NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Genetic analysis in *Haloferax volcanii*.

by

Wan L. Lam

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
at
Dalhousie University
Halifax, Nova Scotia
January, 1991

ISBN 0-315-64544-X

To Ford

Table of contents

List of Figures

List of Tables

ABSTRACT

Genetic tools have been developed for studying the molecular biology of the halophilic archaebacterium, *Haloferax volcanii*. Shuttle vectors and a transformation system facilitate direct cloning of archaebacterial genes, and allow the reintroduction of *in vitro*-altered DNA into *Hf. volcanii*. The archaebacterial genes isolated have in turn provided the necessary genetic markers for further development of strategies for genetic mapping and for direct manipulation of the chromosome.

Several of these genetic techniques are applied to describe genes for tryptophan biosynthesis and mevalonic acid biosynthesis, to determine the organization of functionally-related genes, and to map auxotrophic mutations to the *Hf. volcanii* chromosome.

Abbreviations

A	adenosine
C	cytidine
G	guanosine
T	thymidine
ade	adenine
ala	alanine
AMV	avian myeloblastosis virus
arg	arginine
aro	aromatic amino acid
asn	asparagine
asp	aspartic acid
Bat	bacterio-opsin activator-protein
Bop	bacterio-opsin
bp	basepair
Brp	bacterio-opsin related-protein
C-terminus	carboxy-terminus
cys	cysteine
DHFR	dihydrofolate reductase
DNase	deoxyribonuclease
DPE	distal promoter element
EDTA	ethylenediaminetetra-acetic acid
EMS	ethyl methanesulfonate
gln	glutamine
glu	glutamic acid
gly	glycine
gua	guanine
his	histidine
HMG CoA	hydroxymethylglutaryl-coenzymeA
ile	isoleucine
kb or kbp	kilobasepair
leu	leucine
lys	lysine
met	methionine
mev	mevinolin
mev^r	mevinolin resistance
MMS	methyl methanesulfonate
mRNA	messenger RNA
N-terminus	amino-terminus
Nif	nitrogen fixation
nt	nucleotide
O.D.	optical density
ORF	open reading frame
PEG	polyethylene glycol
phe	phenylalanine
PPE	proximal promoter element
pro	proline
pur	purine
pyr	pyrimidine
RNase	ribonuclease
rRNA	ribosomal RNA
ser	serine
SR	sensory rhodopsin
thr	threonine
Tris	tris(hydroxymethyl)-methylaminomethane
tRNA	transfer RNA
trp	tryptophan
tyr	tyrosine
U	uridine
ura	uracil
UV	ultraviolet
Vac	gas vacoule protein
val	valine

INTRODUCTION

1. Archaebacteria

Archaebacteria represent a monophyletic group of microorganisms distinct from the eubacteria and the eukaryotes (Woese and Fox 1977; Woese 1981, 1987; Woese et al. 1990). Woese and colleagues first proposed this in 1977 on the basis of phylogenetic analyses of archaebacterial ribosomal RNAs (Woese and Fox 1977). The methanogens, certain sulfur-dependent thermophiles and the extreme halophiles make up this group. These diverse divisions of archaebacteria share common features in their translation apparatus (Gupta 1984; Gehrmann et al. 1986; McCloskey 1986; Woese and Olsen 1986) and in their signature ether-linked archaebacterial lipids (De Rosa and Gambacorta 1986; De Rosa et al. 1986) absent from the eubacteria and the eukaryotes. The archaebacteria, although morphologically resembling the common bacteria, actually represent a third primary lineage of life as distinct from the eubacteria and the eukaryotes as the latter two are from each other.

The finding of eukaryotic features in archaebacteria has raised suspicion of a closer relationship between archaebacteria and eukaryotes. Zillig and colleagues have pointed out the resemblance of archaebacterial DNA-dependent RNA polymerases to their eukaryote homologs, in primary sequence and in complexity (Gropp et al. 1986; Leffers et al. 1989; Huet et al. 1983; Puhler et al. 1989; Zillig et al. 1988, 1989). Futhermore, composite phylogenetic trees rooted by gene pairs duplicated prior to the divergence of the three primary kingdoms (namely, the α and β subunits of H^+ATPase and the elongation factors Tu and G) hint that the archaebacteria are more closely related to eukaryotes than to eubacteria (Iwabe et al. 1989; Gogarten et al. 1989; Sudhof et al. 1989).

Archaebacteria have become adapted to live under extreme conditions, such as low pH, boiling temperatures and saturating salinity, and show us how adaptable life on Earth is. Survival under extreme conditions presumably requires unusual adaptations in

biological features and strategies. Therefore, archaebacteria provide a source for discovering new mechanisms and strategies for fulfilling molecular and cell physiological tasks.

More significantly, inquiries into basic features of archaebacterial gene structure and function should provide information necessary for understanding the molecular biology of the last ancestor of the three primary kingdoms -- features common to diverged lineages should have been present in their last common ancestor, and should consequently reveal the biological sophistication of the progenote, the ancestor to all living things on Earth.

It has been more than a decade since the discovery of archaebacteria; but still relatively little is known about their molecular biology. The lack of genetic methods is clearly limiting the understanding of archaebacteria at the DNA level. I will describe here the limited information available on archaebacterial genetics.

2. Genetics in archaebacteria

Genome structure and organization. Although several archaebacterial genomes have been characterized by physical means, genetic maps are not yet available. Archaebacterial chromosomes are similar to those of eubacteria in sequence complexity (Klein and Schorr 1984; Mitchell et al. 1979; Moore and McCarthy 1969a, b; Charlebois et al. 1989; Searcy and Doyle 1975; Jones et al. 1987). Restriction mapping data show that the chromosomes of *Sulfolobus acidocaldarius* and *Thermococcus celer* are circular, like the typical eubacterial chromosome (Noll 1989; Yamagishi and Oshima 1990). The sizes of archaebacterial genomes range from about 4 million bp in halobacteria, similar to the size of the *E. coli* genome, to 8.4 x 10^8 daltons in *Thermoplasma acidophilum*, about 30% of *E.coli* and similar to *Mycoplasma* species (Charlebois et al. 1989; Moore and McCarthy 1969b; Searcy and Doyle 1975).

Gene organization in archaebacteria resembles that found in eubacterial-like. Functionally related genes are often linked, and in some cases arranged in operons. The typical 16S-23S-5S ribosomal gene order found in eubacteria is also common in archaebacteria, but not all archaebacterial examples studied conform to this arrangement. Many thermophilic archaebacteria have unlinked 5S rRNA genes, and the 16S and 23S rRNA genes in *Thermoplasma acidophilum* are not transcriptionally linked (Tu and Zillig 1982). Transfer RNA genes often exist within a tRNA operon, and are sometimes associated with rRNA genes, including the otherwise independent 5S rRNA gene in two species of *Methanococcus* (Larsen et al. 1986; Wich et al. 1984; Brown et al. 1989). Protein-coding genes, such as the genes for halobacterial flagellins (Gerl and Sumper 1988), ribosomal proteins (Auer et al. 1989a, b; Shimmin et al. 1989a), methyl reductase subunits and RNA polymerase subunits are organized in operons (Bokranz et al. 1988; Cram et al. 1987; Leffers et al. 1989; Puhler et al. 1989). The four genes associated with bacterio-opsin synthesis so far identified are also clustered, but only two of the four genes are tandemly transcribed (Betlach et al. 1989).

Understanding of enzymatic activities relevant to DNA replication, topology, and base-modification has led to intriguing discoveries concerning DNA structure in certain archaebacteria. Kikuchi and Asai (1984) identified in *Sulfolobus* a unique topoisomerase which introduces positive helical turns into DNA (Nakasu and Kikuchi 1985; Forterre et al. 1985). The discovery of positively-supercoiled DNA in a virus-like particle in *Sulfolobus*, SSV1, convincingly shows *in vivo* reverse gyrase function (Nadal, et al. 1986). This reverse gyrase activity is widely distributed in hyperthermophilic archaebacterial species (including *Archaeoglobus fulgidus*, *Methanopyrus kandleri*, and also members of the *Sulfolobales*, *Thermoproteales*, *Thermococcales* and certain *Methanobacteriales*) which grow optimally between 85°C and 105°C, but is not found in mesophilic and moderately thermophilic archaebacteria, such as members of *Thermoplasmales, Methanomicrobiales, Methanococcales* (Slezarev 1988; DeLa Tour et

al. 1990). Topoisomerases from *Halobacterium* species show no reverse-gyrase activity, and plasmids pGRB and pHV2 in *Halobacterium* are negatively supercoiled *in vivo* (Sioud et al. 1988a, b; Forterre et al. 1986, 1989).

DNA base composition varies from a G+C content of about 30 mol% in *Methanococcus voltae* and *Methanobrevibacter arboriphilicus* to nearly 70 mol% in certain *Halobacterium* species (Klein, Schnorr 1984; Moore and McCarthy 1969b). High G+C content apparently is not required for protecting double stranded DNA from denaturation by high growth temperature -- DNA in *Thermoplasma acidophilum* is 46 mol% in G+C (Searcy and Doyle 1975). Stable DNA structure in thermophilic archaebacteria has been linked to high internal salt concentration and histone-like DNA-binding proteins (Green et al. 1983; Stein and Searcy 1978).

Genetic instability in extreme halophiles. The genomes of many halophilic archaebacteria can be fractionated by their G+C content into two components, FI and FII. FI DNA (about 68% G+C) represents the major part of the genome and is highly conserved in purple membrane-forming halobacteria (Ebert and Goebel 1985). FII DNA (58% G+C) in *Halobacterium halobium* accounts for about 30% of the genome and consists of the 150 kbp plasmid pHH1, other heterogeneous covalently closed circular DNAs, present in low copy number, and an A-T rich island in the chromosome (Pfeifer and Betlach 1985). FII DNA harbours a variety of insertion (ISH) elements (Pfeifer et al. 1983; Pfeifer 1986; Hofman et al. 1986; Charlebois and Doolittle 1989). These ISH elements are largely responsible for the instability of the *Halobacterium* genome (Charlebois and Doolittle 1989). Their existence also explains the remarkably high frequency of genomic rearrangements involving many families of repeat sequence elements (Sapienza and Doolittle 1982; Sapienza et al. 1982). An insertion element, ISM1, structurally similar to the ISH elements, exists in *Methanobrevibacter smithii*, suggesting that insertion elements are not restricted to the extreme halophiles (Hamilton and Reeve 1985b). However, abundance and variety of such elements have not been

observed outside the extreme halophiles, nor has the presence of two major fractions of DNA of different base composition. The instability of the halophile genome is reflected in the genetic variability of the bacteriophage ϕH of *Hb. halobium*, which also suffers specific deletions, insertions and ISH element-mediated inversions at high frequencies (Schnabel et al. 1982a).

Repeated sequences also arise spontaneously by duplication (and amplification) of segments of the chromosome. Amplification of DNA segments containing the *Haloferax volcanii* (formerly *Halobacterium volcanii*) dihydrofolate reductase (DHFR) gene leads to overproduction of the enzyme it encodes, and allows cells to overcome growth inhibition by DHFR inhibitors (Rosenshine et al. 1987). Similarly, *Hf. volcanii* exposed to an hydroxy-methylglutaryl coenzyme A (HMG CoA) reductase inhibitor often survives by amplifying the HMG CoA reductase gene (see *Results and Discussion*).

Mutants defective in bacterio-opsin synthesis. The abundant active insertion (ISH) elements in *Hb. halobium* have generated a variety of mutants defective in gas vacuole (Vac) production and bacterio-opsin (Bop) synthesis, and thus permitted genetic dissection of these systems (Betlach et al. 1983; DasSarma et al. 1983, 1986; DasSarma 1989; Pfeifer et al. 1984; Pfeifer 1986). In *Hb. halobium*, spontaneous Bop mutants occur at a frequency of 10^{-4} (Pfeifer et al. 1981). Bop$^-$ mutants are readily detectable due to changes in colony appearance (Stoeckenius and Bogomolni 1982; DasSarma 1989). Insertion events are responsible for the majority of the many Bop$^-$ mutants isolated so far, although one deletion has been found (Betlach et al. 1989; DasSarma 1989). Eight halobacterial insertion elements (ISH1, ISH2, ISH23, ISH24, ISH26, ISH27, ISH28 and ISHS1) were detected to have inserted within or upstream of the *bop* gene in individual mutants (Pfeifer et al. 1984; Pfeifer 1986; Ebert et al. 1987; Simsek et al. 1982). Located upstream from *bop* are 2 other genes (*brp* and *bat*) which are transcribed, in this order, in the opposite direction relative to *bop*, and are associated with *bop* gene expression (Betlach et al. 1986, 1989; Leong 1988a, b; Pfeifer et al. 1984).

Bop mutant IV-4 suffers an ISH24 insertion near the 3'-terminus of the *brp* (bacterio-opsin related protein) gene. This insertion abolishes *bat* and *bop*, as well as *brp*, transcription. The only characterized revertant of IV-4 (Bop$^-$), reIV-41, has acquired a second insertion (via translocation of a 588 bp segment from the *Hb. halobium* plasmid pHH1) adjacent to the ISH24 element, proximal to the *bat* gene. This segment (which, when disrupted by further ISH insertions results in Bop$^-$ phenotype again) contains a copy of archaebacterial promoter consensus sequence, and it is presumably this which restores *bat* transcription. The *bop*, truncated *brp* and *bat* mRNAs in this revertant are present at 100%, 75% and 20%, respectively, of the wild-type levels (Betlach et al. 1984; Leong et al. 1988a, b).

Disruption of *bat* (bacterio-opsin activator) results in the loss of detectable levels of *bop* and *brp* mRNA, suggesting that the *bat* gene product is necessary to activate *bop* and *brp* gene expression. Insertion in the 526 bp region between (and upstream of)*bop* and *brp* drastically reduces *brp* and *bop* mRNA levels as well -- perhaps a trans-acting *bat* gene product could affect transcription initiation by binding to this putative regulatory region (Pfeifer et al. 1985; Betlach, et al. 1989). Since *bat* gene expression is affected by upstream insertions in *brp*, while *brp* requires *bat* activation, the *bat* gene would be autoregulated (Betlach, et al. 1989; Leong et al. 1988a, b). A small ORF oriented opposite to *brp* and down stream of *bat* is also suspected to play a role in bacterio-opsin synthesis.

Genetics of gas-vacuole formation. *Hb. halobium* has two similar gas vacuole protein genes. The p-*vac* (or *gvpA*) gene, which is constitutively expressed, is mutated at a remarkable frequency of 10^{-2}. In contrast, *c-vac* (or *gvpB*), which is only expressed during stationary growth, exhibits a much lower mutation frequency of 10^{-5}. The drastic difference in mutation frequency can be explained by the locations of these genes. *c-vac* is situated in the more stable FI DNA of the chromosome, while *p-vac* is found in the ISH element-rich plasmid pHH1 (or pNRC100) which belongs to the variable FII fraction of

the genome (DasSarma et al 1987; Horne et al. 1988). In addition to insertional inactivation by ISH elements (as experienced by the Bop mutants described above), the *p-vac* gene can be lost by a deletion event encompassing the *p-vac* gene region of pHH1 (DasSarma et al. 1988; DasSarma 1989; Pfeifer et al. 1988, 1989; Pfeifer and Blaseio 1989).

Deletion events within the 150 kbp plasmid pHH1 occur at high frequencies during the development of a *Hb. halobium* culture. Certain Vac$^-$ derivatives of pHH1 have acquired more than one deletion, leading to the fusion of non-contiguous pHH1 sequences to give smaller plasmids. Pfeifer and Blaseio (1989) have linked plasmid variability to the action of insertion elements. Each deletion occured exactly at the terminus of an insertion element. Most likely, the deletion variants result from recombination between repeated sequences, and from intramolecular transposition, which breaks a plasmid into two circular DNA molecules of which only one retains the sequences for plasmid maintainance and replication.

Restriction and modification systems. Restriction endonucleases have been found in *Halobacterium, Thermoplasma, Methanococcus* and *Methanobacterium* (McConnell et al. 1978; Prangishvilli et al. 1985; Schmid et al. 1984; Lunnen et al. 1989). Evidence of restriction suggests the presence of base modifications. Modified bases common in eubacteria, like N^4-methylcytosine, 5-methylcytosine and N^6-methyladenine, have been found in archaebacteria (Ehrlich et al. 1985). A virus of *Hb. halobium*, ϕN is, in fact, fully methylated at all cytosine residues in its entire genome (Vogelsang-Wenke and Oesterhelt 1988). Recently, Juez et al. (1990) detected modifications in *Haloferax mediterranei* DNA associated with changes in the salt concentration of the growth medium.

Phages of the *Halobacteria* provided *in vivo* evidence of restriction and modification (Daniels and Wais 1984; Patterson and Pauling 1985; Schnabel et al. 1982b). Phage ϕH forms plaques on *Hb. halobium* strain R1, which carries a deleted version of plasmid

pHH1, and is deficient in vacuole formation. Plating efficiency is reduced by a factor of 10^3-10^4 if wild-type *Hb. halobium* is used as host, while ϕH grown on the wild-type strain infects both R1 and wild-type hosts efficiently, suggesting the existence of a restriction-modification system, perhaps determined by plasmid genes (Schnabel et al. 1982b).

Patterson and Pauling (1985) were able to isolate spontaneus and uv-mutagenized mutants of *Halobacterium cutirubrum* deficient in restriction, in modification or in both processes. Correlation between changes in restriction-modification and in gas vacuolation phenotypes were again observed. They also showed that more than one restriction-modification system exists in *Hb. cutirubrum.* Multiple restriction systems have been found also in *Methanococcus aeolicus* -- 3 restriction enzymes were isolated from this organism (Schmid et al. 1984).

The genes for a type II restriction-modification system in *Methanobacterium wolfei* have been cloned and expressed in *E. coli* (Lunnen et al. 1989). Restriction-modification systems appear to be a wide-spread phenomenon in archaebacteria, as in eubacteria.

Archaebacterial virus-host systems. Viruses are exposed to the same extreme habitats as their hosts, and have adapted accordingly. Despite difficult culture conditions in the laboratory, a variety of DNA viruses have been isolated from *Halobacterium*, *Thermoproteus*, *Sulfolobus*, *Methanobacterium.* and *Methanobrevibacter* (Torsvik and Dundas 1974, 1980; Wais et al. 1975; Pauling 1982; Schnabel et al. 1982b; Janekovic et al. 1983; Rohrmann et al. 1983; Martin et al. 1984; Meile et al. 1989; Jordan et al. 1989). Zillig and colleagues have put together a thorough review of the biology of these viruses (Reiter et al. 1988c).

Archaebacterial viruses have served as model systems for the study of genetic variability, host strain restriction-modification, gene expression and DNA uptake and transformation in archaebacteria. Considerable effort has been directed towards

understanding gene expression in the phage ϕH of *Hb. halobium* (Schnabel et al. 1982b) and viral particle SSV1 of *Sulfolobus* B12 (Martin et al. 1984).

