldarius* encode products bearing similarity to the eucaryal transcription factor TFIIS (Langer and Zillig, 1993; Kaine et al., 1994). This similarity is limited to the C-terminal portions of these peptides, which also share similarity with subunits of eucaryal RNAPI and RNAPII. Interestingly, these eucaryal subunits are more similar to their archaeal counterparts than they are to each other (Kaine et al., 1994), suggesting that the archaeal proteins might be precursors of the eucaryal RNAP subunits and TFIIS. The SwissProt accession numbers of these archaeal molecules, q07271 and q56254, belonging to *S. acidocaldarius* and *T. celer*, respectively, currently identify these proteins as subunit M of archaeal RNAP.

H. The C-Terminal Domain (CTD) of Subunit A″

The largest subunit of RNAPII, RBP1, contains a C-terminal domain (CTD) that, when phosphorylated, alters the conformation and function of the enzyme. The nonphosphorylated form of this enzyme, referred to as IIA, is present during PIC formation, while the phosphorylated form, IIO, is present during the elongation stage of transcription (references in Hampsey, 1998). Subunit A″ of archaeal RNAP does not have a CTD and, unlike eucaryal RNAP II, archaeal RNAP is not phosphorylated (Skorko, 1984; Soppa, 1999).

VI. DNA SITES; *CIS*-ACTING SIGNALS

A. Promoter and Initiator Element

The archetypal archaeal promoter resembles eucaryal RNA polymerase II promoters, consisting of a TATA box (Box A) and a few adjacent bases located 25 to 30 bp upstream of the transcription initiation site (Brown et al., 1989; Reiter et al., 1990; Soppa, 1999). An initiator element (Box B) has also been identified, located approximately 2 bp upstream of the transcription initiation site (Reiter et al., 1990; Soppa, 1999).

B. TATA box or Box A

In eukaryotes, TBP typically interacts with the TATA box for transcription initiation (Kim et al., 1993a; Guzikevich-Guerstein and Shakked, 1996; Soppa, 1999; Lee and Young, 2000; Woychik and Hampsey, 2002). A thorough examination of various species revealed that different subgroups of archaea have distinctive promoters, and consensus sequences were proposed as follows: AAANNTTTATATA for methanogens, AAANNTTTAAA for sulfur-dependent thermophiles, and GANGCCYTTAAGTA for halophiles (where Y = purine, and N = any nucleotide) (Brown et al., 1989). While the consensus sequences of promoters from methanogens and sulfur-dependent thermophiles are somewhat similar, that of halophiles is quite different from both. It must be mentioned, however, that heat-shock promoters have been described for *hsp60* genes (that encode chaperonins or Hsp60 heat-shock proteins), named *cct*, in the extreme halophile *H. volcanii*, and that these promoters resemble the consensus for methanogens rather than the consensus for extreme halophiles (Thompson and Daniels, 1998). The consensus promoter sequences mentioned above (Brown et al., 1989) were generated many years ago, before the genome era. Therefore, it is probable that if a systematic attempt were made today at generating consensus sequences using all genome sequences available, the resulting promoters would differ somewhat from the earlier ones. Until such new consensus sequences are generated, the earlier ones will have to be used as standards. In fact, they have proven useful in our searches to identify putative promoters, which were confirmed whenever experimental data became available (references in Macario et al., 1999). In the studies described in subsequent sections, we have used the old standards to identify putative promoters, and thus provide the starting point for future experiments aiming at mapping transcription-initiations sites and promoters.

C. Box B and the Initiator Element

Box B lies adjacent downstream of the proximal promoter or initiator element that was identified earlier (Reiter et al., 1990) and has been considered to be a part of this element (Soppa, 1999). The sequence of the initiator element is not conserved among archaeal species or among genes within a single species (Reiter et al., 1990; Soppa, 1999; AJLM and EC de M, unpublished results). In agreement with these observations, mutational experiments have demonstrated that extensive mutagenesis is required to abolish the function of the initiator element, indicating that the functional sequence of this element is flexible (Reiter et al., 1990). Despite this lack of conservation, there are often some common characteristics shared by different but related initiator elements. For example, although not conserved in sequence, the initiator elements present in RNA-gene promoters from *Sulfulobus* are rich in A and T (Reiter et al., 1990).

D. Absence of a Downstream Promoter Element (DPE)

Some eucaryal promoters contain a DPE that is thought to interact with TAFs and enhance TFIID-DNA binding in TATA-less promoters (Burke and Kadonaga, 1997; Hampsey et al., 1998; Albright and Tjian, 2000). As of now, no downstream promoter element has been identified among archaea (Soppa, 1999; AJH, EC de M, and AJLM,

unpublished results), which is not surprising because archaea do not have TAFs (or at least no TAF homologs have been reported thus far).

E. TFB-Recognition Element (BRE)

Archaeal TFB binds to DNA via interaction of with the BRE usually located 2 bp upstream of the TATA box (Qureshi and Jackson, 1998; Littlefield et al., 1999; Soppa, 1999; Bell et al., 1999b; see Section III.B). The BRE consensus sequence is RNWAAW (where R = purine; W = either A or T, and N = any nucleotide) as determined by studies with the *S. shibatae* T6 promoter (Bell et al., 1999b) and overlaps the consensus 13-bp promoter for methanogens identified earlier by Brown et al., 1989.

The BRE consensus sequence identified in *Homo sapiens*, SSRCGCC (where S = C or G; and R = purine; Lagrange et al., 1998), is considerably different from the archaeal consensus and is not well conserved among eucarya (Soppa, 1999). Likewise, the archaeal BRE is not well conserved either. Like the TATA box (Box A) and Box B, the BRE differs among different subgroups of archaea, although the latter always has two adenines (Soppa, 1999). The HTH motif contained in the C-terminal portion of TFB (see Sections III.A and III.B) interacts with both specific bases in the BRE as well as with the phosphate backbone of the DNA (Qureshi and Jackson, 1998; Littlefield et al., 1999). Differences among BREs in different genes and/or among the HTH motifs present in the multiple TFBs found in some archaeal species would result in different TFB/BRE interactions, which might be a mechanism of differential gene expression and regulation.

As discussed above, Section III.B, the binding of TFB to BRE is critical for determining the orientation of transcription (Bell et al., 1999b). In some eukaryotic systems the TFIIB-TBP promoter complex is stabilized through direct interaction between BRE and TFIIB, which enhances transcription (Lagrange et al., 1998). However, under different experimental conditions the BRE appeared to play a role in suppressing basal transcription instead of activating it (Evans et al., 2001). These different observations, together with those made with archaeal promoters, show that the BRE sequence and function are poorly conserved. The data suggest that the role of the BRE and the BRE-binding factors in transcription initiation has followed different evolutionary pathways; it may vary among phylogenetic branches and, even in a single species, among different promoters.

VII. REGULATION: *TRANS*-ACTING FACTORS

A. Repressors

1. MDR-1

Although the basal archaeal transcription apparatus is of eucaryal type, its activity seems to be controlled more often than not by bacterial-like factors (Cohen-Kupiec et al., 1997, 1999; Bell et al., 1999a). MDR-1 is a transcriptional repressor found in *A. fulgidus* (Bell et al., 1999a), and is homologous to the bacterial metal-dependent transcription repressor DtxR (Boyd et al., 1990). It represses the transcription of the operon containing its own gene, in addition to genes belonging to an ABC transporter system (Bell et al., 1999a). The genes of the operon are transcribed as a single, polycistronic, mRNA that, when processed, is divided into MRD-1 and ABC transporter transcripts. The MRD-1 protein binds cooperatively to an operator sequence within its own promoter in the

presence of specific metal ions. The binding of this repressor does not interfere with the association of TBP or TFB with DNA but rather imitating a mechanism that is characteristic of some bacterial transcriptional repressors, it interferes with the recruitment of RNA polymerase.

2. Lrp

Archaea also possess homologs of the leucine-responsive regulation protein (Lrp) (Napoli et al., 1999; Ouhammouch and Geiduscheck, 2001), which performs a number of both positive and negative gene-regulatory functions in bacteria (references in Newman and Lin, 1995). In addition, Lrp influences DNA topology and the formation DNA-protein complexes in bacteria (Wang and Calvo, 1993). Lrs14, an archaeal homolog of Lrp from *S. solfataricus* (Napoli et al., 1999), exists as a homodimer in solution (Bell and Jackson, 2000b). This transcriptional regulator binds to the promoter of its own gene and represses its own transcription (Napoli et al., 1999). Unlike the MDR-1 repressor discussed in the preceding section, Lrs14 binds DNA and covers the TATA box and the BRE, obstructing the binding of TBP and TFB to the promoter (Bell and Jackson, 2000b). Recent work has shown that the LrpA from *P. woesei* inhibits transcription by interfering with RNAP recruitment (Dahlke and Thomm, 2002).

Prt1 and Ptr2 are two Lrp homologs found in *M. jannaschii* that, like Lrs14, bind specific DNA sequences within their own promoters. The sequences recognized by these proteins are palindromic, indicating that both molecules interact with DNA as dimers (Ouhammouch and Gieduscheck, 2001).

3. Repressors of Nitrogen-Fixation Genes

nifH and *glnA* are two genes involved in nitrogen fixation and ammonia assimilation, respectively, in the archaeon *Methanococcus maripaludis* (Cohen-Kupiec et al., 1997, 1999). When an abundant source of nitrogen such as ammonia is present, transcription from these genes is repressed. A *lacZ* reporter construct was prepared in order that transcription from the promoter region of *nifH* could be assessed (Cohen-Kupiec et al., 1997). This promoter region contains two palindromic sequences located immediately downstream of the transcription initiation site. Mutations within the first palindrome resulted in derepression of *nifH* in the presence of ammonia, while mutations in the second palindrome had no discernable effect. EMSA showed that a potential repressor bound the first palindrome but not the second. It was inferred that the binding of the repressor to the first palindrome, which is located just downstream of the transcription initiation site, interferes with the assembly of the basal transcriptional machinery, very much like some bacterial repressors do (Cohen-Kupiec et al., 1997).

Unlike *nifH*, *glnA* is transcribed from three different promoters (Cohen-Kupiec et al., 1999). Two of these promoters overlap one another and closely resemble the promoter consensus sequence. The remaining promoter shares less similarity with the consensus and lies further upstream from the translation start codon. The promoter region of *glnA* includes a palindromic sequence that greatly resembles the one found in the promoter region of *nifH*. Mutational analyses demonstrated that this sequence, like the one described for *nifH*, is responsible for the repression of *glnA*. *glnA*, however, is constitutively expressed at low levels even in the presence of ammonia

(Cohen-Kupiec et al., 1999). Furthermore, it has been demonstrated that this gene is essential for the survival of the organism (Cohen-Kupiec et al., 1999).

Under nitrogen-limiting conditions, the two overlapping promoters of *glnA* produced high levels of transcript, while the amount of transcript produced by the upstream promoter was significantly less. In the presence of ammonia, the level of transcript produced from the downstream overlapping promoters was reduced, while that from the upstream promoter was virtually unaffected (Cohen-Kupiec et al., 1999). Therefore, the presence of nitrogen had little if any effect on the upstream promoter, which appears to be the mechanism behind the constitutive low-level expression of this gene. Mutation of the palindromic sequence, which is adjacent to the transcription initiation sites from the two overlapping promoters, reduced the repression of this gene. Thus, it appears that the binding to the palindromic sequence of a repressor, which is likely to be the same molecule that represses *nifH*, is sufficient to block transcription initiation from the downstream promoters, but is not sufficient to stop transcription once it has begun from the upstream promoter (Cohen Kupiec et al., 1999).

B. The Helix-Turn-Helix(HTH) Motif

MDR-1, Lrs14, Prt1, and Prt2 have the HTH motif (Bell et al., 1999a; Napoli et al., 1999; Ouhammouch and Geiduschek, 2001). This motif is also present in the basal transcription factors TFB and TFE (Aravind and Koonin, 1999; Thomsen et al., 2001; see Section III.B), and several prokaryotic regulators (Huffman and Brennan, 2002). Examination of available archaeal genomes and their encoded proteins revealed that although the archaeal HTH domains present in these two transcription factors are more similar to bacterial than to eukaryotic equivalents, many archaeal HTH motifs show archaea-specific characteristics (Aravind and Koonin, 1999). The HTH domains of archaeal basal transcription factors also differ from those found in their eucaryal counterparts (Thomsen et al., 2001). TFB (Plate 3*) and TFE HTH motifs could easily be identified in alignments with HTH canonical consensus sequences, while those of the eucaryal TFIIE *alpha* subunits and TFIIB could not.

C. MBF1

An archaeal homolog of the eucaryal protein MBF1 has been reported (Aravind and Koonin, 1999). The function of this protein is not fully understood, but in some specific instances it can serve as a transcriptional co-activator connecting transcription activators with TBP (Takemaru et al., 1997, 1998). It has been suggested that MBF1 is a required basal transcription factor because of its degree of conservation in eukaryotes and the presence of its homolog in archaea (Aravind and Koonin, 1999).

D. Tfx

Potential transcriptional regulators of eucaryal type have been reported to occur in archaea. One possible example is Tfx, found in *M. thermoautotrophicum* (Hochheimer et al., 1999). The gene encoding this protein, *tfx*, is located upstream of the molybdenum formylmethanofuran dehydrogenase or *fmdECB* operon. Although Tfx interacted

* Plates appear following page 226.

specifically with the promoter of this operon, it did not appear that the gene *tfx* is part of the operon itself because its expression was independent of the expression of the other genes in the operon and produced a monocistronic message. While transcription of the *fmdECB* operon is positively regulated by the presence of molybdate (Hochheimer et al., 1998), transcription of *tfx* was unaffected by this element, although it was repressed in the presence of tungstate. The *fmdECB* operon was transcribed in the presence of both molybdate and tungstate; notably, in the absence of molybdate, tungstate was required for the growth of *M. thermo-autotrophicum*.