Phage ϕH has a linear double-stranded DNA genome which exhibits genetic variability involving specific deletions, insertions or inversions, explainable by the action of insertion elements (Schnabel et al. 1982a). ISH 1.8 elements have been found on either side of a 12 kbp region (called the L segment) which contains the genes expressed in early lytic development. When two copies of ISH 1.8 flank the L-segment in opposite orientations, frequent inversion events are observed. Excision and circularization of this L-segment forms a self-replicating plasmid (pϕHL) which confers immunity to ϕH infection, except that phages carrying an insertion of ISH23 (or ISH50) within the L-region can overcome immunity (Pfeifer et al. 1984; Schnabel 1984; Schnabel and Zillig 1984; Schnabel et al. 1984; Xu and Doolittle 1983).

Transcripts expressed during the early, middle and late stages of lytic growth have been identified, and certain transcripts are not detected in lysogens (Gropp et al. 1989). The specific function, regulation and interaction of these transcripts are yet unknown; nevertheless, ϕH genes contribute to the growing list of sequences from which consensus signals for transcription and translation are deduced. Transcripts identified in ϕH infection account for nearly half of all the characterized protein-coding RNA transcripts in the halophilic archaebacteria.

SSV1 provided Zillig and co-workers (Reiter et al. 1989; Reiter and Palm 1990) information on viral DNA integration, and added more examples of putative transcription signals (Reiter et al. 1988a, b). SSV1 has a 15.5 kbp double-stranded circular DNA genome, which specifically integrates into a site within a $tRNA^{Arg}$ gene on the chromosome, and it is inducible by ultraviolet (uv) radiation. SSV1 integrates via recombination between a 44 bp sequence present in the SSV1 DNA and in a chromosomal $tRNA^{Arg}$ gene, so that the integrated SSV1 DNA is flanked by direct repeats of the 44 bp core attachment sequence (Reiter et al. 1989). A defective SSV1 genome has also been

found integrated in the same manner into a second putative $tRNA^{Arg}$ gene containing a segment nearly identical to the 44 bp core attachment sequence (Reiter and Palm 1990).

The transcription pattern of the small SSV1 genome is simple; at least eight of the transcripts are constitutively expressed. T_{ind} is the only inducible RNA, transcribed in response to uv irradiation (Reiter et al. 1987a, b). All transcript start sites, except for that of T_{ind}, are preceded by a short A-T rich sequence which resembles the consensus putative archaebacterial promoter (Reiter et al. 1988b; also see below). 3'-termini of several transcripts have been located to a pyrimidine rich region following a run of T residues (Reiter et al. 1988a).

Natural mating system. The natural genetic transfer system in the extreme halophile *Hf. volcanii* is the only archaebacterial mating system known. Mevarech and Werczberger (1985) produced prototrophs by cofiltering any two of three auxotrophic mutants -- demonstrating that DNA transfer is not unidirectional, unlike classical eubacterial conjugation. The transfer is an active process, inhibited when one of the two auxotrophic partners is heat killed prior to mating.

Mevarech and colleagues (Rosenshine et al. 1989) followed the process of genetic transfer using immobile plasmids pHV2 and pHV11 as cytoplasmic markers, and found that the cytoplasms of the parental types do not mix during the mating process -- unlike in the fusion of eukaryotic gametes. Chromosomal DNA, but not the cytoplasmic markers, passes through intercellular cytoplasmic bridges which connect *Hf. volcanii* cells into a network. Protoplast fusion could only be induced artificially.

Genetic transformation. Transformation has recently been demonstrated in several species of archaebacteria. Low frequency transformation with exogenous DNA has been reported in two methanogens (Bertani and Baresi 1987; Worrell *et al.* 1988), although these methods have yet to find extensive use. Transformation of halophiles is more widely practised, since the polyethyleneglycol mediated-transformation protocol is efficient and has been well characterized (Cline and Doolittle 1987; Charlebois et al. 1987;

Cline et al. 1989a, b; Lam and Doolittle 1989). To date, no methods for transforming thermophilic archaebacteria have been developed.

Two species of methanogens can be transformed naturally by allowing recipient cells to grow in the presence of marker DNA. Bertani and Baresi (1987) have devised a method patterned after the calcium treatment and heat shock protocol commonly used in the transformation of *E. coli*. Purine or histidine requiring-mutants of *Methanococcus voltae* were transformed to prototrophy at a slightly higher frequency than the reversion rate. Later they found that incubating the same auxotrophs in liquid culture with wild-type DNA alone yielded similar or better results.

Suspecting that physiological factors might affect transformation, Worrell et al. (1988) allowed recipient *Methanobacterium thermoautotrophicum* cells to grow into colonies on non-selective gellan gum plates containing marker DNA, before replicating onto selective medium. They were able to obtain fourteen 5-fluorouracil resistant transformants out of approximately one thousand colonies tested.

Gernhardt and colleagues (Gernhardt et al. 1990) have very recently applied the protocol of Bertani and Baresi to transform *M. voltae* with an *E. coli* vector-construct carrying (i) the puromycin transacetylase gene of *Streptomyces alboniger* fused to the promoter and terminator region of the *M. voltae* methyl reductase operon and (ii) the *M. voltae hisA* gene. Puromycin resistant transformants recovered typically carry the vector-construct integrated into their chromosome at the *hisA* locus. The integrated sequences are stably maintained in the presence of puromycin, but slowly lost in the absence of selective pressure.

Handicapped by the lack of available genetic markers, Cline and Doolittle (1987) relied on plaque assays to score for sucessful transformation events -- using DNA from the lytic phage ϕH to transfect spheroplasts of its natural host, *Hb. halobium*. Transformation is dependent on the physiological state of the cells (early- to mid-log-phase cultures yield higher transformation frequency), inhibited by divalent cations (spheroplasts

are formed by adding sufficient EDTA to chelate free calcium and magnesium ions) and sensitive to high DNA concentrations (DNA precipitates in the presence of PEG and NaCl). The procedure is reliable and efficient, yielding up to 10^7 plaques per microgram of phage DNA. ϕH DNA can also transfect (at reduced frequency, presumably because of restriction) *Hf. volcanii*, which is not a natural host for the phage. This transformation procedure has recently been modified and applied to several species of halophiles (Charlebois et al. 1987; Cline et al. 1989a; Holmes and Dyall-Smith 1990; Lam and Doolittle 1989; S. Cline, personal communication).

In our laboratory, we have sucessfully transformed *Hf. volcanii* with phage, plasmid and chromosomal DNA using a modified version of the transfection protocol (Cline et al. 1989a, b; Charlebois et al. 1987; Lam and Doolittle 1989). Plasmid transformation was first demonstrated using a naturally-occurring *Hf. volcanii* plasmid (pHV2) and a recipient strain previously cured of pHV2 by treatment with ethidium bromide (Charlebois et al. 1987). This plasmid becomes the logical starting material for construction of shuttle vectors maintainable in either halobacteria or *E. coli*, because it transforms *Hf. volcanii* efficiently, and because it has been completely sequenced (Lam and Doolittle 1989; also see *Results and Discussion* below).

Keith Conover of our laboratory has shown that cosmid DNAs containing wild-type *Hf. volcanii* inserts and prepared in *E. coli* can transform halobacterial auxotrophs to prototrophy (Conover and Doolittle 1990). Cosmid DNA cannot replicate autonomously in *Hf. volcanii*, and therefore requires integration into the chromosome for transformation, presumably via homologous recombination. Transformation with cosmid clones (generated and assembled in preparation of a restriction map of the *Hf. volcanii* chromosome) provides a simple strategy for linkage analysis (Charlebois et al. 1989; Conover and Doolittle 1990; Lam et al. 1990b; also see *Results and Discussion*).

Plasmids and potential vectors. Extrachromosomal DNA is common in archaebacteria. Cryptic plasmids have been reported in many archaebacteria (see recent

listing by Brown et al. 1989). Of the many plasmids identified, only a few are associated with obvious phenotypes. Anaerobic autotrophy in *Sulfolobus ambivalens*, and gas vacuole formation and phage immunity in *Hb. halobium* are plasmid-related traits (Zillig et al. 1985; Simon 1978; Weidinder et al. 1979; Schnabel 1984).

Plasmids are potential vectors for DNA transfer. Several low-molecular-weight plasmids from halobacteria are potentially exploitable for vector development, for example, the pϕHL formed by circularization of the L-segment from phage ϕH (Schnabel 1984), the mini-plasmid found in *Halobacterium* strain SB3 (Hackett and DasSarma 1989; Kagramanova et al. 1989), and plasmids isolated from *Hf. volcanii* (Charlebois et al. 1987; Rosenshine and Mevarech 1989). Having an efficient plasmid transformation system in *Hf. volcanii*, it is not surprising that shuttle vectors were first developed for this organism (Charlebois et al. 1987; Lam and Doolittle 1989).

The development and use of shuttle vectors in *Hf. volcanii* is described later in *Results and Discussion.* Some of these vectors have been widely distributed, and in some cases modified by other laboratories for specific purposes. The shuttle vector pWL102, a pHV2-based mevinolin resistance vector, can transform and be maintained in other halophiles, such as *Haloferax* phenon K and *Haloarcula hispanica* (Holmes and Dyall-Smith, personal communication; S.W. Cline, personal communication), but by itself is unable to replicate in *Hb. halobium.* To study features specific for *Hb. halobium,* Blaseio and Pfeifer (1990) have constructed a shuttle vector, maintainable in *Hb. halobium*, *Hf. volcanii* and in *E. coli*, by inserting the replication region of the *Hb. halobium* plasmid pHH1 into a pUC-type plasmid carrying the *Hf. volcanii* mevinolin resistance marker from pWL102. Likewise, to study tRNA-intron splicing *in vivo*, Nieuwlandt and Daniels (1990) have inserted the promoter region of a *Hf. volcanii* $tRNA^{Lys}$ gene into pWL102 to form a tRNA expression vector which allows transcription of tRNA gene clones inserted into sites downstream of this promoter region.

Holmes and Dyall-Smith (1990) have cloned a DNA fragment which confers resistance to novobiocin into a plasmid (pHK2) of *Haloferax* Phenon K, to form a selectable vector which can be maintained in both *Haloferax* Phenon K and its close relative *Hf. volcanii* (Dyall-Smith, personal communication). A shuttle vector replicatable in *E. coli* also has since been constructed. Further analysis of the novobiocin resistance marker reveals that it is in fact a mutant DNA gyrase B (*gyrB*)gene (Holmes and Dyall-Smith 1991).

Complementation of *E. coli* auxotrophs. Several biosynthetic genes (*hisA, hisI, argG, proC, trpB, trpA* and *purE*) have been isolated by complemention of *E. coli* auxotrophs with randomly cloned methanogen DNA (Wood et al. 1983; Cue et al. 1985; Hamilton and Reeve 1985a, b; Beckler and Reeve 1986; Weil et al. 1987; Morris and Reeve 1988; Sibold and Henriquet 1988). Similarities between the products encoded by the complementing genes and their *E. coli* counterparts vary. Complementation is possible because of the occasional incidence of fortuitous eubacterial promoters in the very A+T-rich intergenic spaces of some methanogenic strains (Wood et al. 1983; Bollschweiler et al. 1985; Cue et al. 1985). Perhaps any suboptimal expression of the methanogen genes and the potentially reduced activity of the gene products are compensated by the high copy number imposed by the *E. coli* cloning vectors.

3. Gene structure and function

Our knowledge of archaebacterial gene structure and function is, so far, largely based on comparison of archaebacterial sequences with their eukaryotic and eubacterial homologs. A small number of genes has been isolated by screening libraries of archaebacterial DNA cloned into *E. coli*, using immunological probes (antibodies raised against purified archaebacterial proteins), degenerate oligonucleotide probes (designed using peptide-sequence information of abundant proteins), heterologous gene probes (related eubacterial and eukaryotic genes) and occassionally by complementing *E. coli*

auxotrophs. Signals for gene expression are tentatively identified by looking for consensus sequences or recurring structural motifs flanking the termini of RNA transcripts. As more archaebacterial genes and their transcripts are analysed, the prediction of transcription and translation signals becomes more believable.

Gene structure. The universal genetic code is used in archaebacterial protein-coding genes, and codon usage is generally consistant with the G+C content of organisms. Ribosome-binding sequences (Shine-Dalgarno sequences which complement the 3'-terminal sequence of the 16S rRNA) are found in front of some, but not all, protein-coding genes characterized (Brown et al. 1989; Zillig et al. 1988). In some cases, transcription begins so close to the translation initiation site that an upstream Shine-Dalgarno sequence is impossible (DasSarma et al. 1984). In these cases, Shine-Dalgarno sequences positioned in a hairpin loop structure within the coding region are suspected to facilitate ribosome binding (Betlach et al. 1986; Brown et al. 1989). In other cases, transcription begins sufficiently far upstream that the mRNA contains an untranslated leader -- a 74 nucleotide leader RNA at the beginning of the message for the *H. cutirubrum* L1-L10-L12 ribosomal proteins has been suspected to serve a regulatory function (Shimmin and Dennis 1989). A 111 nucleotide long untranslated leader RNA precedes the sequence which encodes the *H. halobium* surface glycoprotein and its leader peptide (Lechner and Sumper 1987; Sumper et al. 1990).

Introns. Many archaebacterial tRNA genes analysed contain introns in the vicinity of their anticodons, but none of the protein-coding genes characterized so far appears to have an intron (Kaine et al. 1983; Kaine 1987; Kjems et al. 1989; Wich et al. 1987; Daniels et al. 1985; Datta et al. 1989; Zillig et al. 1988). Introns are also found in the 23S rRNA gene of certain thermophilic archaebacteria. The 23S rRNA gene of *Desulfurococcus mobilis* contains a 622 bp intron, which encodes a 194 amino acid protein, within the highly conserved domain IV of the rRNA secondary structure model (Kjems and Garrett

1985). *Staphylothermus marinus* has two smaller introns: a 56 bp intron in domain VI and a 54 bp intron in domain V of the 23S rRNA gene (Kjems and Garrett 1991).

Promoter structure. The typical eubacterial RNA polymerase recognizes promoter sequence elements located 10 bp and 35 bp upstream of the transcript start site (Reznikoff et al. 1985). Transcription in eukaryotes, on the other hand, depends on 3 different types of RNA polymerases which do not generally share promoter recognition sequences. In the RNA polymerase II promoters, the consensus sequences "TATA$^{T}/_{A}$A$^{T}/_{A}$" exists 25 bp upstream of the transcription initiation site (Bucher and Trifonov 1986). Both the stable RNA genes and the protein coding genes of archaebacteria are transcribed by the same RNA polymerase, which resembles eukaryotic polymerases in sequence and in complexity (Huet et al. 1983, Puhler et al. 1989, Zillig et al. 1988).

A number of transcription initiation sites have been identified by primer extension and S1 nuclease protection experiments (see compilation by Zillig et al. 1988; Gropp et al. 1989). Transcription of messenger RNA typically begins very close to the translation start site (AUG). *In vitro* capping experiments have defined the initiation of *bop* (bacterio-opsin) mRNA transcription to two bp upstream of the AUG (DasSarma et al. 1984). Similarily, the two transcripts encoding SSV1 viral proteins VP1, VP2 and VP3 are initiated from the same site eight bp upstream of the AUG (DasSarma et al. 1984; Reiter et al. 1988b).

Extensive comparison of the sequences upstream from the coding region of a variety of archaebacterial genes, including rRNA genes, tRNA genes and protein-coding genes (many of which are viral protein genes), revealed the presence of a short stretch of A-T rich sequences located about 25 bp upstream of the transcript start (Brown et al. 1989; Zillig et al. 1988; Gropp et al. 1989; Thomm and Wich 1988; Reiter et al. 1988b), very much like the eukaryotic pol II "TATA" promoter element, which is also located about 25 bp prior to the transcript start. The consensus sequence (commonly refered to as box A sequence) is believed to constititute the typical archaebacterial promoter. A second

element, weakly conserved in sequence (often TGC or TGA) but always located at or close to the site of transcription initiation, has been termed box B. Transcription often begins at the G residue in box B or at a nearby purine nucleotide (Zillig et al. 1988; Gropp et al. 1989).

Futhermore, RNA polymerase purified from *Methanococcus vannielii* binds and protects, in footprinting experiments, a specific region (upstream of an rRNA gene, a tRNA gene, the methyl reductase operon and the *hisA* gene) which extends from about 30 bp upstream to about 20 bp downstream relative to the transcript initiation site -- this region includes consensus box A and box B sequences (Thomm and Wich 1988; Brown et al. 1988; Thomm et al. 1988; Thomm et al. 1989).

Genetic evidence that box A sequences are part of the archaebacterial promoter now exists. A single base subsitution generated an "up-promoter" mutant which overexpresses the gene for HMG CoA reductase in *Hf. volcanii* (see *Results and Discussion*). The mutant promoter sequence corresponds, in sequence and in location, to the box A consensus sequence of highly expressed halobacterial genes, and to the distal promoter element (DPE) defined by *in vitro* transcription analysis of engineered sequences upstream of the *Sulfolobus B12* rRNA operon (Reiter et al. 1990).

Reiter and colleagues have extensively mutagenized the sequences upstream of the *Sulfolobus* rRNA operon by insertions, deletions and linker substitutions (Reiter et al. 1990). They used a newly developed *Sulfolobus in vitro* transcription system (Hudepohl et al. 1990) to monitor changes in transcription initiation and efficiency. Linker substitution (including minor dinucleotide substititions) in the DPE region (which encompasses box A) drastically reduces transcription efficiency *in vitro*. Alteration of the A-T composition of a region between -11 and -2 (which Reiter et al. [1990] called the proximal promoter element, or PPE) also affects transcription efficiency. The sequence TGC (box B) is not required for transcription efficiency but appears to be involved in

precise start site selection -- when a region containing box B is altered, transcription initiates at multiple sites.

Processing. Considerable effort has been directed towards understanding the processing of polycistronic stable RNA transcripts, but relatively little is known about processing and post-transcriptional modification of peptide-coding RNA transcripts, although polyadenylated RNAs (possibly messenger RNA) have been isolated from *Methanococcus vannielii* and *Hb. halobium* (Brown and Reeve 1985, 1986).

Archaebacterial rRNA and tRNA genes are often transcribed as a polycistronic transcript which is further processed into individual RNAs. The 16S and 23S rRNA are flanked by inverted repeats which could form double stranded stems of varying lengths containing the two small loops, separated on opposite strands by 3 to 4 bp, very much like the eubacterial RNaseIII recognition sites (Dennis 1986; Mankin et al. 1984; Hui and Dennis 1985; Chant and Dennis 1986; Jarsch and Bock 1985; Kjems and Garrett 1987; Kjems et al. 1987). This feature is absent only from the 16S RNA transcript of *T. acidophilum*, which is made independently of the 23S rRNA (Ree et al. 1989).

Processing of archaebacterial tRNA transcripts is virtually unexamined, except that ribonuclease P, the endonuclease which processes the 5'-end of a tRNA from its precursor, has been isolated from *Sulfolobus solfataricus* and from *Hf. volcanii* (Darr et al. 1990; Lawrence et al. 1987). Mature archaebacterial tRNAs contain many modified nucleotides and a terminal CCA which is not encoded by the gene.

The splicing mechanism of the *Desulfurococcus mobilis* 23S rRNA gene intron, at first glance, resembles those of the class III eukaryotic tRNA introns. Sequences flanking the rRNA intron splice site can be folded into a structure which resembles part of a tRNA, such that the intron is located in an anticodon-like loop in the rRNA (Kjems and Garrett 1988; Kjems et al. 1989). However such a tRNA-like structure is not found near the exon junctions of the two introns in the *Staphylothermus marinus* 23S rRNA gene (Kjems et al. 1989; Kjems and Garrett 1991).