The N-terminal portion of Tfx contains an HTH motif believed to be involved in DNA binding, while the C-terminal portion of the protein contains an acidic domain possibly involved in transcriptional activation via interaction with the basal transcription factors, TBP and/or TFB (Hochheimer et al., 1999). This suggests that Tfx might be a eucaryal-like transcriptional regulatory factor; if that were the case, it would be one of the few of this type known in archaea (see the following section).

E. Gas Vesicle Protein (Gvp) Gene Regulation

Another example of an archaeal transcriptional activator is GvpE (Kruger el al., 1998). This factor activates transcription of *gvpA*, whose product is important for the formation of gas vesicles in certain species of halophilic archaea. *gvpA* and *gvpE* are two of 14 genes involved in gas vesicle formation. *H. salinarum* strain PHH1 has two sets of these 14 genes. One set resides in the chromosome, while the other is in the plasmid pHH1. The plasmid allele of *gvpA* (named p-*gvpA*) was expressed constitutively, while transcription from the chromosomal allele (c-*gvpA*) required activation by chromosomally expressed GvpE (c-GvpE) (Krueger et al., 1998).

Molecular modeling and homology analyses revealed that GvpE resembles eukaryotic leucine-zipper proteins that are transcriptional activators. It was predicted from the models that GvpE binds DNA as a dimer with three basic residues in its DNA-binding site interacting with the negatively charged phosphate backbone of DNA. The latter prediction was supported by the fact that mutation of these three residues to either basic or neutral aa resulted in the loss of activity of GvpE (Krueger et al., 1998).

Examination of the promoter regions of the chromosomal and plasmid *gvpA* alleles revealed that the promoter of p-*gvpA* contains a highly conserved TATA box and BRE (Gregor and Pfeifer, 2001). It was suggested that additional factors were not necessary for the transcriptional activation of p-*gvpA* because this gene, which has conserved promoter elements, was readily activated by the basal transcriptional machinery. In contrast, the TATA box present in the promoter of c-*gvpA* is less conserved and the BRE is absent. Due to this lack of conservation by comparison with canonical promoters, it was suggested that the c-*gvpA* promoter is not readily transcribed by the basal transcriptional machinery, and therefore requires activation by chromosomally expressed GvpE—it was not established where in the *gvpA* promoters c-GvpE might bind.

Interestingly, pHH1-expressed GvpE (p-GvpE) was unable to activate transcription from the promoter of c-*gvpA*, while both p-GvpE and c-GvpE could activate and enhance transcription from the p-*gvpA* promoter; c-GvpE appeared to be a stronger activator of transcription than p-GvpE (Gregor and Pfeifer, 2001). A *beta*-galactosidase reporter construct containing a chimeric promoter in which the BRE of

p-*gvpE* was fused to the TATA-box of c-*gvpE* failed to produce *beta*-galactosidase activity in the absence of GvpE, leading to the conclusion that the presence of an intact BRE by itself was not enough to initiate transcription of this gene. The presence of both c-GvpE and p-GvpE, however, induced *beta*-galactosidase expression from the chimeric promoter in contrast to the c-*gvpA* promoter, which produced transcript only in the presence of c-GvpE.

Haloferax mediterranei also has the 14 gas vesicle-protein genes, including *gvpA* (mc-*gvp*A) and *gvpE* (mc-*GvpE*) (Gregor an Pfeifer, 2001; Krueger et al., 1998). mc-GvpE regulates mc-*gvpA* in a similar manner as c-GvpE regulates c-*gvpA*. Interestingly, mc-GvpE can also activate transcription from the p-*gvpA* promoter, but not from the *c-gvpA* promoter (Gregor and Pfeifer, 2001).

Transcription of *mc-gvpA* was repressed by the regulatory protein GvpD; the latter seemed to counteract the activity of GvpE (Pfeifer et al., 2001). The mechanism by which GvpD accomplished this is not fully understood, although several suggestions were made. One hypothesis suggested that repression of *gvpA* is accomplished through the binding of GvpD to its promoter. The alternative idea proposed that GvpD somehow inactivates GvpE or reduces the amount of GvpE produced through interaction with its mRNA.

VIII. TBP INTERACTING PROTEIN (TIP)

A protein that binds TBP, TIP, has been reported to occur in the hyperthermophilic archaeon *Thermococcus* (formerly *Pyrococcus*) *kodakaraensis* (Matsuda et al., 1999). This protein, which was named *Pk*-TIP or TIP26, bound TBP in the absence of TFB and thereby prevented it from binding DNA. When TBP was present, TIP26 formed a quaternary complex, TIP26-TBP-TFB-DNA, as a second step following the initial formation of the ternary complex TBP-TFB-DNA (Matsuda et al., 2001). In the presence of TFB, TIP26 no longer inhibited TBP-DNA binding, a phenomenon that was attributed to an equilibrium shift between TBP-TIP26 and TBP-DNA in favor of the latter complex. High concentrations of TIP26 were required to form the quaternary complex exclusively; at lower concentrations of the protein the quaternary complex was formed but much less efficiently, and when the concentration fell below a certain level no quaternary complex was formed. The nature of the interactions between TIP and its partners and their role in PIC formation is still unclear. The cellular level of TIP26 was found high enough relative to that of TBP to promote the formation of quaternary complexes *in vivo*, which indicates that these complexes and TIB26 might play a physiological role during normal cell growth.

IX. NUCLEAR PROTEIN HOMOLOGS WITH A ROLE IN TRANSCRIPTION

Eukaryotic nucleolar protein homologs are present in archaea (Gautier et al., 1997; Hickey et al., 2000). Some of these proteins in eukaryotes are involved in transcriptional regulation (Newman et al., 2000). Studies of the complexes formed by U14 small nucleolar RNA (snoRNA) in mice revealed that two separate pairs of proteins associate with the box C/D motifs of this molecule. The proteins of one of these pairs are conserved homologs of yeast nucleolar proteins Nop56 and Nop58. These molecules are highly homologous to each other, yet the presence of both proteins is required for viability in yeast (Gautier et al., 1997). Interestingly, archaea have a single protein that is homologous to both Nop58 and Nop56, and that also shares

similarity with another eukaryotic protein, Prp31 (Hickey et al., 2000).

The other pair of proteins that associate with the box C/D motifs of the U14 snoRNA are nucleolar proteins p50 and p55. As is the case for Nop56 and Nop58, p50 and p55 are similar to each other. A single protein that is homologous to both p50 and p55 has been identified in some but not all archaea (Newman et al., 2000; King et al., 2001). The possibility of Nop58 and Nop56, and p50 and p55, forming two separate heterodimers has been suggested. If this were the case, it could also be suggested that their respective archaeal homologs form homodimers, because archaea possess only one copy from each homologous pair (Newman et al., 2000).

It has been demonstrated that the p55 homolog in rats interacts with TBP (Kanemaki et al., 1997). Hence, it was suggested that this protein, designated TIP49 (all these proteins are referred to by many different names throughout the literature), plays a role in the regulation of transcription (Kanemaki et al., 1997). The other nucleolar protein, p50, might be involved in different cellular activities in eukaryotes, including transcriptional regulation of genes pertinent to the cell cycle and ribosome-protein synsthesis (Lim et al., 2000), the production and localization of snoRNA proteins (King et al., 2001), and the Wnt signaling pathway via its interaction with *beta*-catenin in conjunction with its p55 counterpart (Bauer et al., 1998, 2000). However, in snoRNA biogenesis, p50 (or Rvb2 as p50 is referred to in this instance) seems to perform its function also after rather than only during transcription of the snoRNA genes (King et al., 2001).

p50 and p55 have helicase activity; hence, the suggestion that these two nucleolar proteins play a role in DNA replication (Kanemaki et al., 1999; Makino et al., 1999; Newman et al., 2000). Helicase activity is required not only for DNA replication but also for gene transcription. If one or both of these proteins interacted with TBP, then it would seem reasonable to suggest that their helicase activities could aid in melting the double-stranded DNA to allow transcription initiation. In addition, recent work has led to the idea that p50 and p55 regulate transcription in eucarya via chromatin remodeling (Ikura et al., 2000; Shen et al., 2000), a mechanism that could also operate in archaea because some species do have primitive chromatin homologs (references in Sandman and Reeve, 2000; see Section X.B). p50 (Rvp2) would also be involved in unwinding RNA/DNA duplexes or double-stranded RNA and thereby modifying snoRNAs (King et al., 2000). Whether the helicase activity of the archaeal p50 is involved also in transcription remains to be elucidated.

X. DNA STRESS ELEMENTS

A. Stress-Related *cis*-Acting Signals

In addition to the TATA box (Box A) and Box B, other elements might exist in archaea for the expression of stress genes. Comparative and deletional analyses revealed specific sequences just up- and downstream of the TATA box that were important for both basal and heat shock-induced expression of three *cct* heat-inducible genes in *H. volcanii* (Kuo et al., 1997; Thompson and Daniels, 1998). The conserved sequence CGAA located 5 bp upstream of the TATA box could not be deleted or replaced without greatly reducing the basal transcription of the *cct* genes and abrogating heat shock-induced transcription. Interestingly, replacing the first A in this element with a C had no effect on basal transcription, yet heat-shock transcription was reduced. A second

conserved sequence, CAAaC, located six bp downstream of the TATA box also appeared to be important in the regulation of these genes. The replacement of the CA bases with TG had little effect on basal transcription genes but decreased heat-shock transcription. Comparisons of the promoter regions of several genes from various archaeal species revealed that aside from the TATA box, only the sequence GAA contained within the upstream element CGAA was conserved. Therefore, it was suggested that the necessary DNA regulatory elements required for heat-shock expression are located in the core promoter region (Thompson and Daniels, 1998).

Recently, a compilation of heat-shock promoters from several archaeal genomes was generated (Gelfand et al., 2000). Although there were some discrepancies among promoters from different organisms and different families of heat-shock genes, the overall structure of the archaeal heat-shock promoter appeared to be fairly well conserved. Most of the identified promoters contained a recognizable TATA box with the CGAA element upstream, although this element was not always located exactly 5 bp upstream of the TATA box. These putative heat-shock promoters were located at a variety of positions ranging between 10 and 273 bp upstream of the transcription start codon.

We have done an extensive search for heat-shock elements (HSE) typically found in the promoter region of eukaryotic heat-shock genes (Fernandes et al., 1994) and found none (Macario et al., 1999; and unpublished results). The lack of HSE coincides with the absence of archaeal homologs in all archaeal genomes examined (AJLM and EC de M, unpublished results) of heat-shock factors (HSF) typical of eukaryotes, where they play a key role in the regulation of stress-genes along with HSE (Morimoto, 1998).

B. DNA Topology and Stress-Gene Expression

Archaeal gene transcription can be influenced by DNA topology (Yang et al., 1996; Bell et al., 1998; Baliga and DasSarma, 2000). The promoter of the *bop* gene in *Halobacterium salinarum*, for example, contains an 11-bp sequence, known as the RY box, that is sensitive to DNA supercoiling and appears to influence transcription based on its topology (Yang et al., 1996; Baliga and DasSarma, 2000). None of the other elements present in this promoter played a detectable a role in this regulatory mechanism, implying that only the RY box is responsible for topology-based regulation. It was suggested that the RY box acts as a *cis*-acting regulatory element whose structural change, resulting from DNA supercoiling, facilitates the binding of basal trancription factors to DNA and/or the formation of the open-promoter complex (Baliga and DasSarma, 2000).

Similar topological changes might act as a regulatory mechanism during the stress response (Lopez-Garcia and Forterre, 2000). This hypothesis gives weight to the fact that stress could affect DNA topology by melting the double-stranded molecule, and this in turn would influence transcription. Studies on the effects of temperature and DNA topology on transcription in *S. shibatae* demonstrated a transcriptional preference for negatively supercoiled DNA when growth conditions were suboptimal (Bell et al., 1998). At 75°C, the optimal temperature for growth of *S. shibatae*, the topology of the DNA did not affect transcription. When the temperature was lowered to 48°C, transcription of linearized DNA was reduced or blocked, while that of negatively supercoiled DNA was uninhibited. Furthermore, an open promoter complex did not form with linear or positively supercoiled templates at 48°C.

It was suggested that the observed effects were in response to cold rather than to heat stress.

Similar experiments on DNA topology were conducted using the transcriptional machinery of *P. furiosus*, which has an OTG of 95 to 100°C (Hethke et al., 1999). At 70°C, positively and negatively supercoiled and relaxed DNA templates were transcribed *in vitro*, but not with the same efficiency. Negatively supercoiled DNA was the most efficiently transcribed, followed by relaxed DNA, while transcription was weak for positively supercoiled DNA. At 90°C, the transcription of negatively supercoiled and relaxed DNA was significantly reduced, while that of positively supercoiled DNA ceased. Based on these results, one may conclude that under conditions of stress, regardless of whether it is cold or heat stress, negatively supercoiled DNA is preferentially transcribed.

Stress proteins are also known to influence DNA topology, often by counteracting the effects of stress on DNA (references in Lopez-Garcia and Forterre, 2000). Among the many functions performed by Hsp70(DnaK), an important molecular chaperone/heat-shock protein that is discussed in greater detail in Section XII, it was found that it aids in the restoration of DNA negative supercoiling after heat shock in *E. coli* (Ogata et al.,1996). Therefore, Hsp(DnaK) may play a role in the regulation of certain genes, including its own, through its ability to restore DNA negative supercoiling (in addition to the other ways in which this protein regulates itself; Blaszczak et al., 1999; Georgopoulos et al., 1994; Missiakas et al., 1996; Morimoto, 1998).