The intron is precisely excised from the pre-23S RNA giving a 3'-terminal phosphate, and subsequently ligated to form a stable circular RNA *in vivo* (Kjems and Garrett 1988). *In vitro* precursor transcripts are efficiently cleaved, but poorly ligated (Kjems and Garrett 1988, 1991). Surprisingly, cell extracts from a variety of archaebacteria species, which lack introns in their 23S rRNA gene, contain the cleavage enzyme (Kjems and Garrett 1988, 1991).

A structural motif may be common to all archaebacterial intron cleavage sites. Comparison of the intron-exon boundaries of the *D. mobilis* and *S. marinus* 23S rRNA precursors with over ten other intron-containing tRNA precursors from *Desulfurococcus, Sulfolobus, Thermoproteus* and *Haloferax* (Kaine et al. 1983; Kaine 1987; Wich et al. 1987; Daniels et al. 1985; Kjems et al. 1989; Datta et al. 1989; Kjems and Garrett 1991) revealed that, in all cases, the cleavage sites are located within two three-nucleotide loops on opposite strands separated by four basepairs, reminiscent of the cleavage sites in the processing stem of the rRNA precursor (Thompson and Daniels 1988; Thompson et al. 1989; Kjems et al. 1989; Kjems and Garrett 1990). A similar structure is also observed at the splice site of the *Thermoproteus tenax* $tRNA^{iMet}$ precursor, however, the sites are separated by three nucleotides, and the 5' cleavage site lies within a four nucleotide loop (Wich et al. 1987; Thompson et al. 1989).

This staggered loops structure is, in fact, recognized by an intron endonuclease in *Hf. volcanii*. Thompson and coworkers (Thompson and Daniels 1988; Thompson et al. 1989) isolated a *Hf. volcanii* enzyme preparation capable of excising the intron from the halobacterial $tRNA^{Trp}$ precursor, as well as accurately cleaving an *in vitro*-generated $tRNA^{Trp}$ precursor which lacks most of the intervening sequences and retains little more than the putative cleavage structure (opposing three nucleotide loops separated by four basepairs). The cleavage products have 5'-hydroxyl and 2',3' cyclic phosphate termini, the same as the those produced by *D. mobilis* and *S. marinus* rRNA intron cleavage, except that the 3'- phosphate is not cyclic (Kjems and Garrett 1988, 1991).

These similarities in splicing sites and mechanisms raise the possiblity that the same enzymes splice the rRNA and tRNA introns in the archaebacteria. This would explain why crude cell extracts from archaebacteria lacking rRNA introns effectively carry out the cleavage of, and in the case of *Desulfurococcus mucosus* the splicing of, the *D. mobilis* rRNA intron (Kjems and Garrett 1988, 1991).

Termination signals. It is difficult to decide what an archaebacterial terminator looks like. Comparison of sequences and putative structures near the transcription termination sites of a number of stable RNA genes and protein-coding genes fails to reveal a common termination signal. Termination often occurs over a number of consecutive or scattered positions, as illustrated by the 3'-end heterogeneity in most transcripts defined by nuclease protection experiments (Auer et al. 1989b; Leffers et al. 1989; Shimmin and Dennis 1989; Kjems et al. 1987; Kjems and Garrett 1987; Muller et al. 1985; Reiter et al. 1988a). Transcription of *Hb. cutirubrum* L11 equivalent ribosomal protein gene stops at eight separate positions within 100 bp downstream of the stop codon, while the L12 equivalent ribosomal protein gene terminates at essentially a unique site (Shimmin and Dennis 1989).

Termination often, but not always, occurs within or near a segment of T-rich sequences (Kjems et al. 1987; Reiter et al. 1988a). As expected, A-T rich sequences are found close to the end of most, if not all, of the genes in methanogens which generally have DNA composition low in G+C content (Bollschweiler et al. 1985). The *Methanococcus vannielii* rRNA operon and several tRNA transcriptional units terminate at the beginning of the sequence TTTTTAATTTT which bears resemblance to eukaryotic RNA polymerase III termination signals (Wich et al. 1986a, b; Cozzarelli et al. 1983). In addition, several genes for *Sulfolobus* SSV1 viral proteins terminate in pyrimidine rich sequences preceded by a stretch of T's (Reiter et al. 1988a). Similarly, the ribosomal protein L12 mRNA in *Hb. cutirubrum* ends at a short stretch of T's which is preceded by a G+C rich region which could not form secondary structure elements (Shimmin and Dennis 1989).

Inverted repeats, often found near termination sites, are sometimes imperfect, vary in length, and often followed by T-rich sequences. The *bop* gene transcript in *Hb. halobium* terminates over several positions immediately after a nearly perfect stem and loop structure, independent of T-rich sequences (DasSarma et al. 1984). Similarly, each of the multiple termination sites of the unlinked 5S rRNA operon in *Thermococcus celer* is located downstream of a stem and loop structure without neighbouring T-rich sequences (Culham and Nazar 1989). On the other hand, examples of genes which possess structural elements containing consecutive T residues following inverted repeats, similar to eubacterial rho-independent terminators (Rosenberg and Court 1979), are frequently spotted -- for example, the methyl reductase genes (for methanogenesis) in several methanogens and the unlinked ribosomal genes of *Thermoplasma acidophilum* (Bokranz et al. 1988; Cram et al. 1987; Sanangelantoni et al. 1990; Takao et al. 1989; Auer et al. 1989b; Ree and Zimmermann 1990). However, transcription through such a plausible terminator behind the *Hb. cutirubrum* rRNA operon into the downstream $tRNA^{Cys}$ gene has been reported (Chant and Dennis 1986; Dennis 1986).

Furthermore, the lack of obvious structural motifs downstream of certain archaebacterial genes challenges the presumed function of all these putative structures as termination signals (Conover and Doolittle 1990; Hui and Dennis 1985; Kjems and Garrett 1987). Efficiencies of these putative structures in transcription termination have not been directly compared, although read-through transcription has been observed (Leffers et al. 1989; Chant and Dennis 1986; Culham and Nazar 1989). Further understanding of transcription termination (and other transcriptional processes) in archaebacteria requires a vector transformation system and an *in vitro* transcription system to allow *in vivo* and *in vitro* assessment of engineered transcription signals.

Gene expression and regulation. Many archaebacterial genes are inducible or regulated in response to external stimuli. For example, the expression of the *Hb. cutirubrum* superoxide dismutase gene is responsive to paraquat treatment, which

generates oxygen radicals in respiring cells (May and Dennis 1989; May et al. 1989). Similarly, the pattern of the *hisA* transcription changes in *M.voltae* cells exposed to aminotriazole, a chemical which derepresses histidine biosynthesis in *Salmonella* and in yeast (Sment and Konisky 1986). Purine analog-resistant mutants have recently been isolated from several species of methanogens (Bowen and Whitman 1987; Knox and Harris 1988; Worrell et al. 1988; Nagle 1989; Worrell and Nagle 1990). These mutants should prove useful in studying the regulation of nucleotide biosynthesis and in developing pathway genetics.

The purple membrane in *Hb. halobium* and related species is produced constitutively at a low level under aerobic growth conditions. When oxygen tension is reduced and intense light is available, the cells grow photoheterotrophically, and synthesis is increased by 5 fold, covering 50% of cell surface with purple membrane (Oesterhelt and Stoeckenius 1973). Bacterio-opsin (bop) is the protein component of purple membrane. The mechanism by which environmental signals regulate *bop* expression is unclear. However, we do know that *bop* expression depends on two other genes which encode the bactero-opsin related protein (*brp*) and the bacterio-opsin activator protein (*bat*). The *bat* gene product may be a trans-acting factor which binds to the putative regulatory sequence upstream of *bop* and *brp* -- insertion of ISH elements into the *bat* gene or into the spacer between the oppositely transcribed *brp* gene and *bop* gene results in Bop^- phenotype (Betlach et al. 1984, 1986, 1989; Leong et al. 1988a, b).

Autogenous translational regulation of the tri-cistronic L1-L10-L12 mRNA in *Hb. cutirubrum* has been suggested, though no real experimental evidence has been provided (Shimmin and Dennis 1989). The 74 nucleotide long untranslated leader of the L1-L10-L12 mRNA contains a structure almost identical to part of the presumed L1-binding domain of the 23S rRNA. In *E. coli*, excess L1 protein not incooperated into ribosomes can bind to a L1-binding site-like structure in the 5' leader of the L11-L1 mRNA to stop its translation (Thomas and Nomura 1987). At this point, it is premature to guess how

widely this putative regulatory mechanism is distributed among archaebacteria, although translational control has been suggested also for the L1-L1-L12 operon of *M. vannielii* (Baier et al. 1990). Of the small number of archaebacterial ribosomal protein operons sequenced, analyses have been focused on gene organization and sequence similarity with eubacterial and eukaryotic homologs (Shimmin et al. 1989b; Auer et al. 1989a 1989b; Leffers et al. 1989; Spiridonova et al. 1989; Kopke and Wittmann-Liebold 1989; Ramirez et al. 1989).

Regulatory regions upstream of the rRNA operon of *Sulfolobus* sp. B12 have been experimentally identified. Using a recently developed *Sulfolobus in vitro* transcription assay, Reiter and colleagues thoroughly analysed the sequences upstream of the 16S rRNA gene by linker scanning experiments and deletion analysis (Reiter et al. 1990). They observed a weak positive regulatory region between positions -354 and -190, upstream of the distal and proximal promoter elements described above, and more significantly a negative regulatory sequence between position -93 to -38. Deletion of this negative regulatory region leads to an increase of up to 4 fold in transcription efficiency.

Other than these specific examples, and the guesses at regulatory signals prompted by scattered sequence homologies in front of some archaebacterial genes (Daniels et al. 1986; May and Dennis 1989; Shimmin et al. 1989a), gene regulation in archaebacteria is largely unexplored. However, the structural genes for several potentially regulated systems (such as biosynthetic pathways, methanogenesis and nitrogen fixation in the methanogens and phototaxis in certain halophiles) have been isolated, and should provide useful starting points for investigation.

H. halobium has three flagellin glycoproteins which are encoded by a multigene family of five structural genes, arranged in clusters of two (*fla A1 and A2*) and three (*fla B1, B2* and *B3*). The amino acid sequences deduced from these 5 genes are very similar, but exhibit variability in several regions (Gerl and Sumper 1988). Although the regulation

of flagellin synthesis is unexplored, flagella switching in response to phototaxis stimuli has been demonstrated.

Phototaxis in *H. halobium* is mediated by two retinal-containing chromoproteins, sensory rhodopsin (SR) I and II (Spudich and Bogomolni 1988). Signal transduction is mediated by a methylation/demethylation system through photoactivation of SR-I in either its attractant or repellent signaling form, as well as through the repellent receptor SR-II. An SR-I$^+$SR-II$^+$ culture releases methyl groups when either SR-I or SR-II is stimulated, while an SR-I$^-$SR-II$^+$ mutant lacking a 25-kDa chromophoric protein and a 94-kDa methyl-accepting protein fails to respond to SR-I stimulation. Similarly, no response is seen in an SR-I$^-$SR-II$^-$ mutant which lacks both receptors (Spudich et al. 1989). The gene for SR-I has recently been cloned (Blanck et al. 1989), and the amino acid sequence deduced from the gene sequence can be arranged to conform to the secondary structure models previously developed for halobacterial light-driven ion pumps bacteriorhodopsin and halorhodopsin (Oesterhelt and Tittor 1989; Blanck and Oesterhelt 1987; Lanyi et al. 1990).

Several species of methanogens can fix nitrogen and contain DNA sequences homologous to the *nifH* gene (which encodes the nitrogenase iron protein) from *Klebsiella pneumonae* and *Anabaena* strain 7120 (Murray and Zinder 1984; Belay et al. 1984; Sibold et al. 1985; Souillard and Sibold 1986; Scherer 1989; Lobo and Zinder 1990). The putative *nifH* genes from *Methanococcus thermolithotrophicus*, *Methanobacterium ivanovii* and *Methanococcus voltae* have been cloned and sequenced, and show similarites to eubacterial NifH proteins at the amino acid sequence level (Souillard et al. 1988). Surprisingly, the methanogen sequences appear to be as distant from one another as they are from the eubacterial sequences (Souillard et al. 1988; Normand and Bousquet 1989). A plasmid carrying *M. voltae nifH* does not complement *K. pneumoniae nifH* mutants. A second region homologous to *Anabaena nifH* was detected in *M. thermolithotrophicus*. This region contains five open reading frames (ORFs), three of which could encode

proteins similar to eubacterial NifH, NifD and NifK (the structural components of nitrogenase), (Souillard and Sibold 1989). The product of this second *nifH* gene is more similar to the *Clostridium pasteurianum* NifH3 protein and the *Azotobacter vinelandii* NifH3 protein, than it is to the first NifH protein of *M. thermolithotrophicus* -- raising questions of its origin. The *nifH* gene is physically separated from *nifD* (and *K*) by two ORFs which could encode proteins significantly similar in sequence, and is transcribed separately from *nifD* (and *K*). Expression of *nifD,K* and the second *nifH* are responsive to nitrogen-fixing conditions, but the first *nifH* gene is not transcribed. How nitrogenase production is regulated in *M. thermolithotrophicus* is still unclear.

4. To develop molecular genetics

Reason. The value of genetic tools in dissecting biological systems is obvious. The emphasis here is on the urgency for development, as the comparative approach for predicting archaebacterial gene structure and function has reached its potential. Further cloning and sequencing of archaebacterial genes can add little insight into the biological function of predicted open reading frames, putative secondary structures and consensus sequences alike. Deeper understanding demands the development of reliable techniques which allow us to do archaebacterial genetics in archaebacteria (Lam et al. 1990a).

***Hf. volcanii* as a model host.** Halophilic archaebacteria can be easily maintained at moderate temperatures, and require no elaborate growth facilities. *Hb. halobium*, the best studied of the halophiles (primarily because of the tremendous interest in its light-harvesting bacterio-opsin system and because of its unusually high genetic variability), was initially chosen for the development of a transfection protocol.

We (our laboratory) have decided to adapt this protocol for *Hf. volcanii* which we hope to become a model for the archaebacteria (or at least for the halophiles). *Hf. volcanii* has fewer insertion sequences than *Hb. halobium*, and therefore is believed to be more genetically stable (Charlebois and Doolittle 1989). Natural mating and protoplast fusion

have been demonstrated by Mevarech and coworkers (Rosenshine et al. 1989). Wild-type *Hf. volcanii* is able to synthesize all 20 amino acids plus nucleotides, because it can grow on defined medium with ammonia as the sole nitrogen source. Mevarech and Werczberger (1985) showed that adenine-, serine-, methionine- and proline-requiring mutants can be generated by means of EMS treatment.

Rationale of the work reported here. Molecular genetics hinges on the ability to introduce exogenous DNA into a cell, so transformation technology is a priority. A vector system is necessary to facilitate DNA transfer -- preferably, a shuttle vector system that allows direct selection in *E. coli* and in halophiles for cloning archaebacterial genes, and for reintroducing genes altered *in vitro* for functional analyses. Using the endogenous plasmid pHV2 and a strain of *Hf. volcanii* cured of this plasmid, a host vector system was developed (see *Results and Discussion*). The pHV2-based shuttle vectors are marked by mevinolin resistance in *Hf. volcanii* and ampicillin resistance in *E. coli*. These vectors efficiently transform *Hf. volcanii* using a protocol modified from that used to transfect *Hb. halobium*. Isolation of the mevinolin resistance marker has subsequently led to the characterization of the *Hf. volcanii* HMG CoA reductase gene and its promoter, as well as revealing the mechanisms for drug resistance.

Efficient transformation of *Hf. volcanii* allows the development of gene replacement techniques which have proven so useful in targetting foreign DNA into yeast chromosomes, and in simple strain construction in eubacteria. Genes, isolated using the vector transformation system, provided the necessary genetic markers for replacing the *trpB* gene on the *Hf. volcanii* chromosome with the imported mevinolin resistance gene (see *Results and Discussion*).

Methods for genetic mapping are essential for linkage analysis, which is instrumental in elucidating operon structure and genome organization in eubacteria. Newly developed technologies (namely, the ability to generate auxotrophic mutants using ethyl methanesulfonate, the ability to transform auxotrophs to prototrophy using cosmid

clones, and the ability to clone genes using shuttle vectors) permit direct mapping of genetic loci to individual cosmid clones. An ordered set of cosmids has been assembled, in our laboratory, in preparation of a restriction map of the *Hf. volcanii* chromosome (Charlebois et al. 1989; Conover and Doolittle 1990; Lam et al. 1990b). Fine-scale mapping could be achieved by further transformation experiments using subfragments of the cosmid clones containing the gene of interest.

In this thesis, I will describe the development of (i) techniques for introducing plasmid DNA into spheroplasts of *Hf. volcanii*, (ii) shuttle-vectors selectable in *E. coli* and *Hf. volcanii*, (iii) methods for host strain construction by direct manipulation of the chromosome and (iv) methods for mapping genetic markers to the chromosome. Tryptophan (*trp*) biosynthesis is used as a model system to test and to refine the genetic tools developed for *Hf. volcanii*. These tools have been applied to begin a preliminary description of genetics in *Hf. volcanii* at the gene, operon and chromosomal level.

MATERIALS AND METHODS

This section of the thesis describes general methods used throughout my work. Techniques applied to specific experiments are described in the relevant sections in *Results and Discussion.*

Materials. Restriction endonucleases, T4 DNA ligase, nucleases, polymerases, and avian myeloblastosis virus (AMV) reverse transcriptase were purchased from Boehringer Mannheim, New England Biolabs and Pharmacia. Nucleotides were also from Pharmacia. Radioactively labeled nucleotides were either synthesized by J. D. Hofman or purchased from Amersham, Dupont or ICN Biochemicals. Solvents, organic and inorganic reagents came from Anachemia, Sigma and BDH chemicals. Ingredients for growing bacteria were obtained from Difco and Sigma. Mevinolin (1,2,6,7,8,8a-hexahydro-b,d-dihydroxy-2,6-dimethyl-8-[2-methyl-1-oxobutoxyl-1-naphthalene heptanoic acid d-lactone, Alberts et al. 1980) was generously provided by A. Alberts of Merck, Sharp and Dohme International.

Halophile strains, plasmids and phage. Bacteriophage ϕH and its natural host *Hb. halobium* R1 were obtained from W. Zillig, Max-Planck-Institut fur Biochemie, Martinsried. *Hf. volcanii* DS2 (from C. R. Woese, University of Illinois) contains at least two plasmids, including the 90 kbp pHV1 and the 6 kbp pHV2 (Pfeifer et al. 1981). WFD11, a strain cured of pHV2, was derived from DS2 after exposure to ethidium bromide (Charlebois et al. 1987). Auxotrophic and drug resistant mutants are listed and described in the appropiate sections in *Results and Discussion.*

Culture conditions. Both *Hb. halobium* and *Hf. volcanii* cells were grown in minimal or rich medium previously described by Mevarech and Werczberger (1985), except that Tris-HCl, pH 7.2 was added to a final concentration of 50mM and the concentration of $CaCl_2$ was increased to 7 mM. When necessary, minimal medium was supplemented with specific amino acids (50mg/L) or nucleotides (25mg/L). Mevinolin-resistant *Hf.*

volcanii strains were maintained on minimal medium containing 10 μM mevinolin or on rich medium with 40 μM inhibitor. Cultures were incubated at 42°C for agar plates and at 37°C for liquid media with vigorous shaking (at 110 rpm on a platform shaker).