The promoters of some archaeal histone genes are similar to those of heat-shock genes, implying that the former genes are stress inducible and that histones might play a role in the stress response (Gelfand et al., 2000). Histones, which are known to control DNA topology, may also participate in the regulation of other archaeal stress genes by controlling DNA supercoiling. Histones and chromatin remodeling proteins are involved in transcriptional regulation in eukaryotic organisms (Kadonaga, 1998; Spencer and Davie, 1999; Orphanides and Reinberg, 2002). The archaeal nucleosome is less complex than its eukaryotic counterpart, and it is not clear whether its molecular components are able to influence transcription as the eukaryotic nucleosome molecules do (references in Sandman and Reeve, 2000). Pertinent to this point, a recent report indicates that archaeal chromatin-remodeling proteins are able to repress transcription (Bell et al., 2002).

XI. STRESS-GENE EXPRESSION IN *M. MAZEII*

As explained in the Section I, there are several reasons that make *M. mazeii* an attractive target for investigating stress-gene transcription. The first molecular-chaperone machine gene cloned and sequenced from an archaeon was *hsp70(dnaK)* from *M. mazeii* (Macario et al., 1991). This was followed by the cloning and sequencing of the other two components of the machine (*hsp40(dnaJ)* and *grpE*; Conway de Macario et al., 1994) and the flanking genes, and the characterization of their expression patterns under basal (constitutive) conditions and in response to stress (references in Macario et al., 1999). These studies were continued, and today there is considerable information on the pattern of expression, characteristics of the transcripts, initiation and termination sites, promoters, and other useful parameters about the *M. mazeii* genes of the *hsp70(dnaK)* locus. This type of information is essential to decide on strategies and methods for elucidating the molecular mechanisms of transcription initiation and

regulation. The availability of such information pertaining to stress genes is unique in the field of methanogenic archaea. Furthermore, these genes encode proteins very similar to bacterial homologs, but the promoters and mode of transcription are of archaeal type, which resembles that of eukaryotes. It seems that the *M. mazeii* genes were received from a bacterium by horizontal transfer (Gribaldo et al., 1999; Macario and Conway de Macario, 1999). Thus, *M. mazeii* provides a model suitable for investigating how a "bacterial" gene is transcribed and regulated in an organism whose basal transcription factors and DNA elements are different from those from bacteria, and more similar to eucaryal counterparts.

XII. THE *HSP70(DNAK)* LOCUS

A. Presence and Absence of the Locus in Archaea

The organization of the *M. mazeii* S-6 *hsp70(dnaK)* locus is: *5′-orf16-grpE-hsp70(dnaK)-hsp40(dnaJ)-orf11-trkA-3′*, where the first and the last two genes are non-heat shock inducible, but the other three are inducible. The three genes in the middle of the cluster encode the molecular chaperones/stress proteins GrpE, Hsp70(DnaK), and Hsp40(DnaJ), which constitute the molecular chaperone machine (references in Macario et al., 1999). *hsp70(dnaK)* and *hsp40(dnaJ)* are present in all eukaryotes and bacteria but in only some archaea, while *grpE* occurs in all bacteria, only in archaea with *hsp70(dnaK)* and *hsp40(dnaJ)*, and in the organelles of eucarya (Gupta and Singh, 1992; 1994; Bustard and Gupta, 1997; Gupta, 1998; Gribaldo et al., 1999; Macario et al., 1999). Other archaea that also have these stress genes include *M. thermophila*, *M. barkeri*, *M. acetivorans*, *M. thermoautotrophicum*, *T. acidophilum*, *F. acidarmanus*, *T. volcanium*, *Halobacterium marismortui*, and *Halobacterium* sp. NRC-1 (Gupta and Singh, 1992, 1994; Smith et al., 1997; Hofman-Bang et al., 1999; Kawashima et al., 2000; Ng et al., 2000; Ruepp et al., 2000; 2001; Galagan et al., 2002; see also:

- *M. barkeri*: http://www.jgi.doe.gov/JGI_microbial/html/
- *M. acetivorans:* http://www-genome.wi.mit.edu/
- *M. thermoautotrophicum*: http://www.biosci.ohio-state.edu/~genomes/mthermo/
- *T. acidophilum*: http://www.biochem.mpg.de/baumeister/genome/
- *T. volcanium*: http://www.ncbi.nlm.nih.gov
- *F. acidarmanus*: http://www.jgi.doe.gov/JGI_microbial/html/
- *Halobacterium* sp. NRC-1: http://www.ncbi.nlm.nih.gov/

B. Protein-Coding Regions

The *hsp70(dnaK)* loci are organized so that *grpE* lies upstream of the other two stress genes. *hsp70(dnaK)* follows *grpE*, and *hsp40(dnaJ)* lies downstream of *hsp70(dnaK)*. This gene organization is conserved in all known archaeal loci except for that of *Halobacterium* sp. NRC-1, in which an additional open reading frame, *vng0492h*, occurs between *grpE* and *hsp70(dnaK)*.

The *hsp70(dnaK)* locus of *M. mazeii* S-6 also contains the genes *orf16* and *orf11-trkA* located up- and downstream of the stress genes, respectively (references in Macario et al., 1999). *trkA* is the homolog of an *E. coli* gene involved in potassium transport; the function of *orf16* is unknown. Similar genes occur at the same position in the *M. barkeri* and *M. acetivorans* loci. As of now, it is not known whether the *M. thermophila* locus also has an o*rf16* ho-

molog because the nucleotide sequence of the region upstream of *grpE* of this organism is not available, but *M. barkeri*, *M. thermophila*, and *M. acetivorans* do contain homologs of *trkA* in the expected location, downstream of *hsp40(dnaJ)*.

In addition, *M. mazeii* contains an additional open reading frame, *orf11*, that is located just upstream of *trkA*; both genes are co-transcribed (Conway de Macario and Macario, 1995). *M. barkeri*, *M. thermophila*, and *M. acetivorans* contain small open reading frames that encode proteins of 87 (*orf87*), 40 (*orf40*), and 102 (*orf102*) aa, respectively, located upstream of *trkA*. The 40 aa protein shares some similarity with the N-terminal portion of the 87 aa protein, and both share a weak similarity with the protein encoded by *orf11*, as does the 102 aa protein from *M. acetivorans* (data not shown). In *M. thermophila*, *orf40* is located 32 bp upstream of *trkA*, while in *M. barkeri* *orf87* overlaps that gene. If the stop codon of *orf40* were ignored, the resulting peptide would be 67 aa, it would be more similar to *orf87* than the 40 aa protein, and it would overlap *trkA*. Like *orf87*, *orf102* from *M. acetivorans* overlaps the 5′ end of *trkA*.

No *trkA* homologs are present in the *hsp70(dnaK)* locus of the other archaea examined, and no conserved homologs of *orf16* have yet been identified in other archaeal genomes. Upstream of *grpE*, in *M. thermoautotrophicum*, lies *mth1288*, which encodes a putative transcriptional regulator, and 196 bp downstream of *hsp40(dnaJ)* lies *mth1292*, an RNA gene encoding RNase P RNA. The genes located up- and downstream of the stress genes in *T. acidophilum*, *ta1085* and *ta1089* (Ruepp et al., 2000; http://www.biochem.mpg.de/baumeister/genome/), and in *T. volcanium*, *tvn0440* and *tvn0486*, (Kawashima et al., 2000) encode a multidrug-efflux transporter protein and subunit E′ of RNA polymerase, respectively. Examination of the preliminary data available for *F. acidarmanus* revealed an Orf 56 bp upstream of *grpE* that according to BLAST searches of database SWPLUS encodes a protein related to thiamine monophosphate kinase. Two hundred and twenty-seven bp downstream of *hsp40(dnaJ)* lies an Orf encoding a 262 aa protein that is homologous to sulfur oxygenase reductase (SOR) from *Acidianus ambivalens* (Kletzin, 1992). The two genes up- and downstream of *grpE* in *Halobacterium* sp. NRC-1 are hypothetical, and no function has yet been assigned to them, while *vng0448h*, the gene downstream of *hsp40(dnaJ)*, encodes a protein related to acetyl CoA-synthetase (Ng et al., 2000).

C. Gene Expression

The stress genes in the *hsp70(dnaK)* locus of *M. mazeii* S-6 are expressed constitutively at a relatively low level, but after heat shock this level increases considerably (Clarens et al., 1995; Conway de Macario et al., 1995). The stress genes in this locus as well as those in *M. thermophila* (Hofman-Bang et al., 1999) are expressed separately, each with a monocistronic message; whether this is true also in other archaea possessing these genes has yet to be ascertained (see below, Section XII.D.2).

D. Intergenic Regions

1. Promoters

The promoters of *hsp70(dnaK)* locus genes in *M. mazeii* S-6 are shown in Table 4. *grpE* is the only stress gene in this locus that has a complete canonical TATA box as described by Brown et al., 1989. The promoter region of *hsp70(dnaK)* contains a sequence resembling

TABLE 4
Known and Putative Promoters of *hsp70(dnaK)*-Locus Genes in Archaea

Gene	Organism	Sequence (5'-3')[a]	Position[b]	Reference[c]
orf16	*M. mazeii* S-6	aagtAAAagTTTATATA	-51	Macario and Conway de Macario, 1997
	M. barkeri	atgtAAAagTTTATATA	-34	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	M. acetivorans	aagtAAAggTTTATATA	-45	Galagan et al., 2002
grpE	*M. mazeii* S-6	aGAAAActaTTTATAgA	-69	Macario and Conway de Macario, 1997
	M. barkeri	CGAAAActaTTTATAgA	-69	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	M. acetivorans	caaaAActaTTTATAgA	-68	Galagan et al., 2002
	M. thermoautotrophicum	gtCGAAAggTTTATATA	-87	Gelfand et al., 2000
	T. acidophilum	ccaactAtcTTTATATA	-28	http://www.biochem.mpg.de/baumeister/genome/
	T. volcanium	ctcactAtcTTTATATA	-28	Kawashima et al., 2000
	F. acidarmanus	taGAAAAtaTTTATATA	-23	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	Halobacterium sp. NRC-1	ccCGAAAtccgTAaAcc	-28	Ng et al., 2000
hsp70(dnaK)	*M. mazeii* S-6	GAAaAAAcTTTaATtaA	-79	Macario and Conway de Macario, 1997
	M. barkeri	tttaAgtatTTTATtTA	-71	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	M. acetivorans	caaaATAagTTTATtaA	-114	Galagan et al., 2002
	M. thermophila	ctaaGAAvtTTTATcTA	-51	Hofman-Bang et al., 1999
	M. thermoautotrophicum	aGAAAAAatTcTcggTA	+23	Smith et al., 1997
	T. acidophilum	ggagcAAgtTTgATccA	-141	http://www.biochem.mpg.de/baumeister/genome/
	T. volcanium	tgttAAAaggTgATgaA	-1	Kawashima et al., 2000
	F. acidarmanus	ttcaAAAaggTgAaATA	-8	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	Halobacterium sp. NRC-1	ctagcAAggcTTAaAcA	-32	Ng et al., 2000
hsp40(dnaJ)	*M. mazeii* S-6	ccgcAAAccTgTtcAcA	-48	Macario and Conway de Macario, 1997
	M. barkeri	actttAtctTTTtTAaA	-22	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	M. acetivorans	tcctttAttTTTtTAaA	-22	Galagan et al., 2002
	M. thermophila	tccggAgctTTatctTt	-28	Hofman-Bang et al., 1999
	M. thermoautotrophicum	acccAcAttTTTttaTt	-63	Smith et al., 1997
	T. acidophilum	GAAaAAAgcTgatTAaA	-17	http://www.biochem.mpg.de/baumeister/genome/
	T. volcanium	tgGAAAAgaTaTcTAaA	-158	Kawashima et al., 2000
	F. acidarmanus	GAAgAAAgaTgTAgATA	-224	http://www.jgi.doe.gov/tempweb/JGI_microbial/html/
	Halobacterium sp. NRC-1	tctcAActgggTAcATg	-16	Ng et al., 2000
orf11-trkA	*M. mazeii* S-6	TGAAAAAcaTTTATATA	-98	Macario and Conway de Macario, 1997
orf87-trkA	*M. barkeri*	ttGAAAAcaTTTATATA	-106	http://www/jgi/doe.gov/tempweb/JGO_microbial/html/
orf102-trkA	*M. acetivorans*	tGAAAAAcaTTTATATA	-96	Galagan et al., 2002
orf40-trkA	*M. thermophila*	CtaaAAAgaTTTATATA	-106	Hofman-Bang et al., 1999

TABLE 4 (continued)

[a]Boldface, the 13-base promoter with upper and lower case letters indicating identity with consensus (as defined by Brown et al., 1989; see Section VI.B) or lack thereof, respectively, and where the TATA-box (Box A), or vestiges of it, is underlined. The "CGAA" or "GAA" element is boxed.

[b]Distance in bp from the center of the promoter to the translation start codon: upstream and downstream of it, as negative and positive numbers, respectively.

[c]See websites (URL) given in the text (e.g., Section XII.A) for more information.

a TATA box, while that of *hsp40(dnaJ)* has no discernable TATA box. In contrast, the non-heat shock-inducible genes *orf16* and *orf11-trkA* have promoters that completely match the archaeal consensus for non-stress genes as described by Brown et al., 1989 (Macario and Conway de Macario, 1997).

The upstream regions of the genes in the *hsp70(dnaK)*-loci from other archaea were queried with the consensus promoter sequence (Brown et al., 1989; see Section VI.B) to find the best match and the results are displayed in Table 4. The promoters that were identified were compared with each other and with the known heat-shock gene promoters to confirm that these were the best candidates for putative promoters. The promoters of the *orf16* homolog and *orf87-trkA* in *M. barkeri*, and *orf40-trkA* in *M. thermophila*, perfectly match the consensus referred to above (Brown et al., 1989). The upstream region of *trkA* in *M. barkeri* contains the sequence AACTTTTTATATT that closely resembles the consensus and overlaps *orf87*—it is located 100 bps upstream of the translation start codon of *trkA*. This sequence might be too far away from *trkA* to serve as its promoter, but this possibility cannot be ruled out: it is possible that two transcripts from the *M. barkeri trkA* gene are produced, one including the transcript of *orf87* fused with that of *trkA*, and the other encompassing just *trkA*.

The putative promoters identified for *grpE* do not match the consensus entirely, yet they all contain conserved TATA boxes with the exception of the *Halobacterium* sp. NRC-1 gene, which does not have a consensus promoter or anything resembling a TATA-box. *Halobacterium* sp. NRC-1 is a halophile for which a differing consensus promoter has been described (Brown et al., 1989). Therefore, no conclusion can be drawn regarding this particular promoter. The *grpE* promoters in *M. barkeri*, *M. thermoautotrophicum*, and *Halobacterium* sp. NRC-1 contain a potential archaeal heat-shock promoter element, CGAA, upstream of the TATA box, while the promoters in *M. mazeii* and *F. acidarmanus* contain only the sequence GAA in the equivalent location (see above, Section X.A).

The *Methanosarcinae hsp70(dnaK)* promoters, including the TATA box, match the consensus only partially, and the *hsp70(dnaK)* potential promoters identified in the other organisms also bear low resemblance to the consensus (Table 4). The latter promoters, however, are the best match we could find with regard to the consensus and, for the most part, are at positions with respect to the translation start codon where functional promoters are usually found. The *M. mazeii* S-6 *hsp70(dnaK)* promoter has been tested experimentally, but similar data are still needed to determine whether the promoters identified for this gene in other archaea (Table 4) are functional, or they are located somewhere else and have sequences that differ greatly from the consensus for non-stress genes. The latter is a possibility because it would seem likely that stress genes would have distinct promoters from those of other genes with roles in unstressed cells, as suggested from the data already available for *M. mazeii*.

The sequence GAA, a potential archaeal stress element, is present upstream of the *hsp70(dnaK)* promoter identified in *M. mazeii* S-6, *M. thermophila*, and *M. thermoautotrophicum* (Table 4).

Of the eight archaea examined, none the promoters identified for *hsp40(dnaJ)* resembles the canonical consensus sequence (Table 4). The *hsp40(dnaJ)* promoters of *M. thermoautotrophicum* and *Methanosarcinae*, with the exception of that in *M. mazeii*, are T rich. *T. acidophilum*, *T. volcanium*, and *F. acidarmanus* are the only archaea studied that have the GAA element upstream of the *hsp40(dnaJ)* promoter (Table 4).

The promoters of *Halabacterium* sp. NRC-1 listed in Table 4 were identified based on the consensus sequence for non-stress genes in methanogens rather than that for halophiles (both as described by Brown et al., 1989), due to the fact that the three *cct* heat-inducible genes in *H. volcanii*, briefly discussed in Section X.A, have TATA boxes like those of the methanogen's promoters (Thompsen and Daniel, 1998). In addition, we examined the upstream regions of these genes to see if they had promoters resembling the consensus for halophiles, but the candidates found either matched the consensus more weakly than those described in Table 4, or were located far away from the translation start codon (data not shown).

The CGAA or GAA elements that are present in the promoters described in this report either flank or overlap the BRE (see Section VI.E). This observation implies that the BRE and, consequently, also TFB might be involved in the induction of these genes in response to stress.

2. Intergenic Spacing

The degree of conservation of the *hsp70(dnaK)* locus was further assessed by comparing the lengths of intergenic regions. *M. mazeii* S-6, *M. acetivorans*, and *M. barkeri* contain relatively long (over 100 bp) regions separating the three stress genes, whereas the loci of *M. thermoautotrophicum* and *Halobacterium* sp. NRC-1 have only one intergenic region of similar length (i.e., 75 bp) that between *hsp70 (dnaK)* and *hsp40(dnaJ)*. The regions separating the stress genes in *T. acidophilum*, *F. acidarmanus*, and *T. volcanium* are all short or absent ranging from 0 to 20 bp. The regions between *grpE* and *hsp70(dnaK)* in *M. thermoautotrophicum*, *T. acidophilum*, *F. acidarmanus*, and *T. volcanium* are no longer than 14 bp, respectively, which would mean absence of room for promoters or even other shorter transcriptional signals, unless these elements overlap the protein-coding region of the preceding gene. For example, a potential *hsp70(dnaK)* promoter identified in *F. acidarmanus* overlaps by 9 bp the 3′ end of the *grpE* protein-coding region, and a similar situation was found for *T. acidophilum*. In *M. thermoautotrophicum*, a promoter-like sequence could be detected on the 5′ end of the *hsp70(dnaK)* protein-coding region. The positions (especially in *M. thermoautotrophicum*) of these putative promoters and their lack of high homology with the consensus argue against being functional, genuine promoters. However, this has to be investigated further experimentally. Archaea are a constant source of surprises by comparison with the much better known bacteria and eukaryotes, and one must approach them with an open mind and innovative approaches.

If one considers the characteristics of the intergenic regions, it appears likely that at least *grpE* and *hsp70(dnaK)* are transcribed together and would produce a polycistronic message in *M. thermoautotrophicum*, *T. acidophilum*, and *F. acidarmanus*, but, here

again, experimental data are necessary to validate this conclusion.

The position of *trkA* relative to *hsp40(dnaJ)* is different in *M. barkeri* and *M. thermophila* from that in *M. mazeii* S-6. Even if one were to consider the Orfs that precede *trkA* in these organisms, the intergenic region between *hsp40(dnaJ)* and *trkA* is considerably longer in *M. barkeri* and *M. thermophila* than in *M. mazeii* S-6.

3. Repeat Downstream of hsp40(dnaJ) in Methanosarcina barkeri

According to the preliminary data at http://www.jgi.doe.gov/JGI_microbial/html/ , the region between *hsp70(dnaJ)* and *orf87-trkA* in *M. barkeri* contains 19 copies of a 12-nucleotide segment, 5′-TTCCTTTTCTTC-3′, beginning147 bp downstream of *hsp70(dnaK)* translation stop codon. Intriguingly, this sequence is composed exclusively of pyrimidines, which means that the complementary strand has only purines. The function of this sequence, which is not present in the *hsp70(dnaK)* loci of *M. mazeii* S-6 and *M. thermophila*, is not known.

E. Interaction of Basal Transcription Factors with Promoters

1. In Vitro Experiments

The mechanism by which the archaeal *hsp70(dnaK)* locus genes are induced by stress, for example, heat shock, has not yet been fully elucidated. We have began to investigate this mechanism using *M. mazeii* as model system for the reasons mentioned in Section I and XI. One approach has been that of examining the interaction of basal transcription factors with the stress- and non-stress-gene promoters in parallel to allow for comparative analyses. We studied the interaction between basal transcription factors with promoters *in vitro* and *in vivo*, and whenever possible we carried out experiments under conditions imitating both physiological (non-stress) and stress situations. Some of the results of *in vitro* experiments performed with recombinant, purified *M. mazeii* S-6 TBP and TFB are briefly discussed here, while the *in vivo* experiments are presented in Section XII.E.2 below.

It was found that TBP and TBP-TFB undergo conformational changes after interaction with stress-gene promoters, as revealed by circular dichroism (CD) spectroscopy (Thomsen et al., 2001). These conformational changes were distinctive for each promoter and were not observed when the factors bound non-stress-gene promoters (Figure 2). The percentage of *alpha*-helices in TBP and TBP-TFB increased after their interaction with the promoter of *hsp70(dnaK)*, but decreased after their binding to the promoters of *grpE* and *hsp40(dnaJ)*. The binding of these transcription factors to the *hsp40(dnaJ)* promoter produced the most pronounced conformational change. This is interesting considering that of the three stress-gene promoters, the promoter of *hsp40(dnaJ*) matches the consensus the least, and that the promoter for *orf16*, which matches the consensus 100% (Table 4), did not cause conformational changes. Along the same lines, the least amount of conformational change was produced by the *grpE* promoter, which of the three stress-gene promoters bears the most similarity with the consensus (Table 4). Thus, the closer a promoter was to the consensus the least change it induced on the basal transcription factors, and when the promoter matched the consensus completely it did no cause any change. Conversely, the greater the dissimilarity of the

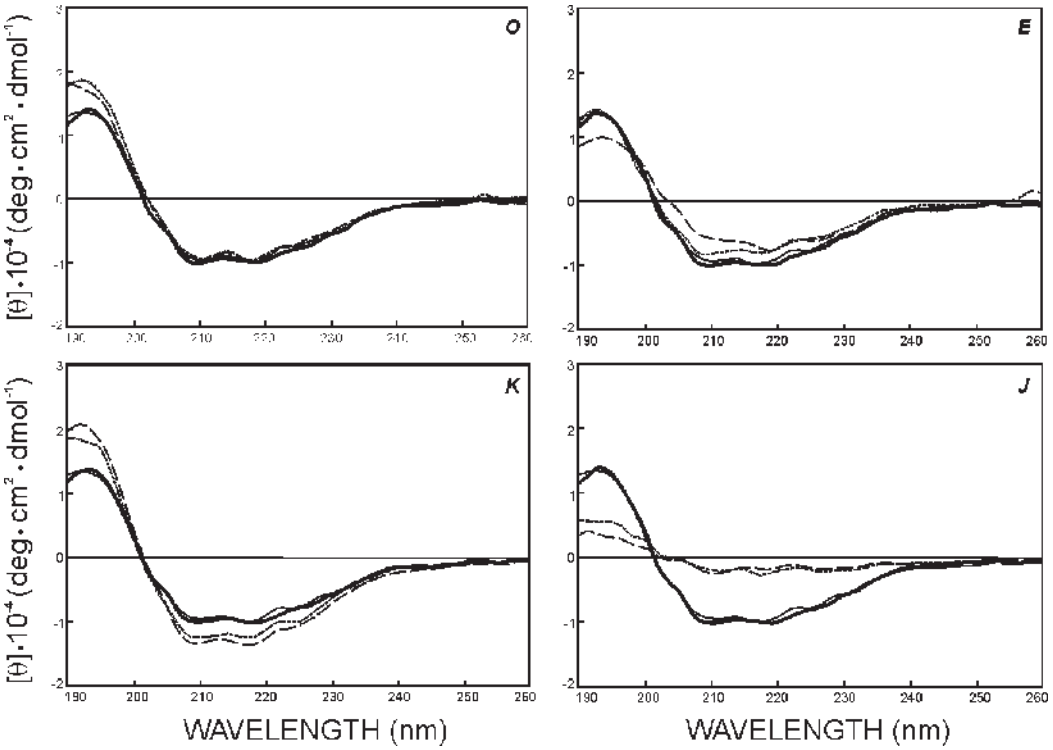

FIGURE 2. Circular dichroism (CD) spectra of TBP alone (thick line) and TBP-TFB (thin line) by themselves (solid line) and in the presence of promoters for the *M. mazeii* S-6 genes *orf16* (*O*), *hsp70(dnaK)* (*K*), *grpE* (*E*), and *hsp40(dnaJ)* (*J*) (broken line, TBP-promoter; dotted line, TBP-TFB-promoter). (Reproduced from Thomsen et al., 2001, with permission.)

promoter with the consensus, the greater the conformational change suffered by the transcription factors. Also, there was an inverse relationship between conformational change and level of gene transcription, inasmuch as the genes responding with the least increase in transcript in response to heat shock were those whose promoters induced the greatest changes (EC de M, and AJLM, in preparation). The role of transcription factor conformation, and its changes after interaction with the promoter in the regulation of constitutive (basal) and stress-induced transcriptions are currently being investigated.

What would be the function of the conformational changes? They may either facilitate or impede transcription. It is possible that the changes result in the transcription of these genes at low levels, for example, in constitutive expression. Under stress (e.g., heat shock), an as yet uncharacterized factor might interact with the "deformed", promoter-bound basal transcription factors, forcing them back into their proper conformations and thus promoting the higher levels of transcription observed during the stress response. If these conformational changes in TBP and TBP-TFB did by themselves impair stress-gene transcription in the absence of stress (low constitutive expression), one may think that another mechanism for increasing transcription in response to stress would be provided by the changes in DNA topology resulting from stress. The changes in DNA topology would enhance the interaction between the "deformed" transcription factors and promoter. (See Section X.B.)

2. In Vivo Experiments

TBP in the cytosol of *M. mazeii* S-6 cells binds promoter DNA as demonstrated

by EMSA using cell lysates (De Biase et al., 2002). Lysates from unstressed cells produced one shifted band, which was shown to contain TBP but no TFB, and lysates from heat-shocked cells produced two bands. One of these bands was identical to that produced by lysates of unstressed cells and the other contained TBP and TFB (Figure 3). Also, stress increased the magnitude of the TBP-containing band, indicating that more TBP bound to the promoter. Taken together, these results suggested that heat stress increased cytosolic TBP and/or TBP-DNA and TBP-TFB interactions. Competitive inhibition experiments showed that TBP from stressed cells bound more firmly to the promoter than did TBP from unstressed cells (Figure 4).