Chemical mutagensis. Typically, a 30 ml culture of *Hf. volcanii* WFD11 (at OD_{600} of 1) was harvested by centrifugation, and resuspended in 30 ml of a salt solution (containing, per liter, 206 g of NaCl, 37 g of $MgSO_4 \cdot 7H_2O$, 4 g of KCl) with 0.2M Tris-HCl pH7.2. One milliliter cell suspensions were shaken with 10μl of liquid EMS or 0.5μl of MMS, or with no mutagen, at 37°C for varying lengths of time. Mutagenesis treatment was terminated by the addition of 40 ml of medium salts. Washed cells were allowed to grow in rich medium for one day, and then diluted and plated onto rich agar medium. Percent survival is calculated using the number of recovered colonies in chemical-treated and untreated experiments (figure 1).

DNA preparations. To obtain genomic DNA, actively growing cells were harvested by centrifugation (at 3,300 x *g* for 15 minutes). Cell pellets were resuspended and lysed in one half growth volume of 10 mM Tris-HCl, pH 8.0/1mM EDTA and extracted twice with phenol. DNA remaining in the aqueous phase was then either spooled into ethanol or precipitated with two volumes of ethanol without addition of any ammonium or sodium salts, and dissolved in buffer.

Phage ϕH DNA was isolated according to the description by Schnabel et al. (1982). Phage particles were precipitated from filtered lysates with polyethyleneglycol (PEG 6000) and further purified twice on cesium chloride gradients. Phage DNA was obtained by solvent extractions followed by extensive dialysis.

Plasmid DNA was isolated from *Hf. volcanii* typically by alkaline extraction (Maniatis et al. 1982). When necessary, plasmid preparations were further fractionated by rate zonal centrifugation (on 10-40% sucrose gradients prepared in 1M NaCl/5mM EDTA/20mM Tris-HCl pH 8) to separate the contaminating 90kbp endogenous pHV1 from the plasmid of interest.

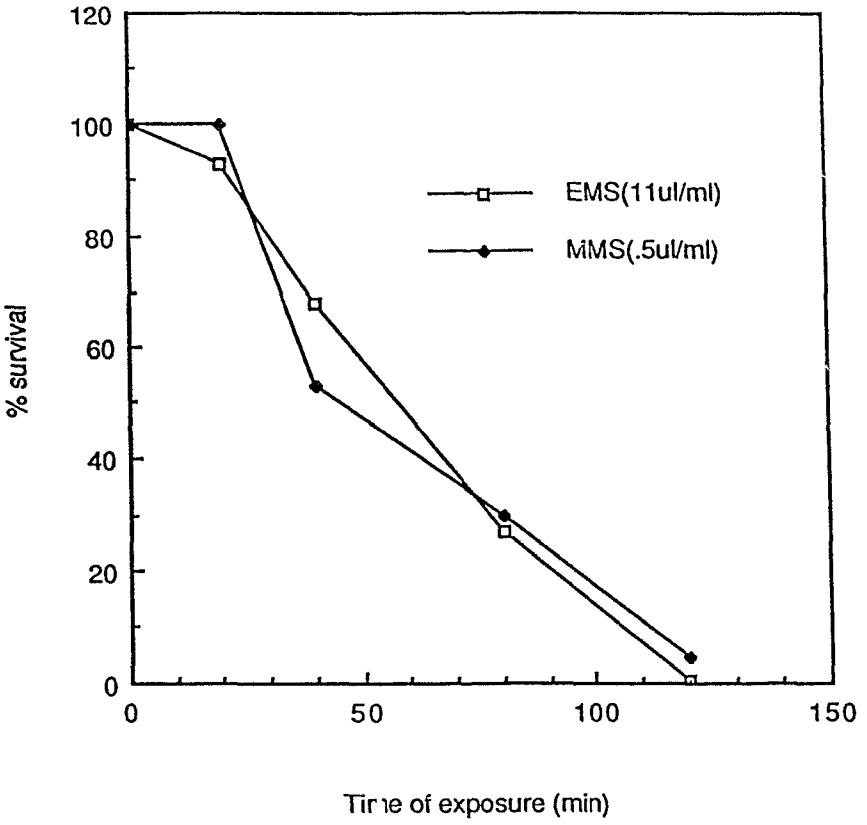


Figure 1. Conditions for mutagenesis. An exponentially growing culture (approximate OD_{600} of 1) was pelleted and resuspended in a solution of medium salts with 0.2M Tris-HCl pH7.2 (to the same cell density). One milliliter cell suspensions were shaken with 10µl of liquid EMS or 0.5µl of MMS, or with no mutagen, at 37°C for varying lengths of time. Mutagenesis treatment was terminated by the addition of 40 ml of medium salts. Washed cells were allowed to grow in rich medium for one day, and then, diluted and plated onto rich agar medium. Percent survival is calculated using the number of recovered colonies in chemical-treated and untreated experiments.

Spheroplast preparation. Exponentially growing cells (between 1 to 1.5 A_{600}) were pelleted and resuspended in spheroplasting solution. Spheroplasting solution for *Hb. halobium* contained 2M NaCl, 27mM KCl, 50mM Tris-HCl (pH 8.75) and 15% sucrose. Spheroplasting solution for *Hf. volcanii* contained 1M NaCl, 25mM KCl, 50mM Tris-HCl (pH 8.2) and 15% sucrose. (*Hf. volcanii* grows optimally at a lower NaCl concentration.) Cells resuspended in spheroplasting solution plus 15% glycerol could be frozen and stored at -70°C (Cline et al. 1989a).

Transformation of *Hf. volcanii* spheroplasts. Input DNAs were prepared in 125mM EDTA, pH 8, typically by adding 5 µl of 0.5M EDTA (pH 8.0) to 15 µl DNA samples (including ligation mixtures). The presence of carrier RNA in the DNA sample does not interfere with transformation. Five minutes after the addition of DNA (20 µl) to the spheroplasts (220µl), an equal volume (240 µl) of polyethylene glycol solution [60% purified PEG 600 (v/v) in spheroplasting solution] was blended into the spheroplast-DNA mixture by repeated gentle inversions. After a further 5 to 20 minute incubation at room temperature, 1 ml of regeneration salt solution [3.5M NaCl, 150mM $MgSO_4$, 50mM KCl, 7 mM $CaCl_2$, 50mM Tris-HCl (pH 7.2), 15% sucrose (w/v)] was added to the 480 µl transformation mixture. Cells were pelleted by centrifugation at 6,500 rpm for 7 minutes at room temperature in an MSE microcentrifuge. In cases that require time for gene expression, cell pellets were resuspended in 1 ml of rich medium containing 15% sucrose and incubated for 6-12 hours at 42°C before plating. 100 µl samples of the appropriate dilutions were mixed with 3 ml of top agar [medium containing 0.7% agar and 15% sucrose (w/v), kept at 60°C], and poured onto selective agar plates. Mevinolin, when necessary, was added to the top agar to a final concentration of 10µM. Mevinolin resistant transformants appears after 10 to 14 days incubation.

Transfection with ϕH DNA. EDTA, phage DNA and PEG solution were sequentially incubated with the spheroplasts as above. A 200 µl inoculum from an *Hb.*

halobium R1 culture was added to each 3 ml of top agar to provide an indicator lawn (for *Hf. volcanii* and *Hb. halobium* transfectants -- since ϕH will not infect *Hf. volcanii*).

PEG purification. The quality of polyethyleneglycol affects transformation efficiency. PEG 600 purchased from Sigma is routinely purified by a method modified from that described by Klebe et al. (1983). PEG 600, dissolved in benzene, was extracted three times with equal volumes of isooctane. PEG was then precipitated from these solvents simply by chilling on ice. Repeated liquid extractions with ether at room temperature rid the PEG of residual benzene and isooctane. Residual ether was evaporated at 60°C in a waterbath.

Transformation using cosmid DNA pools. A multiplex system was devised to handle the large number of transformation mapping experiments. Procedures were carried out at room temperature. Three hundred µl of 0.5M EDTA, pH8.0 was added to 2.4 ml of frozen *Hf. volcanii* spheroplasts of each auxotroph. Eighty µl aliquots of the mixtures were dispensed into 27 microtiter wells containing about 2µg of DNA dissolved in 0.1M EDTA. The DNA preparations were pools of cosmid DNAs, constructed in such a way that each of 154 cosmids was present in two of 25 different pools. Two control transformations -- no DNA and wild-type *Hf. volcanii* DNA, were included. Ninety µl of a PEG solution [60% PEG600 (vol/vol)/0.4M NaCl/10mM KCl/ 6% sucrose(wt/vol)/50mM Tris-HCl, pH7.5] was then added, and mixed by gentle pipetting with a multi-channelled pipettor, to prevent cell lysis. Transformation was terminated by adding 90µl of regeneration solution (described above). One tenth volume of each transformation mixture (containing about 10^8 spheroplasts) was plated in top agar onto minimal agar plates. The small amount of PEG600 and EDTA carried over does not noticeably hinder the regeneration of spheroplasts. Transformants appear after 7 to 10 days.

Recombinant DNA manipulations in *E. coli*. Molecular cloning and analysis of cloned DNA, unless otherwise specified, were performed using standard methods

(Maniatis et al. 1982; Messing 1983; Sambrook et al. 1989). Transformation of *E. coli* cells was carried out according to the method of Chung and Miller (1988).

Hybridizations. DNAs immobilized to GeneScreen Plus™ hybridization transfer membrane or Colony/Plaque Screen™ (Dupont) were hybridized with ^{32}P-labeled DNA probes under conditions recommended by the manufacturer. Conditions for experiments using labeled oligonucleotides as hybridization probe are described by Zeff and Geliebter (1987). Plasmid DNA probes were labeled with [α^{32}P] dATP or [α^{32}P] dGTP by nick-translation (Maniatis et al. 1982) and by random priming (Feinberg and Vogelstein 1983; 1984). M13 probes were synthesized from universal primers which anneal to sequences upstream or downstream of the polylinker region in M13 (Messing 1983). Synthetic oligonucleotides were labeled with [γ^{32}P] ATP using polynucleotide kinase (Geliebter 1987).

Deletion analysis. Specific sections of pHV2-based plasmid clones were deleted with various combinations of restriction endonucleases. After removing the 3'-overhangs using T4 DNA polymerase and filling in the 3'-recessed ends using the Klenow fragment of DNA polymerase (Maniatis et al. 1982), these DNA samples were separately self-circularized at 0.1μg of DNA in 300μl, using 3 units of ligase, under conditions recommended by the manufacturer. Plasmids retrieved from transformed *E. coli* DH5α were analysed for altered restriction patterns. Various deleted forms were tested for their ability to transform *Hf. volcanii* strains to give the expected phenotype.

DNA sequencing and computer analysis. DNA fragments were subcloned into M13 vectors (M13 um vectors from International Biotechnologies Inc. and M13 tg vectors from Amersham). Ordered sets of deletions were generated using ExoIII nuclease (Henikoff 1984) and Mung bean nuclease. Both strands were sequenced by the dideoxy-chain termination method (Sanger et al. 1977) using the Taquenase system (US Biochemicals, Ohio). Deoxyguanosine triphosphate was replaced by its 7-deaza-analog in

all sequencing reactions. Sequencing reaction products were resolved on polyacrylamide gels containing either 50% urea (w/v) or 85% formamide (v/v) as denaturant.

DNA sequences were analyzed for open reading frames, direct and inverted repeats and base composition using standard software (Queen and Korn 1984; Marck 1988). Data bank searches were carried out with FASTA (Pearson and Lipman 1988) and with Microgenie from Beckman (Queen and Korn 1984). Deduced amino acid sequences were aligned using MULTALIN software of Corpet (1988).

RNA isolation and transcript analysis. RNA was prepared from early-to mid-log-phase cultures. Typically, a 40 ml culture (between 0.4-0.6 OD_{600}) was chilled and centrifuged at 10,000g for 5 minutes. Chilled cell pellet was suspended in 0.5 ml of medium salts [3.5M NaCl, 150mM $MgSO_4$, 50mM KCl, 7mM $CaCl_2$, 50mM Tris-HCl (pH 7)], lysed in a guanidiniun-sarcosyl solution, extracted with acidic phenol and chloroform and precipitated repeatly with ethanol, according to the method described by Chomczynski and Sacchi (1987). RNA pellet was dissolved in 0.1M sodium acetate/5mM $MgSO_4$,pH5 and digested with RNase free DNaseI. After phenol extractions and ethanol precipitations, a sample of the dissolved RNA was checked on a 1.2% agarose gel. Primer extension experiments were performed as described by Geliebter (1987) using AMV reverse transcriptase but without actinomycin D.

RESULTS AND DISCUSSION

I. Developing genetic tools

1. Plasmid transformation

The first demonstration of plasmid transformation was the result of the joint effort between a former graduate student and myself (Charlebois et al. 1987). Charlebois provided the sequence of pHV2, a cloned pHV2 probe and a strain of *Hf. volcanii* cured of pHV2 by ethidium bromide treatment. (This cured strain is referred to as strain WFD11.) pHV2 becomes the logical choice to attempt transformation, even though pHV2 does not confer any readily detectable phenotype other than its existence.

Demonstrating plasmid transformation. We used the procedure described for transfecting *Hb. halobium* with phage ϕH DNA (Cline and Doolittle 1987), except with reduced NaCl, to transform *Hf. volcanii* WFD11 with 0.1μg of pHV2 (made from *Hf. volcanii* to avoid any complications caused by the host restriction systems). Regenerated colonies, which appeared after 10 days to 2 weeks of incubation, were transferred with tooth picks to fresh agar plates. Of the 10,100 colonies picked, 18 gave a positive hybridization signal with ^{32}P-labeled-cloned-pHV2 probe after colony lifting (methods given in detail in Charlebois et al. 1987). Further analysis of DNA extracted from these 18 potential transformants verified that 17 of them contained a plasmid of 6.4 kbp. These results were extrapolated to provided an optimistic estimation of transformation efficiency of 10^7 transformants per μg of pHV2, with 2% of the regenerated spheroplasts expected to be transformed.

Confirmation experiment. To rule out the possibility of contamination by pHV2-harbouring cells during transformation, these experiments were repeated using a "tagged" plasmid. The marked plasmid pHV2Δ93 was constructed *in vitro* from native pHV2

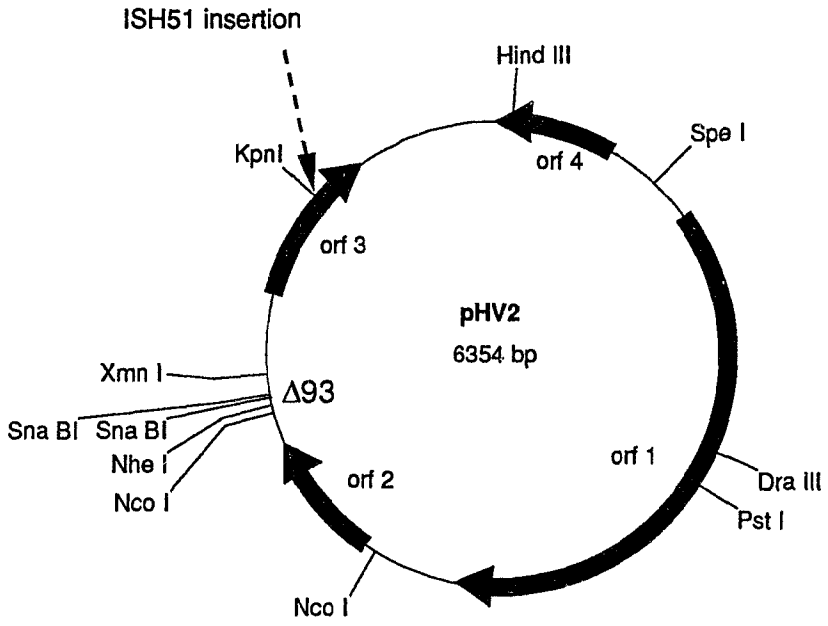


Figure 2. *Hf. volcanii* plasmid pHV2. Arrows indicate the 4 major ORFs predicted by the nucleotide sequence of pHV2 (see Charlebois et al. 1987). Restriction sites relevant for later vector construction , as well as the artificially introduced deletion between the 2 SnaBI sites to generate pHV2Δ93 are indicated. Insertion of an ISH51 element into ORF 3, detected in a natural variant called pHV51, does not interfere with plasmid replication and maintenance.

DNA by deleting a 93 bp SnaBI fragment from the spacer region between ORF 2 and ORF 3 of pHV2 (figure 2). Transformation of WFD11 with 25 ng of circular pHV2Δ93 yielded two positive transformants out of the 1300 regenerated colonies screened by hybridization. Southern analysis confirmed that these two transformants both contained the expected plasmid with the specific 93 bp deletion (Charlebois et al. 1987).

2. Development of shuttle vectors

Selectable marker. Determinants available for direct selection of transformants are few; archaebacteria are insensitive to most of the antibiotics used in eubacterial vector-host systems. In 1986, Cabrera and coworkers reported that mevinolin, an inhibitor of eukaryotic HMG CoA reductases, also strongly inhibits this enzyme in halobacterial extracts, and prevents growth of *Hb. halobium* in liquid media (Cabrera *et al.* 1986). *Hf. volcanii* is also sensitive to mevinolin (figure 3A), and cell growth is completely inhibited at mevinolin concentrations of 1-2 μM and 20-40 μM on plates of minimal and rich agar respectively. Mevalonic acid, the product of the HMG CoA reductase catalyzed reaction, prevents mevinolin inhibition (figure 3B). Mutant colonies resistant to the drug appear spontaneously at a frequency of about one in 10^9 cells plated (figure 3C) and grow well at mevinolin concentrations up to 100 μM on minimal plates. I concluded that mevinolin resistance, if it were due to a stable genetic mutation and could be genetically transferred to sensitive cells, might be employed as a selectable marker for *Hf. volcanii* plasmid vectors.