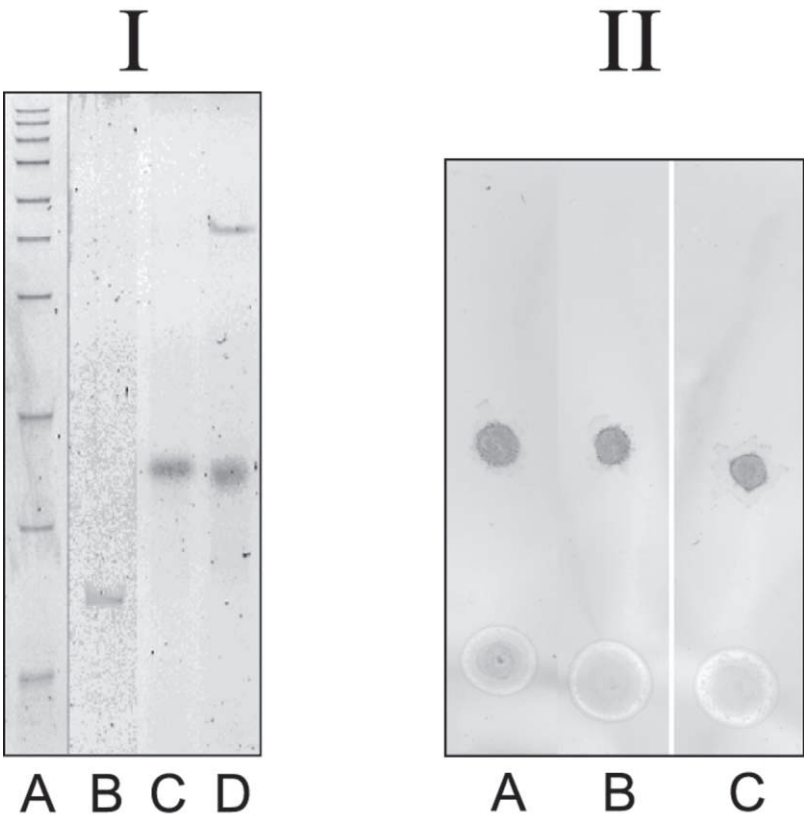

FIGURE 3. Lysates from heat-shocked cells produce an additional shifted band in EMSA with DNAs containing the promoter regions of the *M. mazeii* S-6 *hsp70(dnaK)*-locus stress genes; this band did not appear when extracts from unstressed cells were used. In the experiment shown here, EMSA was carried out with probe ***K*** (with the *hsp70(dnaK)* promoter), labeled at the 5'-end with fluorescein. **Panel I,** EMSA. Lane A: fluorescein-labeled DNA reference ladder (Amersham); Lane B: free probe; Lanes C and D: the probe was incubated with extract from *M. mazeii* cells maintained at the optimal temperature for growth, or from cells heat shocked at 45°C for 15 min, respectively. **Panel II,** Dot blots. A: faster band reacted with anti-TBP antibody (top dot) or normal rabbit serum (bottom dot); B and C**:** slower shifted band reacted with anti-TBP and anti-TFB antibodies, respectively (top dots); bottom dots are negative controls as in A. The anti-TBP and anti-TFB antibodies do not crossreact. (Reproduced from De Biase et al., 2002, with permission.)

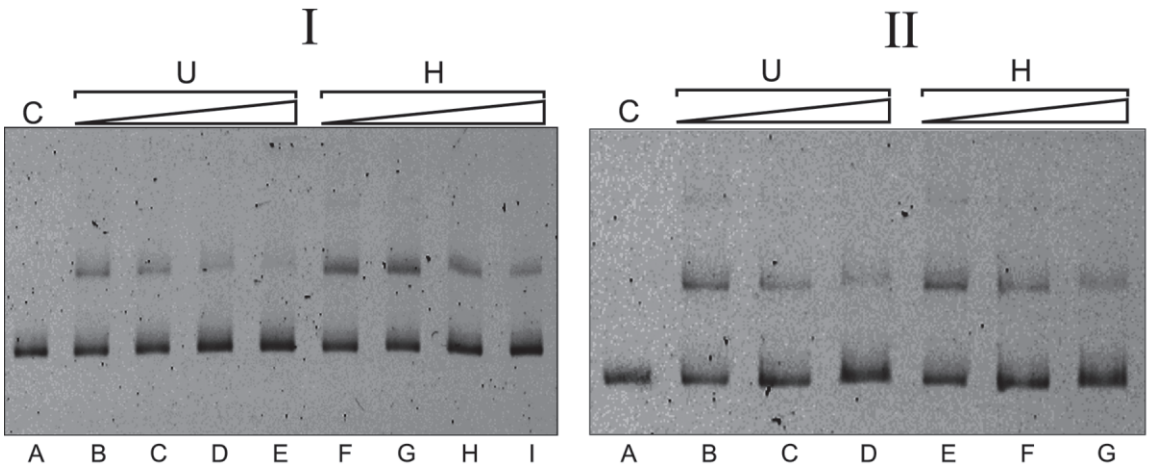

FIGURE 4. EMSA. Competition experiment demonstrating that stress-gene promoter-TBP interaction is stronger with cytosolic TBP from heat-shocked cells when compared with unstressed counterparts. **Panel I**. Competition experiment using unlabeled vs. fluorescein-labeled ***J*** probe (DNA with the *M. mazeii* S-6 *hsp40(dnaJ)* promoter). Lane A: free probe; Lanes B-E: probe incubated with extract from unstressed cells (U); Lanes F-I: probe incubated with extract from heat-shocked cells (H). **C**: competitor added, none (lanes B and F), 2 µl (lanes C and G), 4 µl (lanes D and H), and 8 µl (lanes E and I)**. Panel II**. Competition experiments using unlabeled 29-mer bearing the *hsp70(dnaK)* promoter vs. fluorescein-labeled ***J*** probe; the EMSA was carried out with the latter, as in Panel I. Lane A: free probe; Lanes B-D: probe incubated with extract from unstressed cells (U); Lanes E-G: probe incubated with extract from heat-shocked cells (H). **C**: competitor added, none (lanes B and E), 2 µl (lanes C and F), and 4 µl (lanes D and G). (Reproduced from De Biase et al., 2002, with permission.)

The data from the *in vitro* experiments showed that the basal transcription factors undergo conformational changes after interaction with the promoters of stress genes only (see Section XII.E.1), and the data discussed in this section indicate that stress modifies the interaction of the factors with these promoters to increase the amount and/or stability of the transcription initiation complex. These two lines of observations must be pursued because they provide a promising starting point to elucidate how differential gene expression occurs in *M. mazeii* S-6, and possibly also in other archaea, with such a dearth of basal factors and promoters (essentially only two basal factors, perhaps three if we consider TFE, and a single promoter per gene, because constitutive and stress-induced transcription proceed from the same promoter). *M. mazeii* must manage to initiate transcription of a variety of stress genes using only this restricted battery of basal factors and DNA sites, and yet accomplish differential gene induction. One way to achieve a diversity of mechanisms that would differentially induce genes at different times and locations would be to generate a diversity of conformations in the basal factors, so each conformation would interact preferentially, or exclusively, with just one regulatory factor recruited from a limited set of such factors. In this set, each component would have preference for one conformation and no affinity for the others, which would have their own preferred factor each. Such dynamics of conformation-dependent factor-binding specificity, and the resulting differential combinations with regulatory factors and promoters, can potentially generate a large variety of tran-

scription initiation complexes with different gene-inducing capabilities using very few components (TBP, TFB, promoter DNA, regulatory factor).

F. Conclusions Regarding the *hsp70(dnaK)* Locus

The locus in which the three stress genes of the molecular chaperone machine are clustered is conserved in gene order and orientation. The regions flanking these genes, however, are less conserved, except possibly among organisms belonging to the same genus or that are closely related, as in the case of the *Methanosarcinae*. The intergenic regions between the three stress genes are very short in the *hsp70(dnaK)* loci of *T. acidophilum* and *F. acidarmanus*. Of the three genes in these organisms, only *grpE* has a putative promoter that strongly matches the consensus and contains a fully conserved canonical TATA box. Therefore, its does not seem likely that these genes are transcribed individually as they are in *M. mazeii* S-6 and *M. thermophila*.

The *hsp70(dnaK)* locus of *M. thermoautotrophicum* is unique in that the intergenic region between *gprE* and *hsp70(dnaK)* is small, while that between *hsp70(dnaK)* and *hsp40(dnaJ)* is large. Neither one of these intergenic regions contains a strong promoter sequence by comparison with the consensus, although the promoter identified for *hsp40(dnaJ)* does seem to match the consensus more than the promoter of *hsp70 (dnaK)*. It is possible that *grpE* and *hsp70(dnaK)* are transcribed together, while the transcription of *hsp40(dnaJ)* is independent of the other two.

The *hsp70(dnaK)* locus of *Halobacterium* sp. NRC-1 is unique among the loci examined in that it contains an additional open reading frame between *grpE* and *hsp70(dnaK)*. The intergenic regions in this locus appear large enough to contain promoters. It is possible that genes in this locus are transcribed independently of one another, especially because the locus contains an additional gene between *grpE* and *hsp70(dnaK)*.

Also, it seems likely that the stress genes of the *hsp70(dnaK)* locus in *M. barkeri* are transcribed individually rather than together with production of a single, long mRNA, as they are in *M. mazeii* S-6 and *M. thermophila*, as suggested by the spacing of the genes and the putative promoters we could identify. The promoters for *hsp70(dnaK)* and *hsp40*(*dnaJ)* differed from the consensus but resembled the *M. mazeii* S-6 *hsp70(dnaK)* and *hsp40*(*dnaJ)* promoters (Table 4). Based on the information regarding TFE (see Section IV.B), it is reasonable to hypothesize that these promoters are potentially TFE sensitive. The discrepancies within these promoters are different from those that were recognized by *S. solfataricus* TFE (Bell et al., 2001). The lack of conservation among archaeal TFEs, and the fact that *M. thermoautotrophicum* TFE cannot be substituted for *A. fulgidus* TFE (Hanzelka et al., 2001), suggest that *Methanosarcinae* TFEs would recognize different promoter sequence discrepancies than would *Sufolobus* TFEs. Hence, one might speculate that *Methanosarcinae* TFEs would recognize the variations from the consensus present in the promoters of *hsp70(dnaK)* and *hsp40(dnaJ)* and facilitate the initiation of transcription. A mechanism can also be envisaged that would be complementary to TFE, or constitute an alternative to it, for specific stress-gene induction. For example, a regulatory factor able to recognize certain deviations in the promoter sequence, or a conformational change in the basal factors, could specifically enhance stress-induced transcription.

XIII. CONCLUSION

The study of transcription initiation and regulation in the Archaea is in its infancy.

Wide gaps in the knowledge of molecular mechanisms and regulatory networks are evident compared with what is known for the other prokaryotes, the bacteria, and the eukaryotes. This is also the situation in regard to stress genes. Nonetheless, a considerable amount of information on the basic components of the archaeal transcription machinery is already available, and this information will be instrumental for advancing the field in the years to come.

A review of current data reveals that although archaeal basal transcription factors such as TBP, TFB, and TFE are similar to their eukaryotic equivalents, they also differ from them in various ways. In addition, a certain degree of variation occurs among the archaeal molecules from different species, which suggests a diversity of roles and/or mechanisms of action. Various archaea possess more than one *tbp* gene, whose encoded proteins differ from one another to an extent compatible with diversity of function. The same is true for the *tfb* gene. Interestingly, not every species with more than one *tbp* has a matching number of *tfb* genes, or vice versa. The occurrence of several TBPs and one or more TFBs, or vice versa, in the same species suggests that different TBP-TFB complexes can form perhaps as a means for differential gene expression: a given complex would initiate transcription of only one gene or group of genes. In addition, variations in the TATA box and other promoter elements may also add to the diversity of possible transcription initiation complexes and thus endow gene induction with specificity for a single gene or a group of genes that need to function coordinately. This diversity may be central to stress-gene induction in response to stress, if and when stressors cause alterations in DNA topology and transcription-factor conformation. Both these alterations may provide a further mechanism of specificity for any given combination TBP-TFB-DNA (promoter and promoter elements). Recent data indicate that basal transcription factors change conformation after interaction with the promoters of the stress genes in the *hsp70(dnaK)* locus of *M. mazeii* S-6, and that the change is distinctive for each promoter. Studies *in vivo* revealed that heat stress promotes TBP-TFB and TBP-TFB-DNA (*hsp70(dnaK)*-locus gene promoters) interactions, and that these complexes associate more tightly than they in the absence of stress. The role of stress-induced changes in transcription-factor conformation and DNA-topology in the specific induction of stress genes is currently under investigation.

XIV. PERSPECTIVES

Various territories remain open for investigation within the realm of archaeal stress genes. *In vitro* and *in vivo* studies must be pursued to elucidate intimate molecular interactions and the role of different components (aa and nucleotides) of the transcription machinery in the preferential induction of individual genes at specific times during the cell's cycle and life. Furthermore, interatomic bonds in the transcription initiation complex ought to be characterized to understand the subtle forces that start gene transcription. Also, regulatory networks must be delineated, and groups of stress genes that are coordinately expressed in unstressed cells and in response to stress must be identified. This implies identification of *trans*-acting factors and their sites of interaction on the DNA and basal factors. Studies using imaging techniques for multimolecular complexes, among others, should be performed to unveil the dynamics of transcription initiation complex assembly with and without candidate regulators, the contortions of the complex caused by the interaction of its components and other factors pertinent to energy exchange, and the effect

of stress on these molecular movements and atomic rearrangements. Information at the atomic, molecular, and multimolecular-complex levels will be critical to devise strategies and methods for improving stress response and cell survival. A promising avenue leads to the manipulation of stress genes and their products using genetic and protein engineering techniques to improve cell resistance to stressors, accelerate recovery after stress, and prevent the consequences of stress in situations in which it is possible to anticipate the impact of stressors, for example, in elective surgery and warfare activities. Also, protein engineering strategies ought to help in designing compounds that will either inhibit or enhance transcription factor interaction and gene induction, as required, to maintain cell physiology under normal and stressful conditions.

ACKNOWLEDGMENTS

Part of the authors' work reported here was supported by grants from the Department of Energy, USA, and the National Science Foundation. We thank Nhan Ngo, Tram Ahn Tran, and the Ronald E. McNair Scholars Lolita Simmons and Hermine Voltaire and the staff of the Photography and Illustration Unit for their help and enthusiasm; the Ronald E. McNair Program for support of L. S. and H. V.; and the San Francisco Foundation for partially supporting A. J. H. Data from http://www.jgi.doe/gov/JGI_microbial.htm was provided freely by the US DOE Joint Genome Institute for use in this publication only.