Transformation experiments with genomic DNA from mevinolin-resistant mutants. DNAs from four spontaneous mevinolin-resistant mutants (isolated independently from minimal plates with 40 μg/ml mevinolin) were pooled and used in transformation experiments with spheroplasts of *Hf. volcanii* WFD11, as described in *Materials and Methods*. After 6 hours in rich medium to allow for gene expression, cells

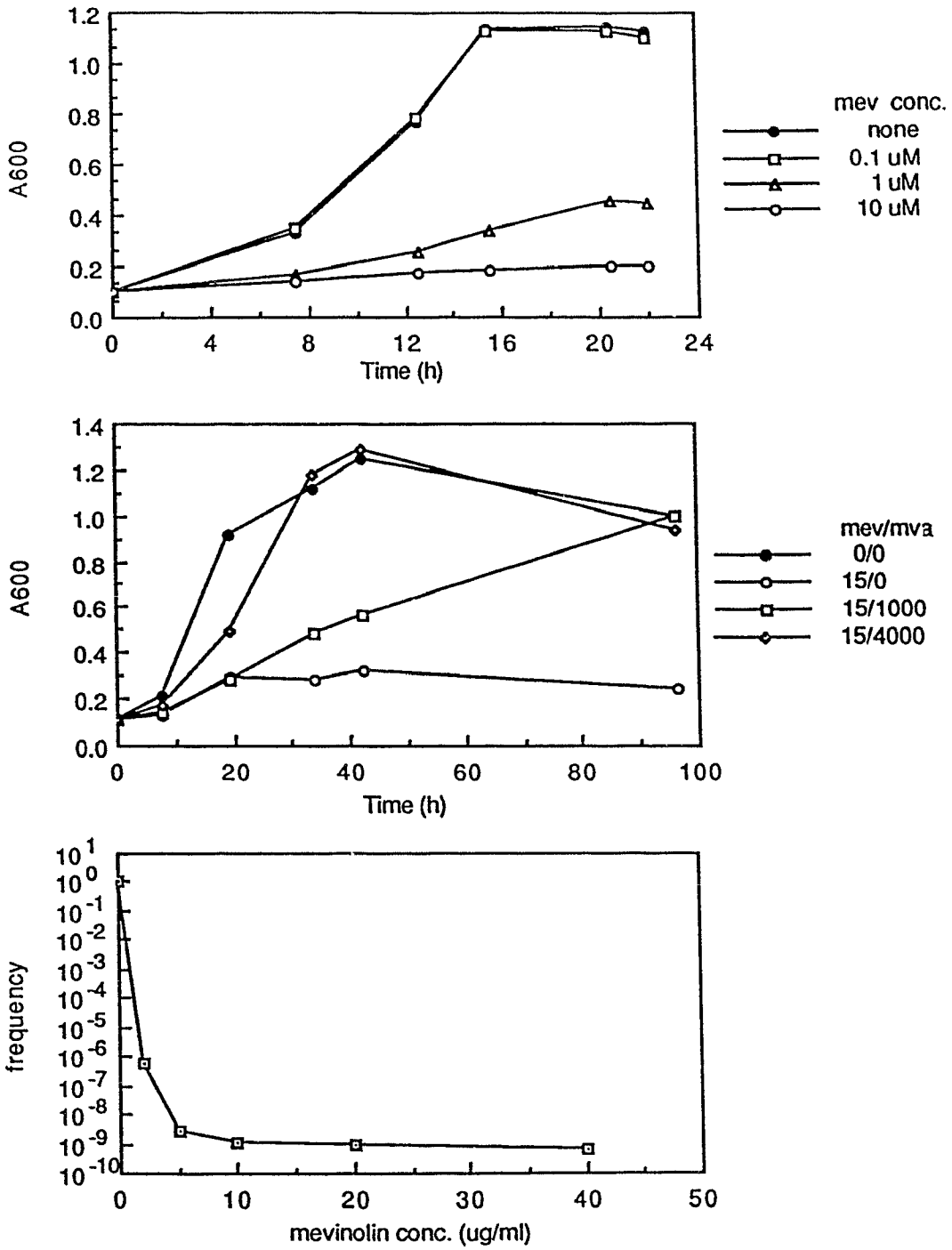


Figure 3. Mevinolin inhibition of *H. volcanii*. (A) shows concentration-dependent growth inhibition of *Hf. volcanii* by mevinolin in minimal liquid medium. Mevinolin (mev) concentrations are indicated. (B) shows the effect of mevalonic acid on mevinolin inhibition. Concentrations of mevinolin (mev) and mevalonic acid lactone (mva) are given in μM. (C) Frequency of spontaneous mutants resistant to mevinolin was determined by spreading a known number of wild-type *Hf. volcanii* cells on solid agar medium containing various concentrations of mevinolin

were plated on minimal regeneration medium containing 10 μM mevinolin. After ten days on plates, resistant transformants were found at frequencies of approximately $4x10^3$ per μg DNA -- more than 2,000 colonies among 1-2 $x10^9$ regenerated spheroplasts, when 0.5 μg of total high-molecular-weight DNA from resistant cells was used (for instance, see Table 1). Control transformations (wild-type or no DNA) yielded resistant colonies at a frequency of 1-5 among 10^9 regenerated spheroplasts, the same as the mutation frequency observed with intact cells. Thus, mevinolin resistance is heritable, and can be transferred.

Isolation of mevinolin resistance marker. A rare variant of pHV2, pHV51, was used as a vector. pHV51 was fortuitously detected during plasmid screening of different colonies of *Hf. volcanii* strain DS2. It resulted from the insertion of a single copy of the transposable element ISH51 at about position 5,500 of the pHV2 sequence near the unique KpnI site within ORF 3 (refer to figure 2 above). A cloned ISH51 probe (Hofman et al. 1986) hybridizes to pHV51. Purified pHV51 and the cloned probe gave identical hybridization patterns against digested genomic DNA. The inserted ISH51 provides convenient sites for cloning with MluI and EcoRI, and further insertions into the ISH51 moiety of pHV51 seem unlikely to disrupt functions essential for plasmid maintenance. Pfeifer has recently published the sequence of the ISH51 element in pHV2 (Pfeifer and Blaseio 1990).

pHV51 DNA was isolated from *Hf. volcanii* cultures, digested with MluI and ligated with MluI-digested DNA from mevinolin-resistant cells. Spheroplasts of *Hf. volcanii* WFD11 were transformed with such ligated DNA preparations (the "MluI library") and plated on minimal agar containing mevinolin. Transformants were readily obtained, with the DNA preparation ligated with pHV51 showing about three times as many transformants as a DNA preparation partially digested with MluI but not ligated to MluI-digested pHV51. Transformants were screened by colony hybridization with

Table 1. Transformation of WFD11 with mevinolin resistance-conferring plasmids from an MluI library.

Source of DNA	µg DNA	transformants	mevinolin resistant transformants per microgram
none	---	$<10^1$	---
mevinolin res.*	0.5	2.2×10^3	4×10^3
M3	0.02	2.3×10^6	1×10^8
M3 - PstI+	0.02	1.3×10^4	6×10^6
M9 (pWL2)#	0.015	1.5×10^6	1×10^8
M9 - PstI+	0.015	1.9×10^4	1×10^6
M9 - HindIII+	0.015	1.4×10^4	9×10^5
M10	0.005	2.8×10^5	6×10^7
M10 - HindIII+	0.005	2.6×10^3	5×10^5

Spheroplasts of strain WFD11 (cured of pHV2) were incubated with total cellular DNA from mevinolin resistant cells ("mevinolin res.") or plasmid DNA preparations and polyethylene glycol, as described in Materials and Methods, and plated on minimal agar in the presence of 10 µg/ml mevinolin, in serial ten-fold dilutions. At least 100 colonies at the appropriate dilution were counted for each transformation (except "no DNA" control).

* DNA from four spontaneous mevinolin-resistant strains, mixed in equal proportions. Source of DNA for preparation of MluI library.

+ DNA digested to completion with indicated restriction endonuclease, to linearize plasmid, before transformation.

The plasmid borne by strain M9 was designated pWL2.

labeled pHV2 DNA; most had reacquired pHV2-related sequences simultaneously with resistance.

To show that these hybridization-positive transformants bore independently-replicating plasmids with chromosomal DNA inserted into pHV51, plasmid DNAs were made from them. In some cases these plasmids bore more than a single inserted MluI fragment, but all of the 24 tested did carry a common 7.9 kbp MluI fragment (figure 4). The 7.9 kbp MluI fragment from the plasmid borne by one such transformant, designated M9 in figure 4, was used to probe the total DNA of two independent transformants, M3 and M10, and itself. Strong hybridization signals confirmed that the cloned fragments were identical (figure 4C, 4E).

Plasmids were also detected in mevinolin-resistant transformants obtained when WFD11 was transformed with a library prepared with KpnI-digested DNA from mevinolin-resistant cells and KpnI-digested pHV2. These contained large (>20 kbp) insertions, and have not been characterized further.

When plasmids from mevinolin-resistant transformants such as those designated M3, M9 and M10 in figure 4 were isolated and used once again to transform WFD11, mevinolin-resistant colonies were obtained at frequencies between 5 and 10 x 10^7 per μg DNA, 10,000-fold higher than the frequency obtained with uncloned total DNA from resistant cells (Table 1). Presumably, this increase reflects primarily the over one-thousand-fold enrichment provided by cloning, although the fact that plasmid-borne markers need not be integrated into the chromosome to be expressed, and that circular DNAs may be more easily taken up (and/or recombined into the chromosome) may well also partly explain this. When plasmids from M3, M9 or M10 were linearized by digestion with PstI or HindIII (which cut once within pHV51 and not within the inserted mevinolin-resistance fragments; Table1), transformation frequencies dropped about 100-fold.

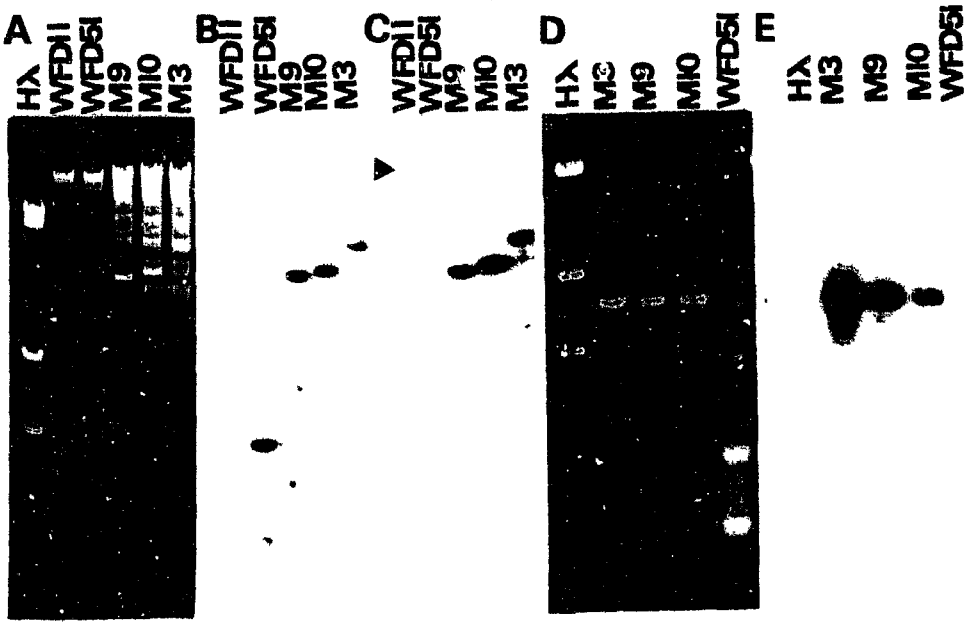


Figure 4. Characterization of mevinolin-resistant transformants obtained with MluI library. (A) Total DNAs from transformants M3, M9 and M10 (see Table 1), WFD51(which contains pHV51) and WFD11(which has been cured of pHV2) were digested with HindIII and resolved on a 0.5% agarose gel. Hλ indicate HindIII digested λ DNA as size markers. (B) The gel shown in A, blotted and probed with ^{32}P-labeled pHV2 DNA. (C) A duplicate blot probed with a 7.9 kbp Mlu fragment of chromosomal DNA excised from the mevinolin-resistance-conferring plasmid extracted from strain M9. Dark arrow indicates position of faintly-hybridizing band (visible on darker exposures) we assume to be the homologous chromosomal region of the recipient, which would be present in 6-fold fewer copies (Charlebois et al. 1987). (D) Plasmid DNAs isolated from mevinolin-resistant transformants, digested with MluI plus PstI, and resolved on a 0.7% agarose gel. (E) The gel shown in D, blotted and probed with the same 7.9 kbp MluI fragment described in C above.

Construction and transformation of shuttle vectors. Plasmids from transformants M3, M9 and M10 (see above) can be used as selectable vectors for cloning into *Hf. volcanii*, though recovery of these low copy number plasmids from *Hf. volcanii* is inefficient. For further development of a shuttle vector the mevinolin resistance-conferring plasmid pWL2 borne by strain M9 was selected as starting material. A restriction map of pWL2 is shown in figure 5A; the chromosomal insertion determining resistance is 7.9 kbp, and pWL2 has unique sites for NheI, PstI, SpeI, DraIII and HindIII. EcoRI and SnaBI insertions can also be made easily.

Hybrid vectors for transformation into *E. coli* were prepared by ligation of pWL2 with the pBR322 derivative pAT153 (Twigg and Sherratt,1980), which lacks 622 bp of pBR322, but retains its ampicillin- and tetracycline-resistance determinants. These hybrid constructs are also shown in figure 5A. For pH455, HindIII-digested pWL2 was ligated with HindIII-digested pAT153 (disrupting its tetracycline-resistance determinant). In p74, the ampicillin-resistance gene is disrupted by ligation of PstI-digested plasmids. Plasmids pE3- and pE3+, obtained by ligation of EcoRI-digested plasmids, retain both resistance determinants, and differ in the orientation of their pAT153 moieties.

Competent *E. coli* DH5α cells were transformed with the hybrid constructs shown in figure 5A. Plasmids were prepared from the ampicillin- or tetracycline-resistant *E. coli* transformants obtained and used to transform *Hf. volcanii* WFD11 spheroplasts, selecting for resistance to mevinolin (Table 2). As one control, I "reconstructed" pWL2 *in vitro* by HindIII digestion of *E. coli*-propagated pH455, removal of pAT153, and religation. (This plasmid, designated pΔAT in Table 2) should lack modifications characteristic of DNA prepared from *Hf. volcanii* cells.) With pΔAT, or with any of the constructs containing pAT153, mevinolin-resistant transformants were obtained at frequencies of about 10^4 per microgram DNA. When DNA was prepared from such transformants, plasmids of approximately 20 kbp (the expected size) were detected.

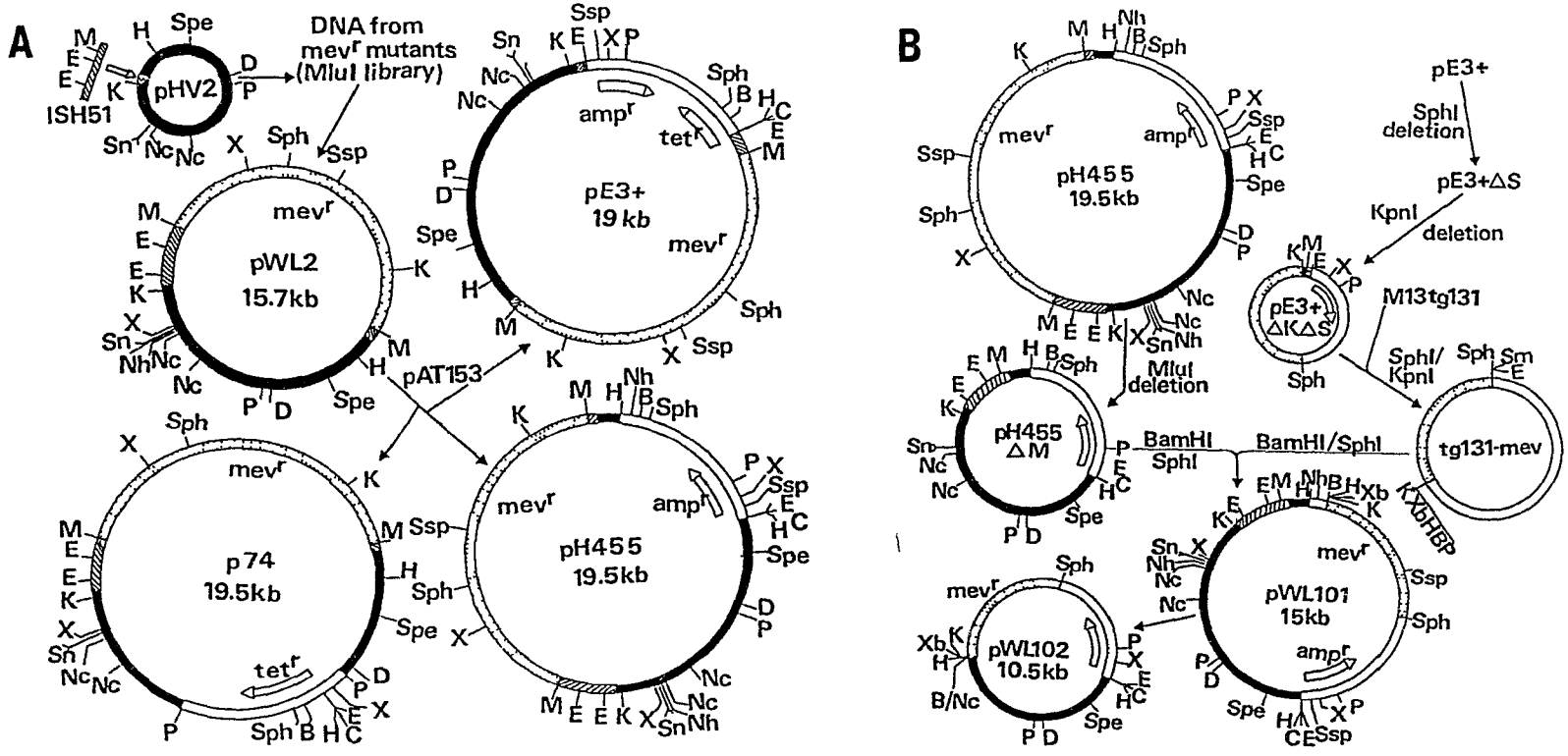


Figure 5. Construction of shuttle vectors. (A) shows the restriction maps of pWL2 and hybrids constructed with pWL2 and pAT153. pWL2 is the mevinolin-resistance-conferring plasmid isolated from transformant M9. Solid shading indicates pHV2 sequence, hatched shading indicates ISH51 sequence, and the dotted region is the mevinolin-resistance-conferring MluI fragment of chromosomal DNA. The unshaded region in pH455, p74 and pE3+/pE3- is the pBR322 derivative pAT153, linearized with HindIII, PstI or EcoRI, respectively, and ligated into the corresponding sites on pWL2. (B) shows the construction of pWL101. The mevinolin-resistance fragment was excised from pH455 with MluI, to produce pH455ΔM. This was digested with BamHI and SphI, and ligated with BamHI-SphI digested M13-tg131-mev, which was obtained by ligation of the 3.5 kbp KpnI/SphI mevinolin resistance determinant of pE3+ into M13-tg131 vector from Amersham. Restriction endonuclease designations, in this and subsequent figures are: B, BamHI; C, ClaI; D, DraIII; E, EcoRI; H, HindIII; K, KpnI; M, MluI; Nc, NcoI; Nh, NheI; P, PstI; Spe, SpeI; Sph, SphI; Sm, SmaI; Sn, SnaBI; Ssp, SspI; X, Xmn; Xb, XbaI.

Table 2. Transformation of WFD11 with pWL2-pAT153 hybrid constructs (figure 5A), prepared from *Escherichia coli.*

Source of DNA	µg DNA	transformants	mevinolin resistant transformants per microgram
none	----	1×10^1	----
mevinolin res.*	0.5 µg	2.0×10^3	4×10^3
pH455	2 µg	2.9×10^4	1×10^4
p74	2 µg	3.0×10^4	2×10^4
pE3-	2 µg	2.4×10^4	1×10^4
pΔAT**	1 µg	2.1×10^4	2×10^4

* DNA from four spontaneous mevinolin-resistant strains, mixed in equal proportions. Source of DNA for preparation of MluI library.

** Product of removal of pAT153 sequences from pH455 grown in *E. coli.* Should be identical in sequence to pWL2 (designated M9 in Table 1) but lack *Hf. volcanii*-specific modifications.

Crude alkaline plasmid preparations of pH455-bearing *Hf. volcanii* strains transformed *E. coli* DH5α to ampicillin resistance, and the ampicillin-resistant transformants acquired plasmids with the same restriction endonuclease digestion pattern as pH455 -- confirming its utility as a shuttle vector.

The relatively low transformation frequencies obtained with hybrid constructs prepared from *E. coli* (compared to around 1 x 10^8/μg obtained with *Hf. volcanii*-propagated pWL2) is probably caused by restriction in the archaebacterial host -- a similar reduction is observed when DNA from phage ϕH grown in *Hb. halobium* is used to transfect *Hf. volcanii*, and yet burst sizes in the two hosts are the same (Charlebois *et al.* 1987). Although a mutant of *Hf. volcanii* lacking this restriction system(s) has not yet been isolated, utility of the hybrid constructs as shuttle vectors is, for most purposes, not seriously compromised by restriction. A microgram of vector DNA prepared in *E. coli* will produce thousands of mevinolin-resistant *Hf. volcanii* transformants in a typical experiment, while on average less than one new spontaneously-resistant mutant appears on plates with comparable numbers (10^8) of regenerated spheroplasts after mock (no DNA) transformations.

Tailoring of shuttle vectors. To reduce the size and complexity of subsequent generations of shuttle vectors, specific regions of pWL2 and of the hybrid constructs were experimentally deleted.

The 7.9 kbp plasmid-borne mevinolin resistance marker could recombine with the chromosome of the sensitive recipient, a problem which could be reduced but not eliminated by shortening the region of homology. A deletion (resulting also in the loss of tetracycline resistance) between the two SphI sites of the construct pE3+ (figure 5A) did not destroy the ability to transform strain WFD11 to mevinolin resistance. Even a subsequent deletion between KpnI sites, which removed more of the mevinolin-resistance region and all but a few hundred base pairs of the pHV51 moiety, allowed transformation

to resistance, albeit at very low frequency (presumably reflecting recombination of the mevinolin-resistance region with the chromosome).

The 3.5 kbp SphI - KpnI fragment which, from these experiments, must contain a functional mevinolin resistance determinant, was recloned into pH455ΔM (pH455 religated after removal of the 7.9 kbp MluI fragment bearing the mevinolin resistance determinant; see figure 5B). The resultant plasmid, pWL101 contains complete pHV51 and pAT153 sequences (figure 5B). It was subjected to functional analyses by deletion into the pHV2 moiety from the unique ClaI and BamHI sites near the boundaries of the pAT153 component. (pAT153 replication and ampicillin resistance functions remain intact in such deletions.) Results of these analyses are presented in figure 6. A deletion of 762 bp of pHV2 between ClaI and SpeI sites reduced transformation frequency by over 20 fold. Deletions from the opposite direction up to the NcoI site do not affect ability to transform WFD11 to mevinolin resistance, but a further deletion, extending to the DraIII site does, again reducing transformation frequency by 20 fold. These experiments presumably have defined a region(s) important to plasmid replication in *Hf. volcanii*. The smaller number of transformants obtained with plasmids lacking this region(s) (or indeed lacking all pHV2-derived material) result from integration of the mevinolin-resistance determinant into the chromosome of WFD11 (see below).

DNA samples made from mevinolin resistant *Hf. volcanii* transformants obtained with the deleted vectors designated ΔBM, ΔK, ΔBSn, and ΔBNc in figure 6, contain circular plasmids, in all of twenty cases. Digestion with ClaI produced linearized plasmids of the expected sizes, with no obvious insertions or deletions (figure 6C). Southern analysis showed these plasmids bear pAT153 sequences (see figure 6D). No such plasmids were seen in preparations of DNA from the few mevinolin-resistant transformants obtained with deletions designated ΔBD and ΔBSpe, but colony hyridization detected pAT153 in some of these transformants, as might be expected from

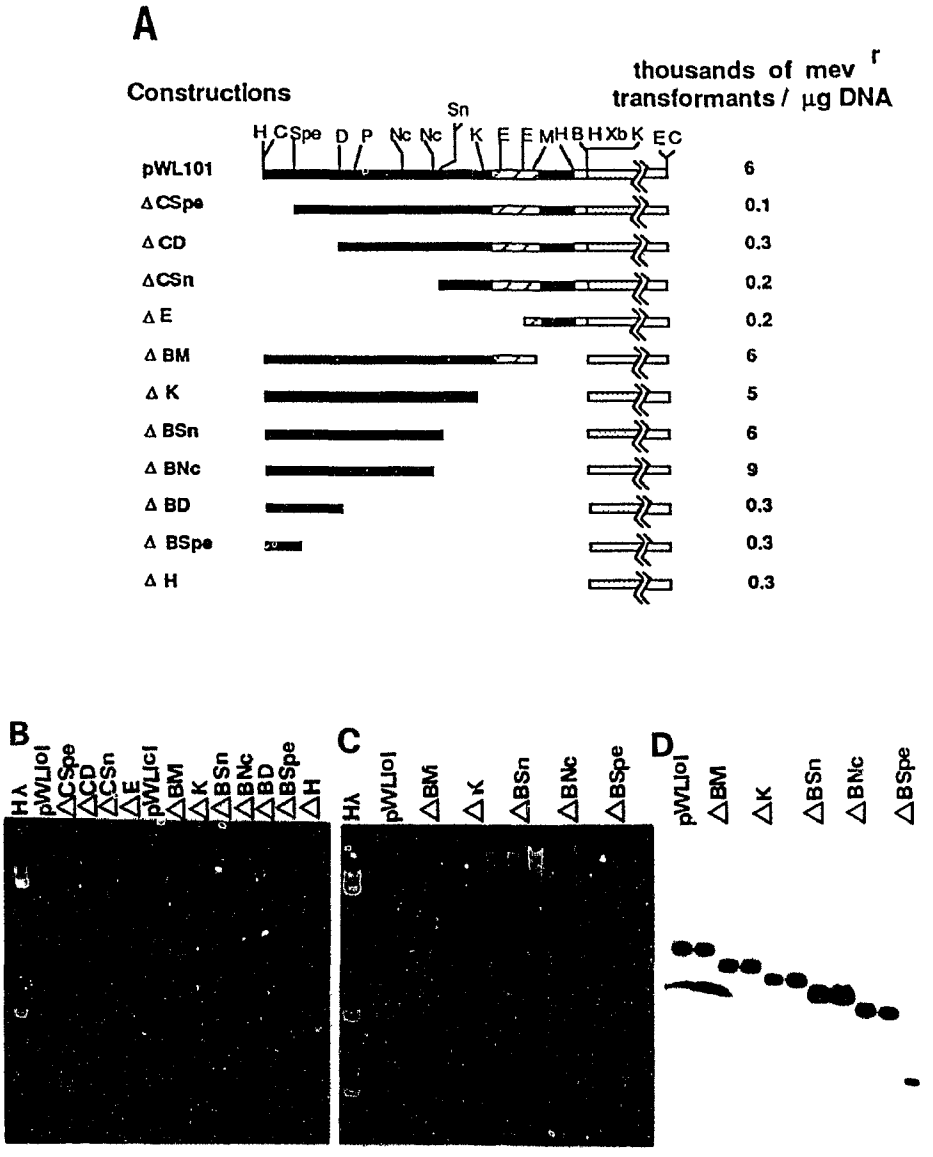


Figure 6. Deletion analysis of pWL101. Deletions were made by digesting pWL101 with the indicated enzymes. religating and transforming *E. coli* DH5α, with selection for ampicillin resistance. (A) Schematic representation of DNA remaining after deletion. pWL101 is shown as linearized at its unique ClaI site. Mevinolin resistance region, and ampicillin-resistance region of pAT153 (between broken lines) are not displayed, as these are retained in all deletions. Shading as in figure 5. Frequencies of transformation of WFD11 to mevinolin resistance obtained with each deleted plasmid are indicated. (B) Size analysis of deleted plasmids, prepared in *E. coli* DH5α. Plasmids were linearized and resolved on a 0.7% agarose gel. (C) ClaI digestions of total DNA obtained from mevinolin-resistant *H. volcanii* colonies from transformation with some of the deleted plasmids shown in B. Two independent transformants are presented for each deleted plasmid. (D) Southern hybridization of gel shown in C, with pAT153 DNA probe.

chromosomal integration, by a single crossover, of non-replicating but circular input DNA. With two such ΔBSpe transformants, Southern hybridization signals obtained when ClaI-digested total DNA was probed with pAT153 suggest that there are indeed chromosomal copies of this eubacterial sequences (figure 6D). It should prove easy to use recombination with non-replicating selectable markers for insertional and gene replacement mutagenesis (see below).

The 11.2 kbp pWL102 is an effective vector for shuttling between *E. coli* and *Hf. volcanii*, with selection for ampicillin- and mevinolin-resistance, respectively. It bears no ISH51 sequences, and contains unique sites for KpnI, XbaI, SphI, ClaI, and EcoRI. Insertions can be made into any of these sites without disrupting plasmid maintenance or resistance functions. The 15.0 kbp pWL101 has additional (BamHI, SnaBI and MluI) sites for cloning. These vectors, as they are, have proven useful in isolating genes in the leucine and tryptophan biosynthetic pathways (see below). Further tailoring of vectors and hosts could increase utility. Pfeifer and co-worker have recently transferred the mevinolin resistance marker from pWL102 into a plasmid of *Hb. halobium* and have shown that the mevinolin selection can also be used in *Hb. halobium* (Blaseio and Pfeifer 1990).

3. Characterization of mevinolin resistance

Identity of the resistance gene. DNA sequencing of the 3.5 kbp KpnI-SphI fragment, which contains the mevinolin resistance gene in the shuttle vector pWL102 (figure 5B) revealed a 1212 basepair (bp) open reading frame encoding a protein of approximately 4.1x 10^4 MW (figure 7). The deduced amino acid sequence aligns coherently with the C-terminal half of animal and the soluble plant HMG CoA reductases (Chin et al. 1984; Learned and Fink 1989), but lacks the variable N-terminal membrane spanning domains (Basson et al. 1988; Woodward et al. 1988) found in animal and yeast

```
          CTCGAGTCGTCCCACGTGCTCTCGTCGTAGCCGTAGCCGAGCACCCAGCCGGCCTCGTCGCCGGTCT
CGTCGTCGGATTCGACCTCGGCGGCGCGCTCGCGGAGCAGTTCGACCGCCTCCGCCGGGGAGCCGGCGGCGGAGAGGTC
GGCGTCCACGAGCGAGCGGCCGACCATCGGGAGGTGCGTGTGGGCGTCGATGAAGCCGGGGAGGACGACGCGGCCGCCG
CAGTCGACGACCCGCGTCTCGACGCCGGCGAGCAGTCCCACGTCGAACGCGCGGCCGACGCGGACGACCTCGCCGTCGC
GGACGGACGGCTTCGTGCGTCTCGTCGGGCGACCCGAGGGTGTGTACCTCCGCGTTCGTCAGTACGAGGTCCGCTGCCG
CGGTCATACGTTCGACCCCGACGCGGGAGGGCAAAACGGTTCGGGAGGCGGCGGGGTGGCCGAGCGGCGGTCGGTCGTC  462
                                                    G                    ⬇
GCGGGTGACGCCGGCCGGGTCGGTGGCGGCCGTTGGAAGAACCGGAACGCTTTAGGGCCGGCGCGGGAAACGTCGGAGC  541

ATG ACA GAC GCC GCG TCC CTC GCA GAC CGC GTT CGG GAG GGC GAC CTC CGC CTC CAC GAA  601
Met thr asp ala ala ser leu ala asp arg val arg glu gly asp leu arg leu his glu

CTC GAA GCG CAC GCC GAC GCC GAC ACC GCC GCC GAG GCG CGC CGA CTG CTC GTC GAA TCG  661
leu glu ala his ala asp ala asp thr ala ala glu ala arg arg leu leu val glu ser

CAG TCC GGC GCG TCG CTC GAC GCG GTC GGG AAC TAC GGC TTC CCC GCG GAG GCC GCC GAG  721
gln ser gly ala ser leu asp ala val gly asn tyr gly phe pro ala glu ala ala glu

TCC GCC ATC GAG AAC ATG GTC GGC TCG ATC CAG GTG CCG ATG GGC GTC GCC GGC CCC GTC  781
ser ala ile glu asn met val gly ser ile gln val pro met gly val ala gly pro val

AGC GTC GAC GGC GGC TCC GTC GCC GGC GAG AAG TAC CTC CCC CTC GCG ACC ACC GAG GGC  841
ser val asp gly gly ser val ala gly glu lys tyr leu pro leu ala thr thr glu gly

GCG CTC CTC GCG TCG GTC AAC CGC GGT TGC TCG GTC ATC AAC AGC GCC GGC GGC GCG ACC  901
ala leu leu ala ser val asn arg gly cys ser val ile asn ser ala gly gly ala thr

GCC CGC GTC CTC AAG TCC GGG ATG ACC CGC GCG CCG GTG TTC CGC GTC GCC GAC GTT GCC  961
ala arg val leu lys ser gly met thr arg ala pro val phe arg val ala asp val ala

GAG GCC GAG GCG CTC GTC TCG TGG ACC CGC GAC AAC TTC GCG GCG CTG AAG GAG GCC GCG  1021
glu ala glu ala leu val ser trp thr arg asp asn phe ala ala leu lys glu ala ala

GAG GAG ACG ACG AAC CAC GGC GAA CTC CTC GAC GTG ACG CCG TAC GTC GTC GGC AAC TCG  1081
glu glu thr thr asn his gly glu leu leu asp val thr pro tyr val val gly asn ser

GTG TAC CTG CGA TTC CGC TAC GAC ACC AAG GAC GCG ATG GGG ATG AAC ATG GCC ACC ATC  1141
val tyr leu arg phe arg tyr asp thr lys asp ala met gly met asn met ala thr ile

GCC ACC GAG GCC GTC TGC GGC GTC GTC GAA GCC GAG ACG GCC GCC TCG CTC GTC GCC CTC  1201
ala thr glu ala val cys gly val val glu ala glu thr ala ala ser leu val ala leu

TCG GGC AAC CTC TGT TCC GAC AAG AAG CCC GCC GCC ATC AAC GCC GTC GAG GGG CGC GGC  1261
ser gly asn leu cys ser asp lys lys pro ala ala ile asn ala val glu gly arg gly

CGG AGC GTC ACC GCC GAC GTT CGA ATC CCG CGC GAG GTC GTC GAA GAA CGC CTG CAC ACC  1321
arg ser val thr ala asp val arg ile pro arg glu val val glu glu arg leu his thr

ACG CCC GAA CGC GGT CGC GAA CTC AAC ACA CGC AAG AAC CTG GTC GGC TCC GCG AAG GCC  1381
thr pro glu arg gly arg glu leu asn thr arg lys asn leu val gly ser ala lys ala

GCG AGC CTC GGG TTC AAC GCC CAC GTC GCC AAC GTC GTC GCC GCG ATG TTC CTC GCC ACC  1441
ala ser leu gly phe asn ala his val ala asn val val ala ala met phe leu ala thr

GGG CAG GAC GAG GCG CAG GTC GTC GAG GGC GCG AAC GCC ATC ACG ACC GCC GAG GTG CAG  1501
gly gln asp glu ala gln val val glu gly ala asn ala ile thr thr ala glu val gln

GAC GGC GAC CTC TAC GTC TCG GTC TCC ATC GCC TCC CTC GAA GTC GGC ACC GTC GGC GGC  1561
asp gly asp leu tyr val ser val ser ile ala ser leu glu val gly thr val gly gly

GGC ACG AAA CTC CCG ACG CAG TCC GAG GGC CTC GAT ATC CTC GGC GTC AGC GGC GGC GGC  1621
gly thr lys leu pro thr gln ser glu gly leu asp ile leu gly val ser gly gly gly

GAC CCC GCC GGC TCC AAC GCC GAC GCC CTC GCC GAA TGC ATC GCC GTC GGT TCC CTC GCG  1681
asp pro ala gly ser asn ala asp ala leu ala glu cys ile ala val gly ser leu ala

GGC GAA CTC TCC CTT CTC TCC GCG CTC GCC TCG CGG CAC CTC TCC AGC GCC CAC GCG AAC  1741
gly glu leu ser leu leu ser ala leu ala ser arg his leu ser ser ala his ala asn
          ------> <-----
TCC GGT CGG TAACTCGCCGGGTTCATTTTCACCGACAGCAGGGACGGCTGGACTGCCTTTTCTTCGCCGTTGTTCC  1817
ser gly arg OCH

GGTCTGTTCCCTCCGACTTCGACCTCGTCACGTTTCACTTTCACTTTCACTCCGCACCTGTTACGGGTCGCCAGTTTTA  1996
```

Figure 7. DNA sequence of the mevinolin resistance marker and the deduced amino acid sequence of *Hf. volcanii* HMG CoA reductase. Arrow denotes the start of transcription. G to T substitution in the promoter region is indicated.

enzymes (figure 8). Cabrera et al. (1986) have located a salt-dependent HMG CoA reductase activity in the soluble fractions of *Hb. halobacterium* lysates.

Comparison to eukaryotic and eubacterial HMG COA reductases. The archaebacterial sequence is strikingly similar to the catalytic domain (C-terminal half) of eukaryotic HMG CoA reductases, sharing between 42% to 45% identical residues with the corresponding sequences of hamster, yeast and *Arabidopsis*. Therefore these enzymes must share a common origin. The eukaryotic sequences are slightly more similar to each other (about 55% sequence identity in the region shown in figure 8), reflecting the deep divergence between these primary kindoms.

The NADH-dependent HMG CoA reductase of *Pseudomonas mevalonii* (Beach and Rodwell 1989) is the only eubacterial sequence available for comparison. The *Pseudomonas* enzyme bears very limited resemblance to either the eukaryotic or the archaebacterial HMG CoA reductases at the sequence level. Sequence similarity is more recognizable near the region corresponding to the last 75 residues of the archaebacterial enzyme (figure 8). Unlike the eukaryotic and the archaebacterial NADPH-dependent reductases, the primary role for this NAD enzyme from mevalonate-grown *Pseudomonas* is catabolic, converting mevalonic acid to HMG CoA. The *Pseudomonas* reductase shares only 15% sequence identity with the halobacterial enzyme, but over half of the identical residues are conserved among *Hf. volcanii*, *P. mevalonii* and all the eukaryotic examples compared, suggesting that they are true homologs. All but two of the 30 conserved residues are clustered within two regions (figure 8). The clustering may reflect functional importance of these regions.