REFERENCES

Albright, S.R., Tjian, R. 2000. TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* **242**, 1–13.

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

Baliga, N.S., DasSarma, S. 2000. Saturation mutagenesis of the haloarchaeal bop gene promoter: identification of DNA supercoiling sensitivity sites in the absence of TFB recognition element and UAS enhancer activity. *Mol. Microbiol*. **36**, 1175–1183.

Barns, S.M. Delwiche, C.F., Palmer, J.D., Pace, N.R. 1996. Perspectives on archaeal diversity, thermophily, and monophyly from environmental rRNA sequences. *Proc. Natl. Acad. Sci. USA* **93**, 9188–9193.

Bauer, A., Chauvet, S., Huber, O., Usseglio, F., Rothbacher, U., Aragnol, D., Kemler, R., Pradel, J. 2000. Pontin52 and reptin52 function as antagonistic regulators of beta-catenin signalling activity. *EMBO J*. **19**, 6121–6130.

Bauer, A., Huber, O., Kemler, R. 1998. Pontin52, an interaction partner of *beta*-catenin, binds to the TATA box binding protein. *Proc. Natl. Acad. Sci. USA* **95**, 14787–14792.

Bell, S.D., Botting, C.H., Wardleworth, B.N., Jackson, S.P., White, M.F. 2002. The interaction of Alba, a conserved archaeal chromatin remodeling protein, with Sir2 and its regulation by acetylation. *Science* **296**, 148–151.

Bell, S.D., Brinkman, A.B., van der Oost, J., Jackson, S.P. 2001. The archaeal TFIIE *alpha* homologue facilitates transcription initiation by enhancing TATA-box recognition. *EMBO Reports* **21**, 133–138.

Bell, S.D., Cairns, S.S., Robson, R.L., Jackson, S.P. 1999a. Transcriptional regulation of an archaeal operon *in vivo* and *in vitro*. *Mol. Cell* **4**, 971–982.

Bell, S.D., Jackson, S.P. 2000a. The role of transcription factor B in transcription initiation and promoter clearance in the archaeon *Sulfolobus acidocaldarius*. *J. Biol. Chem.* **275**, 12934–12940.

Bell, S.D., Jackson, S.P. 2000b. Mechanism of autoregulation by an archaeal transcriptional repressor. *J. Biol. Chem.* **275**, 31624–31629.

Bell, S.D., Jaxel, C., Nadal, M., Kosa, P.F., Jackson, S.P. 1998. Temperature, template topology and factor requirements for archaeal transcription. *Proc. Natl. Acad. Sci. USA* **95**, 15218–15222.

Bell, S.D., Kosa, P.L., Sigler, P.B., Jackson, S.P. 1999b. Orientation of the transcription preinitiation complex in archaea. *Proc. Natl. Acad. Sci. USA* **96**, 13662–13667.

Berg, J.M. 1990. Zinc fingers and other metal-binding domains. *J. Biol. Chem*. **265**, 6513–6516.

Bergqvist, S., O'Brien, R., Ladbury, J.E. 2001. Site-specific cation binding mediates TATA binding protien-DNA interaction from a hyperthermophilic archaeon. *Biochemistry* **40**, 2419–2425.

Bergqvist S., Williams M.A., O'Brien R., Ladbury J.E. 2002. Reversal of halophilicity in a protein-DNA interaction by limited mutation strategy. *Structure* **10**, 629–637.

Best. A.A., Olsen, G.J. 2001. Similar subunit architecture of archaeal and eukaryal RNA polymerases. *FEMS Microbiol. Lett*. **195**, 85–90.

Blaszczak, A., Georgopoulos, C., Liberek, K. 1999. On the mechanism of FtsH-dependent degradation of the sigma-32 transcriptional regulator of *Escherichia coli* and the role of the DnaK chaperone machine. *Mol. Microbiol*. **31**, 157–166.

Boyd, J., Oza. M.N., Murphy, J.R. 1990. Molecular cloning and DNA sequence analysis of diphtheria *tox* iron-dependent regulatory elements (*dtxH)* from *Corynebacterium diphtheriae*. *Proc. Natl. Acad. Sci. USA* **87**, 5968–5972.

Brown, J.W., Daniels, C.J., Reeve, J.R. 1989. Gene structure, organization, and expression in archaeabacteria. *CRC Crit. Rev. Microbiol.* **16**, 287–338.

Bult, C.J., White, O., Olsen, G.J., Zhou, L., Fleischmann, R.D., Sutton, G.G., Blake, J.A., Fitzgerald, L.M., Clayton, R.A., Gocayne, J.D., Kerlavage, A.R., Dougherty, B.A., Tomb, J.-F., Adams, M.D., Reich, C.I., Overbeek, R., Kirkness, E.F., Weinstock, K.G., Merrick, J.M., Glodek, A., Scott, J.L., Geoghagen, N.S.M., Weidman, J.F., Fuhrmann, J.L., Nguyen, D., Utterback, T.R., Kelley, J.M., Peterson, J.D., Sadow, P.W., Hanna, M.C., Cotton, M.D., Roberts K.M., Hurst M.A., Kaine, B.P., Borodovsky, M., Klenk, H.-P., Fraser, C.M., Smith, H.O., Woese, C.R., Venter, J.C. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073.

Burke, T.W., Kadonaga, J.T. 1997. The downstream core promoter element, DPE, is conserved from *Drosophila* to humans and is recognized by $TAF_{II}60$ of *Drosophila*. *Genes Dev*. **11**, 3020–3031.

Bustard, K., Gupta, R.S. 1997. The sequences of heat shock protein 40 (DnaJ) homologs provide evidence for a close evolutionary relationship between the *Deinococcus-Thermus* group and cyanobacteria. *J. Mol. Evol*. **45**, 193–205.

Chen, H-T., Legault, P., Glushka, J., Omichinski, J.G., Scott, R.A. 2000. Structure of a (Cys_3His) zinc ribbon, a ubiquitous motif in archaeal and eucaryal transcription. *Prot. Sci*. **9**, 1743–1752.

Clarens, M., Macario, A.J.L., Conway de Macario, E. 1995. The archaea *dnak-dnaJ* gene cluster: organization and expression in the methanogen *Methanosarcina mazei*. *J. Mol. Biol*. **250**, 191–201.

Cohen-Kupiec, R., Blank, C., Leigh, J.A. 1997. Transcriptional regulation in archaea: *in vitro* demonstration of a respressor binding site in a methanogen. *Proc. Natl. Acad. Sci. USA* **94**, 1316–1320.

Cohen-Kupiec, R., Marx, C.J., Leigh, J.A. 1999. Function and regulation of *glnA* in the

methanogenic archaeon *Methanococcus maripaludis*. *J. Bacteriol*. **181**, 256–261.

Conway de Macario, E., Clarens, M., Macario, A.J.L. 1995. Archaeal *grpE*: transcription in two different morphologic stages of *Methanosarcina mazei* and comparison with *dnaK* and *dnaJ*. *J. Bacteriol*. **177**, 544–550.

Conway de Macario, E., Dugan, C.B., Macario, A.J.L. 1994. Identification of a *grpE* heat-shock homolog in the archaeon *Methanosarcina mazei*. *J. Mol. Biol*. **240**, 95–101.

Conway de Macario, E., Guerrini, M., Dugan, C.B., Macario, A.J.L. 1996. Integration of foreign DNA in an intergenic region of the archaeon *Methanosarcina mazei* without effect on transcription of adjacent genes. *J. Mol. Biol*. **262**, 12–20.

Conway de Macario, E., Macario, A.J.L. 1995. Transcription of the archaeal *trkA* homolog in *Methanosarcina mazei* S-6. *J. Bacteriol*. **177**, 6077–6082.

Cramer, P., Buchnesll, D.A., Fu., J., Gnatt, A.L. Maier-Davis, B., Thompson, N.E., Burgess., R.E., Edwards, A.M., David., P. R., Kronberg, R.D. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. *Science* **288**, 640–649.

Dahlke, I., Thomm, M. 2002. A *Pyrococcus* homolog of the leucine-responsive regulatory protein, LprA, inhibits transcription by abrogating RNA polymerase recruitment. *Nucl. Acids Res*. **30**, 701–710.

Darcy, T.J., Hausner, W., Awery, D.E., Edwards, A.M., Thomm, M., Reeve, J.N. 1999. *Methanobacterium thermoautotrophicum* RNA polymerase and transcription *in vitro*. *J. Bacteriol*. **181**, 4424–4429.

De Biase, A., Macario, A.J.L., Conway de Macario, E. 2002. Effect of heat stress on promoter binding by transcription factors in the cytosol of the archaeon *Methanosarina mazeii*. *Gene* **282**, 189–197.

DeDecker, B.S., O'Brien R., Fleming, P.J., Geiger, J.H., Jackson, S.P., Sigler, P.B. 1996. The crystal structure of a hyperthermophilic archaeal TATA-box binding protein. *J. Mol. Biol*. **264**, 1072–1084.

Deppenmeier, U., Johann, A., Hartsch, T., Merkl, R., Schmitz, R.A., Martinez-Arias, R., Henne, A., Wiezer, A., Baumer, S., Jacobi, C., Brueggemann, H., Lienard, T., Christmann, A., Boemeke, M., Steckel, S., Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H-P., Gunsalus, R.P., Fritz, H-J., Gottschalk, G. 2002. The genome of *Methanosarcina mazei*: evidence for lateral gene transfer between bacteria and archaea. *J. Mol. Microbiol. Biotechnol*. **4**, 453–461.

Evans, R., Fairley, J.A., Roberts, S.G.G. 2001. Activator-mediated disruption of sequence-specific DNA contacts by the general transcription factor TFIIB. *Genes Dev*. **15**, 2945–2949.

Fernandes, M., O'Brien, T., Lis, J.T. 1994. Structure and regulation of heat shock gene promoters. In: *The Biology of Heat Shock Proteins and Molecular Chaperones* (Morimoto, R.I., Tissieres, A., and Georgopoulos, C., Eds.), Cold Spring Harbor Laboratory Press, Plainview, NY, pp. 375–393.

Fitz-Gibbon, S.T., Ladner, H., Kim, U.-J., Stetter, K.O., Simon, M.I., Miller, J.H. 2002. Genome sequence of the hyperthermophilic crenarchaeon *Pyrobaculum aerophilum*. *Proc. Natl. Acad. Sci. USA* **99**, 984–989.

Galagan, J.E., Nusbaum, C., Roy, A., Endrizzi, M.G., Macdonald, P., FitzHugh, W., Calvo, C., Engels, R., Smirnov, S., Atnoor, D., Brown, A., Allen, N., Naylor, J., Stange-Thomann, N., DeArellano, K., Johnson, R., Linton, L., McEwan, P., McKernan, K., Talamas, J., Tirrell, A., Ye, W., Zimmer, A., Barber, R.D., Cann, I., Graham, D.E., Grahame, D.A., Guss, A., Hedderich, R., Ingram-Smith, C., Kuettner, H.D., Krzycki, J.A., Leigh, J.A., Li, W., Liu, J., Mukhopadhyay, B., Reeve,

J.N., Smith, K., Springer, T., Umayam, L.A., White, O., White, R.H., Conway de Macario, E., Ferry, J.G., Jarrell, K.F., Jing, H., Macario, A.J.L., Paulsen, I., Pritchett, M., Sowers, K.R., Swanson, R.V., Zinder, S.H., Lander, E., Metcalf, W.W., Birren, B. 2002. The genome of *Methanosarcina acetivorans* reveals extensive metabolic and physiological diversity. *Genome Res.* **12**, 532–542.

Gautier, T., Berges, T., Tollervey, D., Hurt, E. 1997. Nucleolar KKE/D repeat proteins Nop56p and Nop58p interact with Nop1p and are required for ribosome biogenesis. *Mol. Cell Biol*. **17**, 7088–7098.

Gelfand, M.S., Koonin, E.V., Mironov, A.A. 2000. Prediction of transcription regulatory sites in archaea by a comparative genomic approach. *Nucl. Acids Res.* **28**, 695–705.

Georgopoulos, C., Liberek, K., Zylicz, M. and Ang, D. 1994. Properties of the heat shock proteins of *Escherichia coli* and the autoregulation of the heat shock response. In: *The Biology of Heat Shock Proteins and Molecular Chaperones* (Morimoto, R. I., Tissiéres, A., and Georgopoulos, C., Eds.), pp. 209–249, Cold Spring Harbor Laboratory Press, Plainview, NY;

Gohl, H. P., Groendahl, B., Thomm, M. 1995. Promoter recognition in archaea is mediated by transcription factors: identification of transcription factor aTFB from *Methanococcus thermolithotrophicus* as archaeal TATA-binding protein. *Nucl. Acids Res*. **23**, 3837–3841.

Gregor, D., Pfiefer, F. 2001. Use of a halobacterial *bgaH* reporter gene to analyse the regulation of gene expression in halophillic archaea. *Microbiol*ogy **147**, 1745–1754.

Gribaldo, S., Lumia, V., Creti, R., Conway de Macario, E., Sanangelantoni A., Cammarano, P. 1999. Discontinuous occurrence of the *hsp70(dnaK)* gene among archaea and sequence features of HSP70 suggest a novel outlook on phylogenies inferred from this protein. *J. Bacteriol.* **181**, 434–443.

Gupta, R.S. 1998. Protein phylogenies and signature sequences: a reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Micrbiol. Mol. Biol. Rev.* **62**, 1435–1491.

Gupta, R.S., Singh, B. 1992. Cloning of the HSP70 gene from *Halobacterium marimortui*: relatedness of archaebacterial HSP70 to its eubacterial homologs and a model of the evolution of the HSP70 gene. *J. Bacteriol*. **174**, 4594–4605.

Gupta, R.S., Singh, B. 1994. Phylogenetic analysis of 70kD heat shock protein sequences suggests a chimeric origin for the eukaryotic cell nucleus. *Curr. Biol*. **4**, 1104–1114.

Guzikevich-Guerstein, G., Shakked, Z. 1996. A novel form of the DNA double helix imposed on the TATA box by the TATA binding protein. *Nature Struct. Biol*. **3**, 32–37.

Hampsey, M. 1998. Molecular genetics of the RNA polymerase II general transcription machinery. *Microbiol. Mol. Biol. Rev*. **62**, 465–503.

Hanzelka, B.L, Darcy, T.J., Reeve, J.N. 2001. TFE, an archaeal transcription factor in *Methanobacterium thermoautotrophicum* related to eucaryal transcription factor TFIIE *alpha*. *J. Bacteriol*. **183**, 1813–1816.

Hausner, W., Thomm, M. 1993. Purification and characterization of a general transcription factor, aTFB, from the archaeon *Methanococcus thermolithotrophicus*. *J. Biol. Chem.* **268**, 24047–24052.

Hausner, W., Wettach, J., Hethke, C., Thomm, M. 1996. Two transcription factors related with the eucaryal transcription factors TATA-binding protein and transcription factor IIB direct promoter recognition by an archaeal RNA polymerase. *J. Biol. Chem.* **271**, 30144–30148.

Hethke, C., Bergerat, A., Hausner, W., Foterre, P., Thomm, M. 1999. Cell-free transcription at

95°: thermostability of transcriptional components and DNA topology requirements of *Pyrococcus* transcription. *Genetics* **152**, 1325–1333.

Hickey, A.J., Macario, A.J.L, Conway de Macario, E. 2000. Identification of genes in the genome of the archaeon *Methanosarcina mazeii* that code for homologs of nuclear eukcaryotic molecules involved in RNA processing. *Gene* **253**, 77–58.

Hochheimer, A., Hedderich, R., Thauer, R. 1998. The formylmethanofuran dehydrogenase iosenzymes in *Methanobacterium wolfei* and *Methanobacterium thermoautotrophicum*. Induction of molybdenum isoenzyme by molybdate and constitutive synthesis of tungsten isoenzyme. *Arch. Microbiol*. **170**, 389–393.

Hochheimer, A., Hedderich, R., Thauer, R. 1999. The DNA binding protein Tfx from *Methanobacterium thermoautotrophicum*: structure, DNA binding properties and transcription regulation. *Mol. Microbiol.* **31**, 641–650.

Hofman-Bang, J., Lange, M., Conway de Macario, E., Macario, A.J.L, Ahring, B.K. 1999. The genes encoding for the *hsp70(dnaK)* molecular chaperone machine occur in the moderate thermophillic archaeon *Methanosarcina thermophila* TM-1. *Gene* **238**, 387–395.

Huffman, J.L., Brennan, R.G. 2002. Prokaryotic transcription regulators: more than just the helix-turn-helix motif. *Curr. Opin. Struct. Biol.* **12**, 98–106.

Ikura, T. Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshii, M., Scully, R., Qin, J., Nakatani, Y. 2000. Involvement of te TIP60 histone acetylase comples in DNA repair and apoptosis. *Cell* **102**, 463–473.

Inostroza, J., Flores, O., Reinberg, D. 1991. Factors involved in specific transcription by mammalian RNA polymerase II: purification and functional analysis of general transcription factor IIE. *J. Biol. Chem.* **266**, 9304–9308.

Jackson-Fisher, A.J., Chitikila, C., Mitra, M., Pugh, B.F. 1999. A role for TBP dimerization in preventing unregulated gene expression. *Mol. Cell* **3**, 717–727.

Kadonaga J.T 1998. Eukaryotic transcription: an interlaced network of transcription factors and chromatin-modifying machines. *Cell* **92**, 307–313.

Kaine, B.P., Mehr, I. J., Woese, C.R. 1994. The sequence, and its evolutionary implications, of a *Thermococcus celer* protein associated with transcription. *Proc. Natl. Acad. Sci. USA* **91**, 3854–3856.

Kanemaki, M. Kurokawa, Y., Matsu-ura, T., Makino, Y., Masani, A., Okazaki, K., Morishita, T., Tamura, T. 1999. TIP49b, a new RuvB-like DNA helicase, is included in a complex together with another RuvB-like DNA helicase, TIP49a. *J. Biol. Chem.* **274**, 22437–22444.

Kanemaki, M., Makino, Y., Yoshida, T., Kishimoto, T., Koga, A., Yamamoto, K., Yamamoto, M., Moncollin, V., Egly, J.-M., Muramatsu, M., Tamura, T. 1997. Molecular cloning of a rat 49–kDa TBP-interacting protein (TIP49) that is highly homologous to the bacterial RuvB. *Biochem. Biophys. Res. Comm.* **235**, 64–68.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y., Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H., Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K., Nakamura, Y., Nomura, N., Sako, Y., Kikuchi, H. 1999. Complete genome sequence of an aerobic hyper-thermophilic crenarchaeon, *Aeropyrum pernix* K1. *DNA Res.* **6**, 83–101.

Kawarabayasi, Y., Hino, Y., Horikawa, H., Jin-no, K., Takahashi, M., Sekine, M., Baba,

S., Ankai, A., Kosugi, H., Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Otsuka, R., Nakazawa, H., Takamiya, M., Kato, Y., Yoshizawa, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Masuda, S., Yanagii, M., Nishimura, M., Yamagishi, A., Oshima, T., Kikuchi, H. 2001. Complete genome sequence of an aerobic thermoacidophilic crenarchaeon, *Sulfolobus tokodaii* strain 7. *DNA Res*. **8**, 123–140.

Kawarabayasi, Y., Sawada, M., Horikawa, H., Haikawa, Y., Hino, Y., Yamamoto, S., Sekine, M., Baba S.-I., Kosugi H., Hosoyama A., Nagai, Y., Sakai, M., Ogu, K., Otsuka, R., Nakazawa, H., Takamiya, M., Ohfuku, Y., Funahashi, T., Tanaka, T., Kudoh, Y., Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K.-I., Yoshizawa, T., Nakamu,Y., Robb, F.T., Horikoshi, K., Masuchi, Y., Shizuya, H., Kikuchi H. 1998. Complete sequence and gene organization of the genome of a hyper-thermophilic archaebacterium, *Pyrococcus horikoshii* OT3. *DNA Res*. **5**, 55–76.

Kawashima, T., Amano, N., Koike, H., Makino, S., Higuchi, S., Kawashima-Ohya, Y., Watanabe, K., Yamazaki, M., Kanehori, K., Kawamoto, T., Nunoshiba, T., Yamamoto, Y., Aramaki, H., Makino, K., Suzuki, M. 2000. Archaeal adaptation to higher temperatures revealed by genomic sequence of *Thermoplasma volcanium*. *Proc. Natl. Acad. Sci. USA* **97**, 14257–14262.

Kim, J.L., Nikolov, B.B., Burley, S.K. 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**, 520–527.

Kim, Y., Geiger, J.H., Hahn, S., Sigler, P.B. 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**, 512–520.

King, T.H., Decatur, W.A., Bertrand, E., Maxwell, E.S., Fournier, M.J. 2001. A well-connected and conserved nucleoplasmic helicase is required for production of box C/D/ and H/ACA snoRNAs and localization of snoRNP proteins. *Mol. Cell Biol.* **21**, 7731–7746.

Klenk, H.-P., Clayton, R.A., Tomb, J-F., White, O., Nelson, K.E., Ketchum, K.A., Dodson, R.J., Gwinn, M., Hickey, E.K., Peterson, J.D., Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides, N.C., Fleischmann, R.D., Quackenbush, J., Lee, N.H., Sutton G.G, Gill, S. Kirkness, E.F., Dougherty, B.A. McKenney, K. Adams, M.D., Loftus, B. Peterson, S. Reich, C.I., McNeil, L.K., Badger, J.H., Glodek, A. Zhou, L., Overbeek, R. Gocayne, J.D., Weidman, J.F., McDonald, L., Utterbach, T., Cotton, M.D., Spriggs, T., Artiach, P. Kaine, B.P., Sykes, S.M., Sadow, P.W., D'andrea, K.P., Bowman, C., Fujii, C., Garland, S.A., Mason, T.M., Olsen, G.J., Fraser, C.M., Smith H.O. Woese, C.R., Venter, J.C. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* **390**, 364–370.

Kletzin, A. 1992. Molecular characterization of the *sor* gene, which encodes the sulfur oxygenase/reductase of the thermoacidophilic Archaeum *Desulfurolobus ambivalens*. *J. Bacteriol*. **174**, 5854–5859.

Kosa, P.F., Ghosh, G., DeDecker, B.S., Sigler, P.B. 1997. The 2.1–Angstrom crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. USA* **94**, 6042–6047.

Kuldell, N.H., Buratowski, S. 1997. Genetic analysis of the large subunit of yeast transcription factor IIE reveals two regions with distinct function. *Mol. Cell Biol.* **17**, 5288–5298.

Krueger, K., Hermann, T., Armbruster, V., Pfeifer, F. 1998. The transcriptional activator GvpE for the Halobacterial gas vesicle genes resembles a basic region leucine-zipper regulatory protein. *J. Mol. Biol*. **279**, 761–771.

Kuo, Y.P., Thompson, D.K., St. Jean, A., Charlebois, R.L., Daniels, C.J. 1997.

Characterization of two heat shock genes from *Haloferax volcanii*: a model system of regulation in the archaea. *J. Bacteriol.* **179**, 6318–6324.

Lagrange, T., Kapanidis, A.N., Tang, H., Reinberg, D., Ebright, R.H. 1998. New core promoter element in RNA polymerase II-dependent transcription: sequence-specific DNA binding by transcription factor IIB. *Genes Dev.* **12**, 34–44.

Langer, D., Hain, J., Thuriaux, P., Zillig, W. 1995. Transcription in archaea: similarity to that in eucarya. *Proc. Natl. Acad. Sci. USA* **92**, 5768–5772.

Langer, D., Zillig, W. 1993. Putative *tfIIs* gene of *Sulfolobus acidocaldarius* encoding an archaeal transcription elongation factor is situated directly downstream of the gene for a small subunit of DNA-dependant RNA polymerase. *Nucl. Acids Res.* **21**, 2251.

Lee, T.I., Young, R.A. 2000. Transcription of eucaryotic protein-coding genes. *Annu. Rev. Genet.* **34**, 77–137.

Lemon, B., Tjian, R. 2000. Orchestrated response: a symphony of transcription factors for gene control. *Genes Dev.* **14**, 2551–2569.

Littlefield, O., Korkhin, Y., Sigler, P.B. 1999. The structural basis for the oriented assembly of a TBP/TFB/promoter complex. *Proc. Natl. Acad. Sci. USA* **96**, 13668–13673.

Lim, C.R. Kimata, Y., Ohdate, H., Kokubo, T., Kikuchi, N., Horigome, T., Kohno, K. 2000. The *Saccharomyces cerevisiae* RuvB-like protein, Tih2p is required for cell cycle progression and RNA polymerase II-directed transcription. *J. Biol. Chem.* **275**, 22409–22417.

Lopez-Garcia, P., Forterre, P. 2000. DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *BioEssays* **22**, 738–746.

Macario, A.J.L., Conway de Macario, E. 1997. Stress genes: an introductory overview. *Stress* **1**, 123–134.

Macario, A.J.L., Conway de Macario, E. 1999. The archaeal molecular chaperone machine: peculiarities and paradoxes. *Genetics* **152**, 1277–1283.

Macario, A.J.L., Conway de Macario, E. 2001. The molecular charperone system and other anti-stress mechanisms in archaea. Frontiers in *Bioscience* **6**, d262–283. http://www.bioscience.org/2001/v6/d/macario/fulltext.htm

Macario, A.J.L., Dugan, C.B., Conway de Macario, E. 1991. A *dnaK* homolog in the archaebacterium *Methanosarcina mazeii* S6. *Gene* **108**, 133–137.

Macario, A.J L., Lange, M., Ahring, B.K., Conway de Macario, E. 1999. Stress genes and proteins in the Archaea. *Microbiol. Mol. Biol. Rev.* **63**, 923–967.

Makino, Y., Kanemaki, M., Kurokawa, Y., Koji, T., Tamura, T.A. 1999. A rat RuvB-like protein, TIP49a, is a germ cell-enriched novel DNA helicase. *J. Biol. Chem.* **274**, 15329–15335.

Magill. C.P., Jackson. S.P., Bell, S.D. 2001. Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB. *J. Biol. Chem.* **276**, 46693–46696.

Marsh, T.L., Reich, C.I., Whitelock, R.B., Olsen, G.J. 1994. Transcription factor IID in the archaea: sequences in the *Thermococcus celer* genome would encode a product closely related to the TATA-binding protein of eukaryotes. *Proc. Natl. Acad. Sci. USA* **91**, 4180–4184.

Matsuda, T. Morikawa M., Haruki, M., Higashibata, H., Imanaka, T., Kanaya, S. 1999. Isolation of TBP-interacting protein (TIP) from a hyperthermophilic archaeon that inhibits the binding of TBP to the TATA-DNA. *FEBS Lett.* **457**, 38–42.

Matsuda, T., Fujikawa, M., Haruki, M., Tang, XF., Ezaki, S. Imanaka, T., Morikawa, M., Kanaya, S. 2001. Interaction of TIP26 from a hyperthermophilic archaeon with TFB/TBP/DNA ternary complex. *Extremophiles* **5**, 177–182.

Maxon, M.E., Goodrich, J.A., Tjian, R. 1994. Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev*. **8**, 515–524.

Maxon, M.E., Tjian, R. 1994. Transcriptional activity of transcription factor IIE is dependent on zinc binding. *Proc. Natl. Acad. Sci. USA* **91**, 9529–9533.