The eukaryotic enzymes have probably diverged from the archaebacterial enzyme after their common ancestor diverged from the eubacterial HMG CoA reductase. Horizontal gene transfer is unlikely, since such a transfer has to predate the divisions within these primary kingdoms. However, we cannot rule out the possibility that the

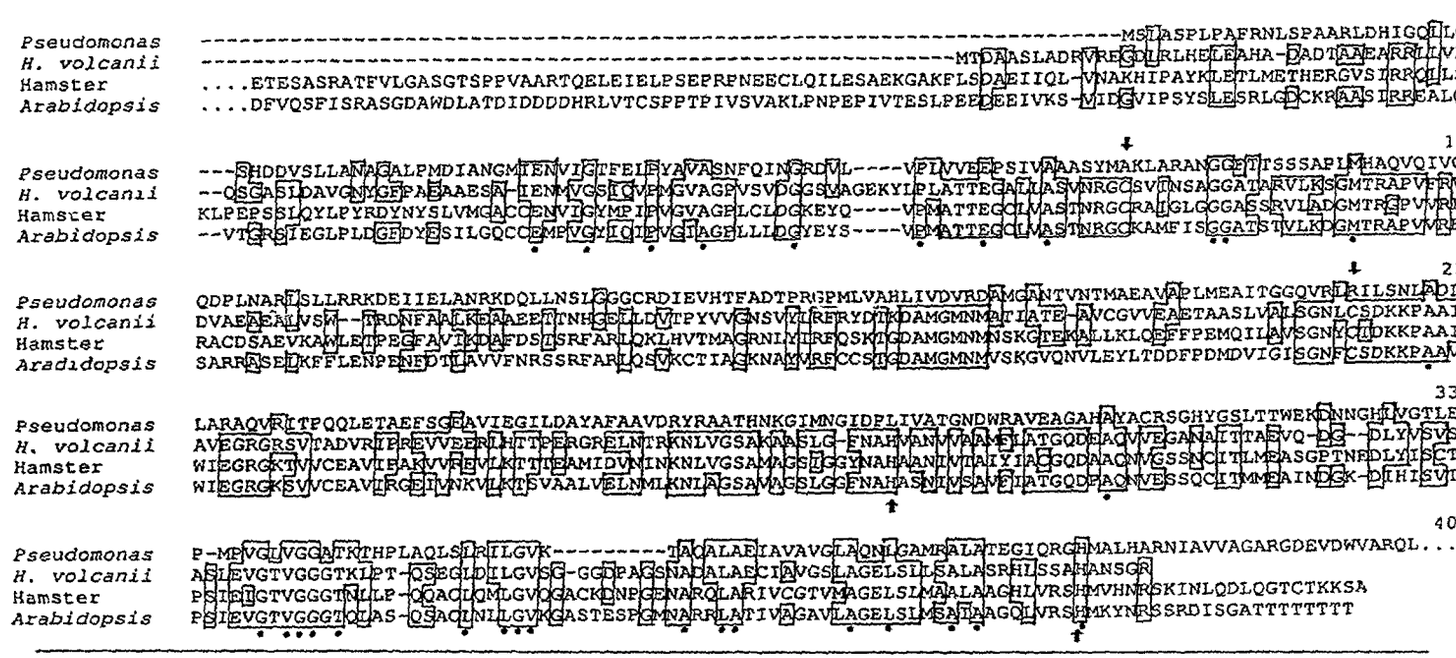


Figure 8. Comparison of HMG CoA reductase sequences. Numbers refer to positions in the *Hf. volcanii* sequence. Residues identical to those found in the *Hf. volcanii* sequence are boxed. (---) indicates gaps introduced to accommodate alignment, while (...) represents sequence not shown. Cys (⬇) and his (⬆) residues conserved in the eukaryotic and the archaebacterial sequences are marked by arrows. Conserved residues are marked by a dot.

Pseudomonas example may be atypical or that there is another eubacterial (paralogous) HMG CoA reductase gene yet to be discovered. The close resemblance between archaebacterial and eukaryotic enzymes is in line with the growing suspicion that the archaebacteria are more closely related to the eukaryotes than to the eubacteria (see *Introduction*), based on the findings of eukaryotic features (such as introns) in archaebacteria, and the similarities in eukaryotic and archaebacterial key metabolic enzymes (for example, H^+ATPase, elongation factors and RNA polymerase subunits; refer to Iwabe et al. 1990 and Puhler et al. 1989).

Residue-specific chemical modification experiments suggest the presence of histidine and cysteine residues in the active site of rat liver and yeast HMG CoA reductases, although the catalytic residues have yet to be identified (Roitelman and Shechter 1989; Dugan and Katiyar 1986). Sequence alignment of HMG CoA reductases from hamster (Chin et al. 1984), human (Luskey and Stevens 1985), sea urchin (Woodward et al. 1988), *Schistosoma* (Rajkovic et al. 1989), *Drosophila* (Gertler et al. 1988) yeast (Basson et al. 1988), and plant (Learned and Fink 1989) show five conserved cysteine and two conserved histidine residues. The *Pseudomonas* enzyme has only 2 cysteine residues; however, and they are not present at the corresponding positions in either the eukaryotic or the archaebacterial enzymes (figure 8). Replacement of these two cysteine residues with alanine residues shows no major effect on enzyme specificity or enzyme activity (Jordon-Starck and Rodwell 1989a, b). With the archaebacterial enzyme, two cysteine and two histidine residues are common to all the eukaryotic sequences compared above. Cys-110, cys-225 and his-288 are situated in conserved stretches of amino acid sequences. His-398, near the C-terminus of the enzyme, is also found in the *Pseudomonas* sequence. Site-direct mutagenesis experiments are required to clarify the role of these and other conserved residues in enzyme activity.

Up-promoter mutation. In an initial attempt to determine whether mevinolin resistance could be due to gene amplification, DNAs isolated from the wild-type (strain DS2) and

from resistant mutants (from which the resistance gene was isolated) were compared in Southern hybridization experiments using the 7.9 kbp MluI fragment containing the resistance-determining gene as hybridization probe (figure 9A). Although no amplification was detected in these first four mutants, unexpectedly the loss of an MluI site was evident. Two MluI fragments (2.9 and 5.0 kbp) in the wild-type DNA appear as a single 7.9 kbp fragment in the mutant DNA. No other obvious rearrangements were detected; both DNA samples indicate a single copy of the HMG CoA reductase gene (figure 9A, B). Is this mutation, destroying the MluI recognition site, responsible for mevinolin resistance?

To answer this question, I isolated the gene for HMG CoA reductase from wild-type (mevinolin sensitive) *Hf. volcanii* cells, using the resistant gene as a probe. The DNA sequence in the coding region of the wild-type gene is identical to that isolated from the resistant mutant. A single base substitution (from G to T, which coincidentally destroys the recognition site for the restriction enzyme MluI), detected 29 bp upstream of the translation initiation site, is apparently responsible for drug resistance (figure 9 and 10).

A TATA like sequence preceeding the transcript initiation site by about 25 bp resembles the archaebacterial consensus promoter described by others, (figure 10; Thomm and Wich 1988; Zillig et al. 1988). The HMG CoA reductase messenger RNA begins at the C residue 6 nt upstream of the translation start site (figure 11), placing the mutation within this putative promoter (figure 10). Quantitative primer extension experiments confirm that this mutation at position -23 is in fact an "up promoter" mutation, increasing the expression of the gene for HMG CoA reductase (figure 11). The cell becomes resistant to mevinolin, presumably, by titrating the inhibitor with excess

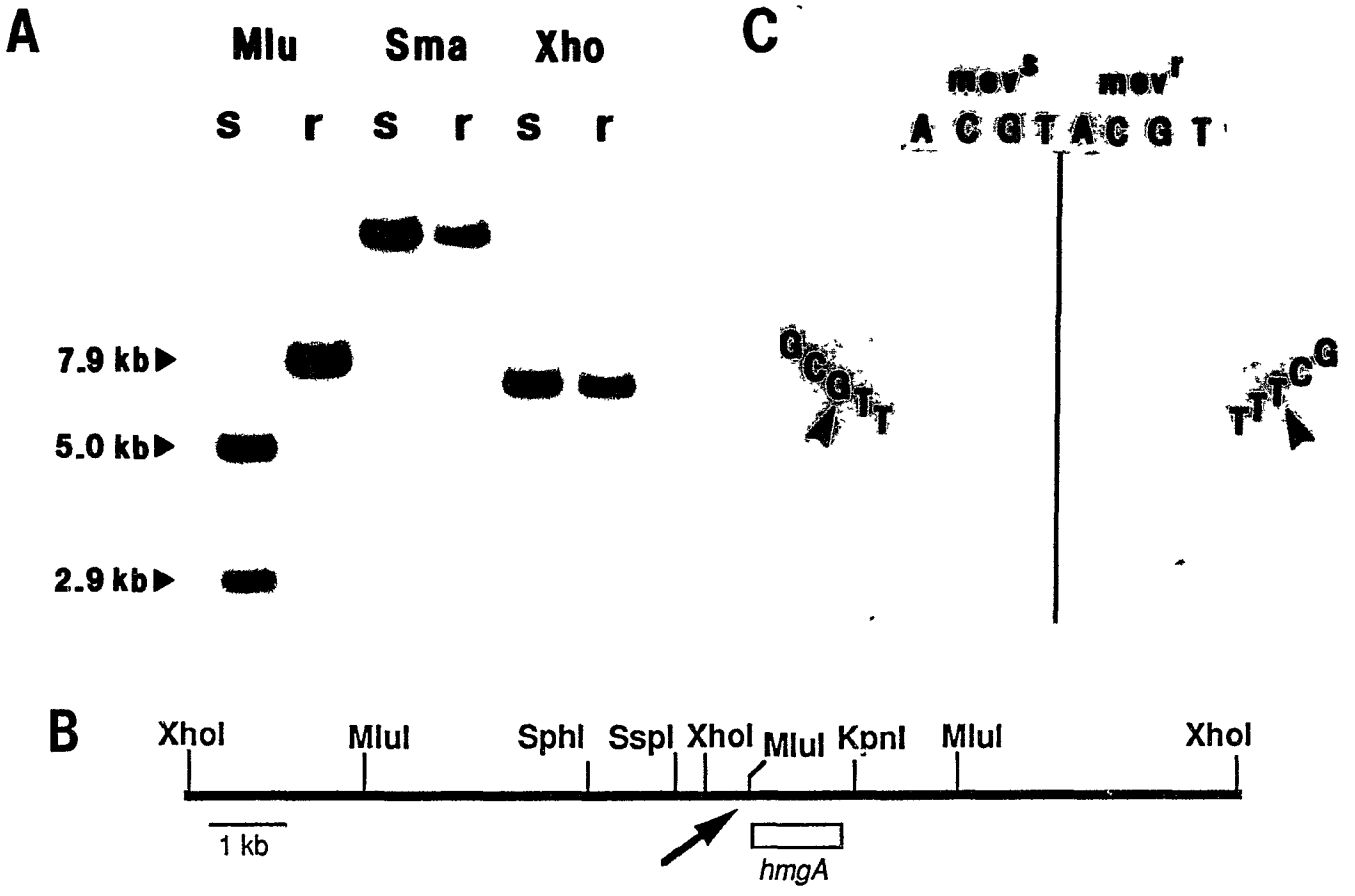


Figure 9. Chromosomal region containing the wildtype and the mutant HMG CoA reductase gene. (A) Restriction digests of genomic DNA from resistant mutants (r, mixture of DNA isolated from 4 mutants, source of DNA for the MluI library used for cloning the mevinolin resistance marker) and from wildtype strain WFD11 (mevinolin sensitive, s), were compared by Southern hybridization technique using the 7.9 kb MluI fragment of pWL2 as probe. (B) Restriction map of the chromosomal region. Arrow points at the MluI site missing in mutants. (C) Comparison of DNA sequence upstream of the wildtype and the resistance gene reveals a G to T base subsitution in the mutant sequence.

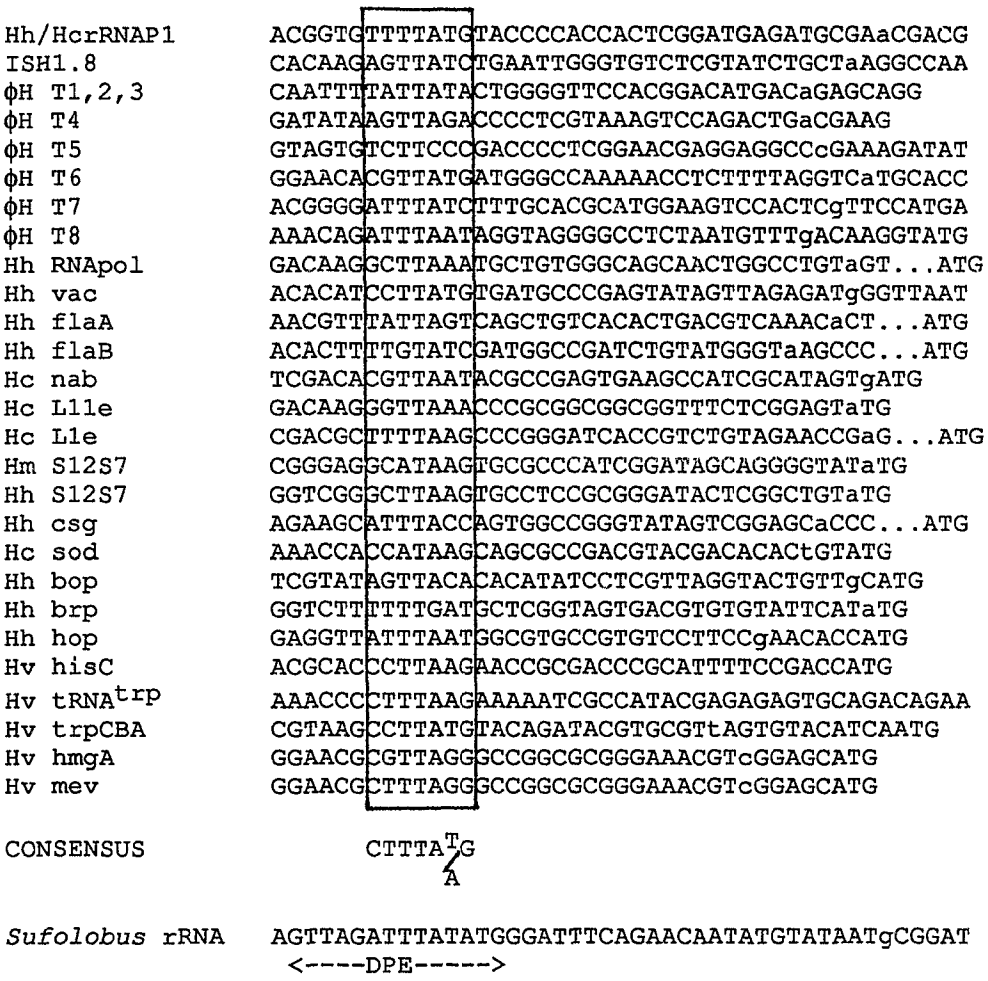


Figure 10. Consensus promoter sequence. Sequences upstream of protein-coding genes of *Halobacteria* were compared. Lower case letter indicates transcription start site. DPE is the distal promoter element upstream of *Sulfolobus* B12 rRNA operon defined by *in vitro* transcription experiments (Reiter et al. 1990).

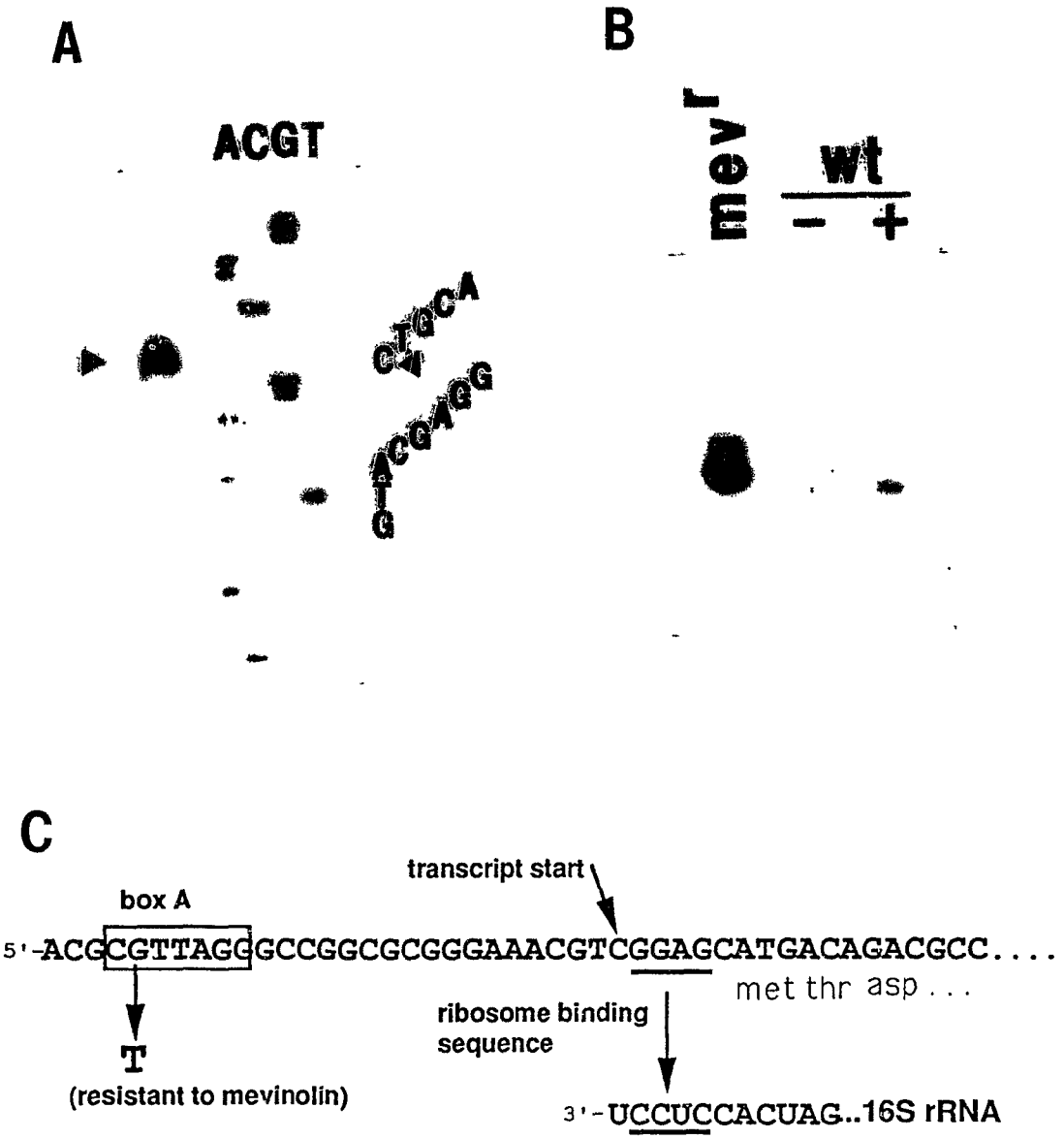


Figure 11. Initiation of the *Hf. volcanii* HMG CoA reductase RNA transcript. (A) Determination of 5'-terminus of the transcript by primer extension using 15 µg of total RNA isolated from a culture of *Hf. volcanii* WFD11 grown on minimal liquid medium. Arrowhead indicates transcription initiation site. (B) The same primer extension assay applied to WFD11 (wt) cultures, grown with (+) and without (-) supplement of 10mM mevalonic acid lactone, and to a up-promoter mutant culture in minimal medium. (C) Summary of features in the sequence upstream of the *hmgA*.

enzyme. Overexpression as a means to overcome mevinolin inhibition has been well documented in mammalian cells. Amplification of the gene for HMG CoA reductase permits Chinese Hamster Ovary cells to grow in the presence of mevinolin (Chin et al. 1984). I have also observed similar amplification mutants of *Hf. volcanii* (see below).