Mayerhofer, L E., Macario, A.J.L., Conway de Macario, E. 1992. Lamina, a novel multicellular form of *Methanosarcina mazei* S-6. *J. Bacteriol*. **174**, 309–314.

Metcalf, W.W., Zhang, K.K., Apolinario, E., Sowers, K.R., Wolfe, R.S. 1997. A genetic system for Archaea of the genus *Methanosarcina*: liposome-mediated transformation and construction of shuttle vectors. *Proc. Natl. Acad. Sci. USA* **94**, 2626–2631.

Miller, J., McLachlan, A.D., Klug, A. 1985. Repetive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J*. **4**, 1609–1614.

Missiakas, D., Raina, S., Georgopoulos, C. 1996. Heat shock regulation, In: *Regulation of Gene Expression in Escherichia coli*, E.C.C. Lin and A.S. Lynch, Eds. Chapman & Hall, New York, NY, pp. 481–501.

Morimoto, R.I. 1998. Regulation of the heat shock transcriptional response: cross-talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes Dev*. **12**, 3788–3796.

Napoli, A., Van der Oost, J., Sensen, C.W., Charlebois, R.L., Rossi, M., Ciaramella, M. 1999. An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfactaricus* which binds to its own promoter. *J. Bacteriol*. **181**, 1474–1480.

Newman, E.B., Lin, R. 1995. Leucine-responsive regulatory protein, a global regulator of gene expression in *E. coli*. *Annu. Rev. Microbiol*. **49**, 747–775.

Newman, D.R., Kuhn, J.F., Shanab, G.M., Maxwell E.S. 2000. Box C/D snoRNA-associated proteins: two pair of evolutionarily ancient proteins and possible links to replication and transcription. *RNA* **6**, 861–879.

Nikolov, D.B., Chen, H., Halay, E.D., Hoffmann, A., Roeder, R.G., Burley, S.K. 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA* **93**, 4862–4867.

Nikolov, D., Hu, S-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua., N-H., Roeder, R.G., Burley, S.K. 1992. Crystal structure of TFIID TATA-box binding protein. *Nature* **360**, 40–46.

Ng, W.V., Ciufo, S.A., Smith, T.M., Bumgarner, R.E., Baskin, D., Faust, J., Hall, B., Loretz, C., Seto, J., Slagel, J., Hood, L., DasSarma, S. 1998. Snapshot of a large dynamic replicon in a halophilic arhcaeon: megaplasmid or minichromosome? *Genome Res*. **8**, 1131–1141.

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Berquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N.S., Thorsson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welti, R., Goo, Y.A., Leithauser, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddocks, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale,H., Isenbarger, T.A., Peck, R.F., Pohlschroder, M., Spudich, J.L., Jung, K-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L., DasSarma, S. 2000. Genome sequence of *Halobacterium* sp. NRC-1. *Proc. Natl. Acad. Sci. USA* **97**, 12176–12181.

O'Brien, R., DeDecker, B., Fleming, K.G., Sigler, P.B., Ladbury, J.E. 1998. The effects of salt on the TATA Binding Protein-DNA interaction from a hyperthemophilic archaeon. *J. Mol. Biol.* **279**, 117–125.

Ogata, Y., Mizushisma, T., Kataoka, K., Kita, K. Miki, T., Sekimizu, K. 1996. DnaK heat shock protein of *Escherichia coli* maintains the negative supercoiling of DNA against thermal stress. *J. Biol. Chem*. **271**, 29407–29414.

Orphanides, G., Reinberg, D. 2002. A unified theory of gene expression. *Cell* **108**, 439–451.

Ouhammouch, M., Geiduschek. E.P. 2001. A thermostable platform for transcriptional regulation: the DNA-binding properties of two Lrp homologs from the hyperthermophilic archaeon *Methanococcus jannaschii*. *EMBO J*. **20**, 146–156.

Ouzounis, C., Sander C. 1992. TFIIB, an evolutionary link between the transcription machineries of archaebacteria and eukaryotes. *Cell* **71**, 189–190.

Pastor, N., Weinstein, H. 1995. Electrostatic analysis of DNA binding properties in lysine to leucine mutants of TATA-box binding proteins. *Prot. Eng*. **8**, 543–549.

Perler, F.B., Davis, E.O., Dean, G.E., Gimble, F.S., Jack, W.E., Neff, N., Noren, C.J., Thorner, J., Belfort., M. 1994. Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucl. Acids Res*. **22**, 1125–1127.

Pfeifer, F., Zotzel, J., Kurenbach, B., Röder, R., Zimmermann, P. 2001. A p-loop motif and two basic regions in the regulatory protein GvpD are important for the repression of gas vesicle formation in the archaeon *Haloferax mediterranei*. *Microbiology* **147**, 63–73.

Pugh, B.F. 2000. Control of gene expression through regulation of the TATA-binding protein. *Gene* **255**, 1–14.

Puhler, G., Lottspiech, F., Zillig, W. 1989. Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius*. *Nucl. Acids Res*. **17**, 4517–4534.

Qureshi, S.A., Bell, S.D., Jackson, S.P. 1997. Factor requirements for transcription in the archaeon *Sulfolobus shibatae*. *EMBO J*. **16**, 2927–2936.

Qureshi, S.A., Jackson, S.P. 1998. Sequence-specific DNA binding by the *S. shibatae* TFIIB homolog, TFB, and its effect on promoter strength. *Mol. Cell* **1**, 389–400.

Qureshi S.A., Khoo, B., Baumann, P., Jackson, S.P. 1995. Molecular cloning of the transcription factor TFIIB homolog from *Sulfolobus shibatae*. *Proc. Natl. Acad. Sci. USA* **92**, 6077–6081.

Rashid, N., Morikawa, M., Imanaka, T. 1995. An abnormally acidic TATA-binding protein from a hyperthermophilic archaeon. *Gene* **166**, 139–143.

Reiter, W-D., Hudelpohl, U., Zillig, W. 1990. Mutational analysis of an archaebacterial promoter; essential role of a TATA box for transcription efficiency and start-site selection *in vitro*. *Proc. Natl. Acad. Sci. USA* **87**, 9509–9513.

Roberts, S.G.S., Greene, M.R. 1994. Activator-induced conformational change in general transcription factor TFIIB. *Nature* **371**, 717–720.

Rowlands, T., Baumann, P, Jackson, S.P. 1994. The TATA-binding protein: a general transcription factor in eukaryotes and archaebacteria. *Science* **264**, 1326–1329.

Ruepp, A., Graml, W., Santos-Martinez, M-L., Koretke, K.K., Volker, C., Mewes, H.W. Frishman, D., Stocker, S., Lupas, A.N., Baumeister, W. 2000. The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513.

Ruepp, A., Rockel, B., Gusche, I., Baumeister, W., Lupas, A.N. 2001. The chaperones of the archaeon *Thermoplasma acidophilum*. *J. Struct. Biol*. **135**, 126–138.

Sandmanm K., Reeve, J.N. 2000. Structure and functional relationships of archaeal and eukaryal histones and nucleosomes. *Arch. Microbiol.* **173**, 165–169.

Scandurra, R., Consalvi, V., Chiaraluce, R., Politi, L., Engel, P.C. 2000. Protein stability in extremophilic archaea. *Frontiers in Bioscience* 5, d787–795. http://www.bioscience.org.v5/d/scan/fulltext.htm

Schmidt, J.E., Ahring, B.K. 1995. Granular sludge formation in upflow anaerobic sludge blanket (UASB) reactors. *Biotechnol. Bioeng.* **49**, 229–246.

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.K, 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., Van der Oost, J. 2001. The complete genome of the crenarchaeon S*ulfolobus solfataricus* P2. *Proc. Natl. Acad. Sci. USA* **98**, 7835–7840.

Shen, X., Mizuguchi, G., Hamiche, A., Wu, C. 2000. A chromatin remodeling complex formed involved in transcription and DNA processing. *Nature* **406**, 541–544.

Shimkets, L.J., Brun, Y.I. 2000. Prokaryotic development: Strategies to enhance survival. In: *Prokaryotic Development* (Brun, Y.I. & Shimkets, L.J., Eds.), pp. 1–7, American Society for Microbiology Press, Washington, DC, USA.

Skorko, R. 1984. Protein phosphorylation in the archaeabacterium *Sulfolobus acidocaldarius*. *Eur. J. Biochem*. **145**, 617–622.

Slesarev, A.I., Mezhevaya, K.V., Makarova, K.S., Polushin, N.N., Shcherbinina, O.V., Shakhova,V.V., Belova, G.I., Aravind, L., Natale, D.A., Rogozin, I.B., Tatusov, R.L., Wolf, Y.I., Stetter, K.O., Malykh, A.G., Koonin, E.V., Kozyavkin, S.A. 2002. The complete genome of hyperthermophile *Methanopyrus kandleri* AV19 and monophyly of archaeal methanogens. *Proc. Natl. Acad. Sci. USA* **99**, 4644–4649.

Smith, D.R., Doucette-Stamm, L.A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., Gilbert, K., Harrison, D., Hoang, L., Keagle, P., Lumm, W., Pothier, B., Qiu, D., Spadafora, R., Vicaire, R., Wang, Y., Wierzbowski, J., Gibson, R., Jiwani, N., Caruso, A., Bush, D., Safer, H, Patwell, D, Prabhakar, S., McDougall, S., Shimer, G., Goyal, A., Pietrokovski, S., Church, G.M., Daniels, C.J., Mao, J., Rice, P., Nölling, J., Reeve, J.N. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum*) ΔH**:** functional analysis and comparative genomics. *J. Bacteriol.* 179, 7135–7155.

Soppa, J. 1999. Transcription initiation in archaea: facts, factors and future aspects. *Mol. Microbiol*. **31**, 1295–1305.

Soerensen, A.H., Torsvik, V.L., Torsvik, T., Poulsen, L.K., Ahring, B.K. 1997. Whole-cell hybridization of *Methanosarcina* cells with two new oligonucleotide probes. *Appl. Environ. Microbiol*. **63**, 3043–3050.

Spencer, V.A., Davie, J.R. 1999. Role of covalent modifications of histones in regulating gene expression. *Gene* **240**, 1–12.

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

Takemaru, K., Harashima, S., Ueda, H., Hirose, S. 1998. Yeast coactivator MBF1

mediates GCN4–dependent transcriptional activation. *Mol. Cell. Biol.* **18**, 4971–4976.

Thompson, D.K., Daniels, C.J. 1998. Heat shock inducibility of an archaeal TATA-like promoter is controlled by adjacent sequence elements. *Mol. Microbiol.* **27**, 541–551.

Thomsen, J., De Biase, A., Kaczanowski, S., Macario, A.J.L., Thomm, M., Zielenkiewicz, P., MacColl, R., Conway de Macario, E. 2001. The basal transcription factors TBP and TFB from the mesophillic archaeaon *Methanosarcina mazeii*: structure and conformational changes upon interaction with stress-gene promoters. *J. Mol. Biol.* **309**, 589–603.

Todone, F., Brick, P., Werner, F., Weinziert, R.O.J., Onesti, S. 2001. Structure of an archaeal homolg of the eukaryotic RNA polymerase II RPB4/RPB7 complex. *Mol. Cell* **8**, 1137–1143.

Um, M., Manley, J.L. 2000. The *Drosophila* TATA binding protein contains a strong but masked activation domain. *Gene Expression* **9**, 123–132.

Werner, F., Eloranta, J.J., Weinzierl, R.O.J. 2000. Archaeal RNA polymerase subunits F and P are bona fide homologs of eukaryotic RBP4 and RBP12. *Nucl. Acids. Res.* **28**, 4299–4305.

Wang, Q. Calvo, J.M. 1993. Lrp, a major regulatory protein in *Escherichia coli*, bends DNA and can organize the assembly of a higher-order nucleoprotein structure. *EMBO J.* **12**, 2495–2501.

Woychik, N.A., Hampsey, M. 2002. The RNA polymerase II machinery: structure illuminates function. *Cell* **108**, 453–463.

Yao, R., Macario, A.J.L., Conway de Macario, E. 1992. Immunochemical differences among *Methanosarcina mazei* S-6 morphologic forms. *J. Bacteriol.* **174**, 4683–4688.

Yang, C-F., Kim, J., Molinari, E., DasSarma, S. 1996. Genetic and topological analysis of the *bop* promoter of *Halobacterium halobium*: stimulation by DNA supercoiling and non-B-DNA structure. *J. Bacteriol.* 178, 840–845.

Yokomori, K., Verrijzer, C.P., Tjian, R. 1998. An interplay between TATA box-binding protein and transcription factors IIE and IIA modulates DNA binding and transcription. *Proc. Natl. Acad. Sci. USA* **95**, 6722–6727.

Zhang, J.L., White, A.K., Kuettner, H.C., Boccazzi, P., Metcalf, W.W. 2002. Directed mutagenesis and plasmid-based complementation in the methanogenic archaeon *Methanosarcina acetivorans* C2A demonstrated by genetic analysis of proline biosynthesis. *J. Bacteriol.* **184**, 1449–1454.

Zhu, W., Zeng, Q., Colangelo, C.M., Lewis, L. M., Summers, M.F., Scott, R.A. 1996. The N-terminal domain of TFIIB from *Pyrococcus furiosus* forms a zinc ribbon. *Nature Struct. Biol.* **3**, 122–124.

Zillig, W., Palm P., Langer, D., Klenk, H-P., Lanzendorfer, M., Hudepohl, U., Hain, J. 1992. RNA-polymerases and transcription in archeabacteria. *Biochem. Soc. Symp.* **58**, 79–88.

Zinder, S.H. 1993. Physiological ecology of methanogens. In: *Methanogenesis* (Ferry, J.G., Ed.), pp. 161–206, Chapman & Hall, New York, NY.