Genetic demonstration of an archaebacterial promoter. The G->T mutation alters the efficiency of transcription initiation and thus locates and functionally defines the archaebacterial promoter *in vivo* -- this is the first genetic demonstration of an archaebacterial promoter. The increase in promoter strength is consistent with the fact that the mutant sequence (TTTAGG) resembles more closely the consensus sequence (Zillig et al. 1988) located at the same distance upstream of the transcript start sites of highly expressed halobacterial genes, such as ribosomal RNA genes of *Hb. halobium* (TTTATG) and the $tRNA^{Trp}$ gene of *Hf. volcanii* (TTTAAG). Furthermore, this mutant promoter corresponds (figure 10), both in sequence and in location, to the distal promoter element (DPE) upstream of the *Sulfolobus B12* ribosomal RNA operon which Reiter and co-workers have recently defined by linker substitutions to influence the efficiency of *in vitro* transcription (Reiter et al. 1990).

Unlike the typical eubacterial promoter, which depends on signals positioned at -35 and at -10 relative to the transcript start site for RNA polymerase binding and transcription initiation (Reznikoff et al. 1985), the A-T rich archaebacterial promoter sequence (especially prominent in G+C rich halophile DNA) is very similar to the A-T rich "TATA box" sequence of eukaryotic promoters for RNA polymerase II, which is also located about 25 bp upstream from the starting point of RNA synthesis (figure 10). Zillig and colleagues (Puhler et al. 1989; Zillig et al. 1988) have repeatedly pointed out the similarities between the archaebacterial and the eukaryotic RNA polymerases.

Gene amplification. A second mechanism leading to mevinolin resistance was discovered when additional resistant mutants were surveyed for rearrangement in regions flanking the gene for HMG CoA reductase.

Hf. volcanii DS2 and its derivatives WFD11, T9 and T11 are sensitive to mevinolin and all display the wild-type two banded hybridization pattern -- the 7.9 kbp MluI fragment containing the resistant gene hybridizes to a 2.9 kbp and a 5.0 kbp fragment in an MluI digest of wild-type DNA (see figure 9A above). Surprisingly, DNA isolated from various resistant mutants originating from these four sensitive parental strains exhibited two types of hybridization patterns, only one of which could be explained by the loss of an MluI site to give a single signal of 7.9 kbp. The second pattern looks more complex, but can be explained simply. Tandem amplification of a specific segment of the chromosome containing the gene for HMG CoA reductase could account for the intense hybridization signal not found in the wild-type DNA (figure 12), as well as for retention of 2.9 and 5.0 kbp signals. In some mutants, the amplified segment encompasses the 2.9 kbp and the 5.0 kbp MluI fragments entirely. Different segments of DNA have been amplified in each mutant.

This is not the first example of gene amplification in *Hf. volcanii.* Mevarech and coworkers (Rosenshine etal., 1987) have correlated resistance to dihydrofolate reductase inhibitors to the amplification of specific DNA sequences. Gene amplification is also responsible for mevinolin resistance in Chinese Hamster ovary cells (Chin et al. 1984).

In summary, *Hf. volcanii* utilizes two independent mechanisms to escape growth inhibition by mevinolin. An up-promoter mutation, which also provided the first genetic evidence of an archaebacterial promoter, and the amplification of the target gene both lead to mevinolin resistance via overexpression of HMG CoA reductase. I have found several occurrences of each type of mutation, although only one has been characterized at the sequence level.

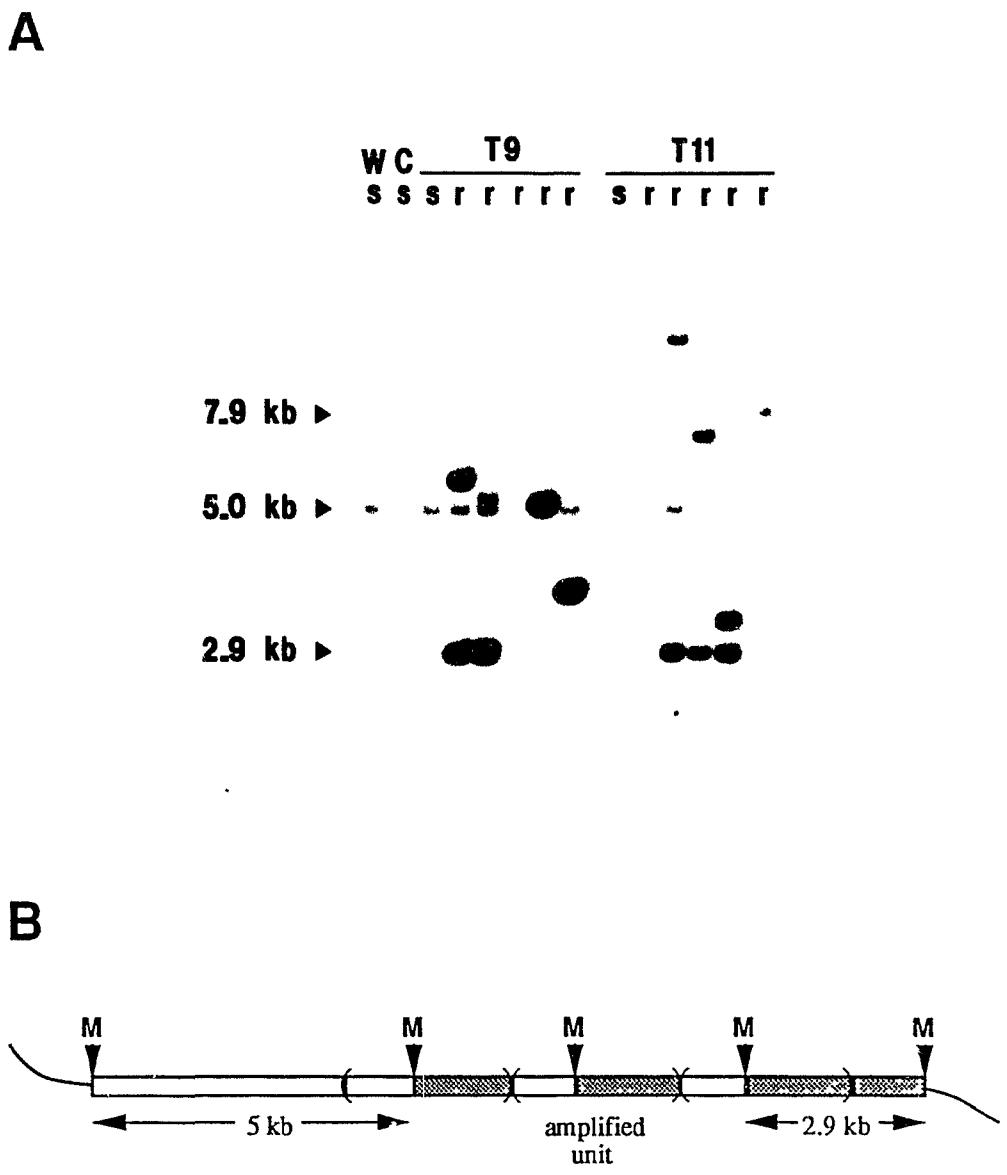


Figure 12. Characterization of mevinolin resistant mutants of *Hf. volcanii*. (A) shows hybridization of ^{32}P-labeled 7.9 kb mevinolin resistance marker to MluI-digested DNA samples, isolated from four mevinolin sensitive (s) strains (W, wildtype strain DS2; C, cured strain, WFD11; T9 and T11, two tryptophan auxotrophic strains) and ten resistant (r) mutant strains generated from T9 and T11. (B) shows a schematic representation of gene amplification. M represents MluI sites. DNA derived from the 2.9 kb MluI fragment is shaded.

Gene expression. I undertook some preliminary experiments to examine the regulation of expression of the HMG CoA reductase gene. Transcription of this gene appears to be insensitive to exogenous mevalonic acid (the product of the reaction catalyzed by the reductase). Total RNA isolated from a WFD11 culture grown with 10 mM of mevalonic acid lactone gave a signal, in primer extension experiments, no less intense than the signal obtained from mevalonate-starved cells (figure 11). (Mevelonic acid at this concentration should prevent mevinolin growth inhibition; see figure 3B above.) This is consistent with the findings of Cabrera et al. (1986) who reported that exogenous mevalonate did not suppress *Hb. halobium* reductase activity over a wide range of concentrations. It is premature to conclude constitutive expression since regulation by more complex mevalonate-derived products has not been tested. Mevalonate-derived products (like cholesterol) are responsible for feedback inhibition in mammals (Brown and Goldstein, 1980; Goldstein and Brown 1990).

The biochemical pathway for mevalonic acid biosynthesis in halobacteria is unclear. Ekiel et al. (1986) proposed that biosynthesis in *Halobacterium* involves carbon atoms from lysine and the pathway may be different from that of the eukaryotes. However, the same group found that the lipid labeling pattern in another archaebacterium, *Methanospirillum hungatei*, is consistent with the pattern expected from mevalonic acid synthesized by the standard eukaryotic pathway (Ekiel et al. 1983).

Annalee Cohen (of our laboratory) and I are trying to isolate mutants in the early steps of the pathway. Since exogenous mevalonate has been shown to be incorperated into the lipids of *Halobacterium* (Kates et al. 1968; Cabrera et al. 1986), we may be able to generate mevalonate auxotrophs. In addition, mammalian HMG CoA synthase inhibitors (if they work on halophiles) could be used for generating resistant mutants in *Hf. volcanii*, and for isolating the gene using the same strategy employed to clone the HMG CoA reductase gene.

4. Host strain development

Integration of a non-replicating selectable marker into the *Hf. volcanii* chromosome has been observed in transformation experiments (using deleted forms of pWL101, such as those designated ΔBD and ΔBSpe in figure 6). Such a recombination events should allow the application of gene replacement techniques, which have been proven useful in targeting genes and in introducing *in vitro*-manufactured mutations into yeast chromosomes, to *Hf. volcanii.* This section describes experiments which demonstrate that gene replacement techniques can, in fact, be applied to *Hf. volcanii.* Useful strains, which will aid detailed dissection of biological functions like tryptophan biosynthesis, are still under construction.

Gene replacement. I set out to replace the chromosomal *trpB* gene with a non-replicatable mevinolin resistance (*mev*^{r}) marker. An artificial construct (pmevΔtrp), built by fusing the cloned *mev*^{r} gene to the *trpB* flanking sequences (figure 13), was linearized with PstI, and used to transform *Hf. volcanii* WFD11 (which is Trp$^+$Mev^S). Mevinolin resistant transformants recovered were maintained on rich medium containing mevinolin, and then tested for tryptophan requirement. Trp$^-$Mev^r transformants should reflect recombination (via two crossover events) at the regions flanking the *trp* locus, resulting in the replacement of the *trpB* gene on the chromosome by the *mev*^{r} gene from the artificial construction (figure 13). Southern analysis of genomic DNA isolated from Trp$^-$Mev^r transformants verified the occurrence of this replacement event. The 8 kbp *trpB*-containing EcoRI fragment on the chromosome of WFD11 has changed, as expected, in seven of the eight Trp$^-$Mev^r transformants analysed, to become a 10 kbp fragment (figure 14A) which also hybridizes to the *mev*^{r} gene (figure 14B).

A circular pmevΔtrp should integrate via a single crossover event into the chromosome at either the *mev* locus or at the regions flanking the *trpB* locus of interest, to give mevinolin resistant transformants. The complex integration structure

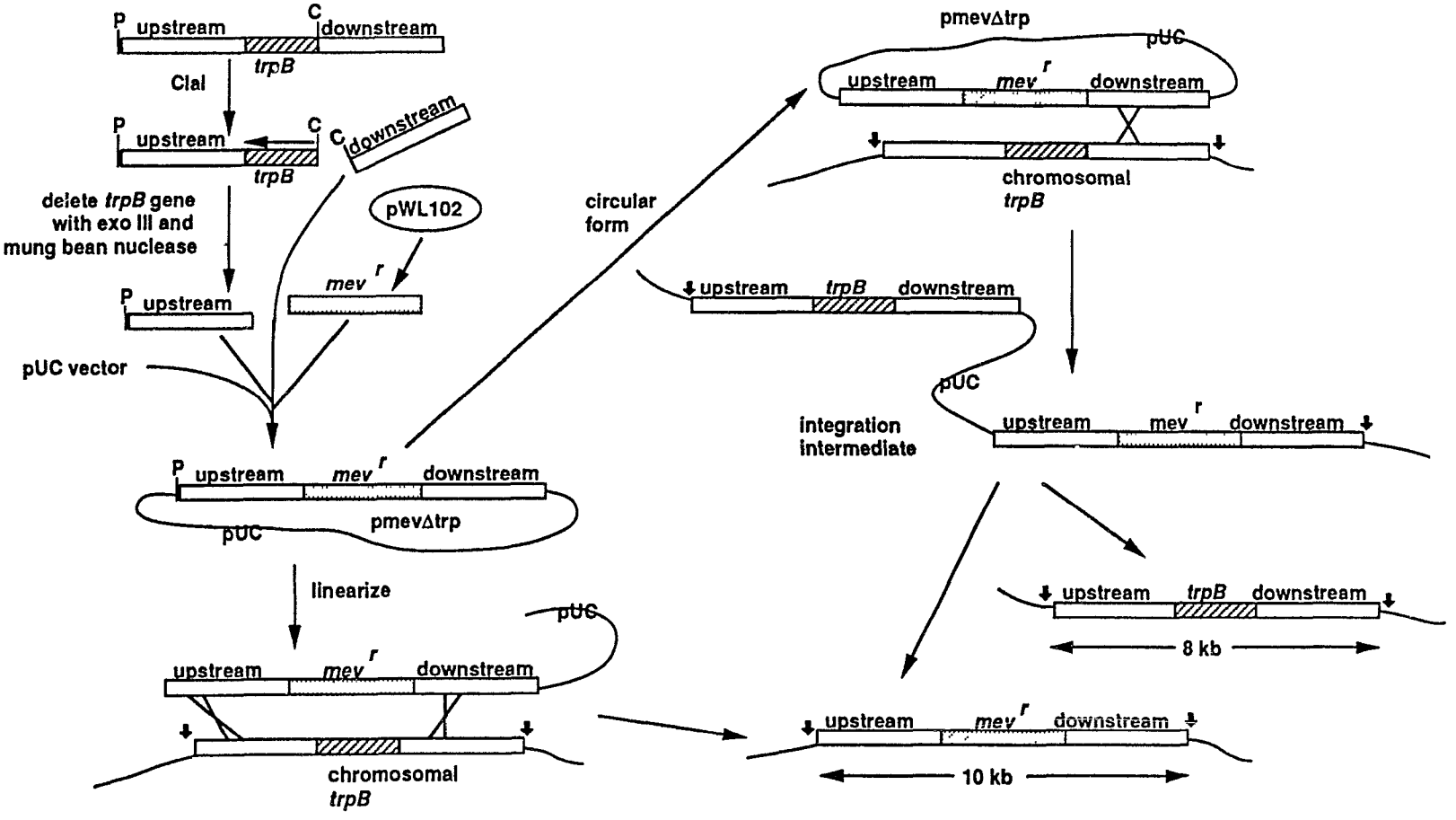
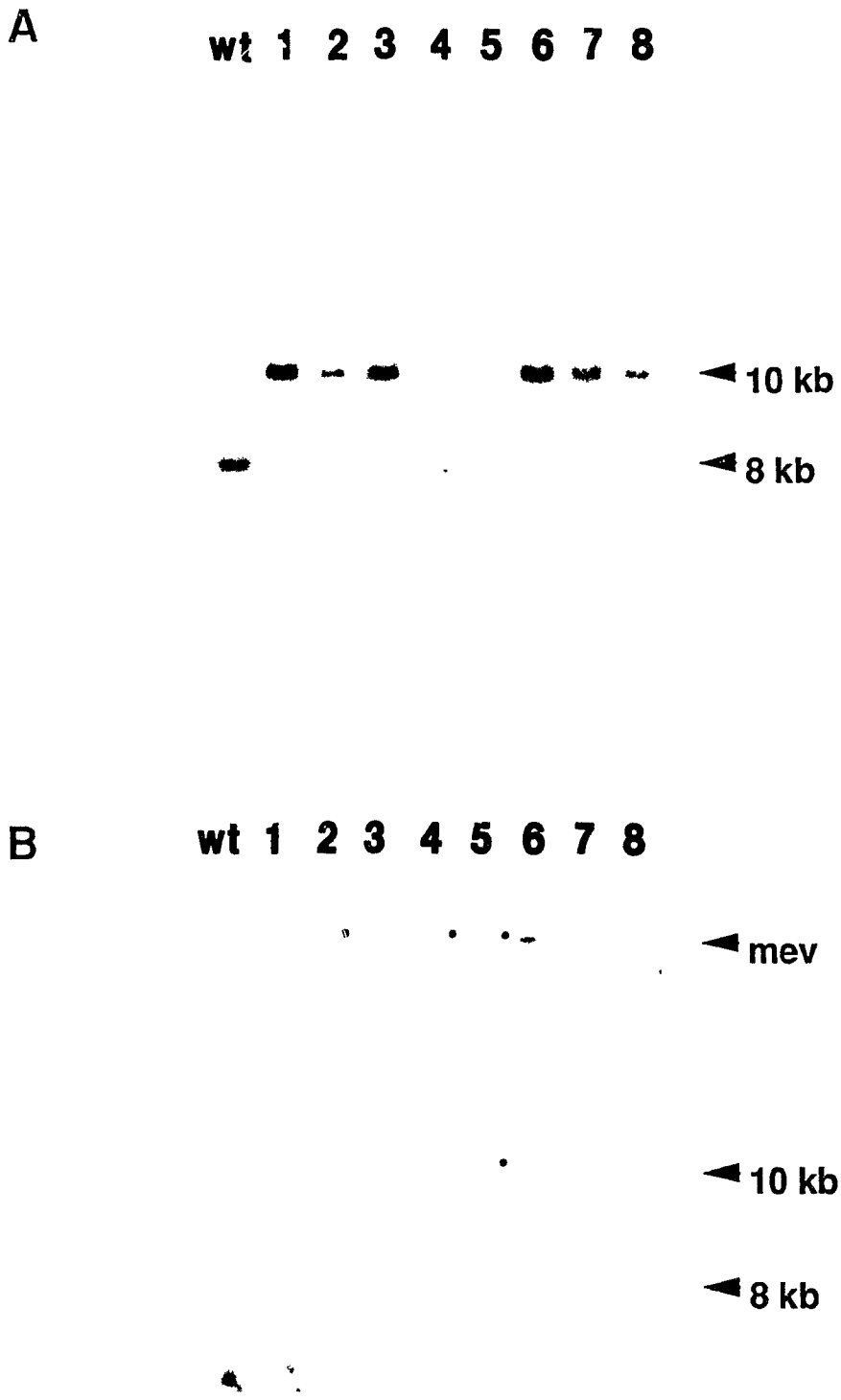
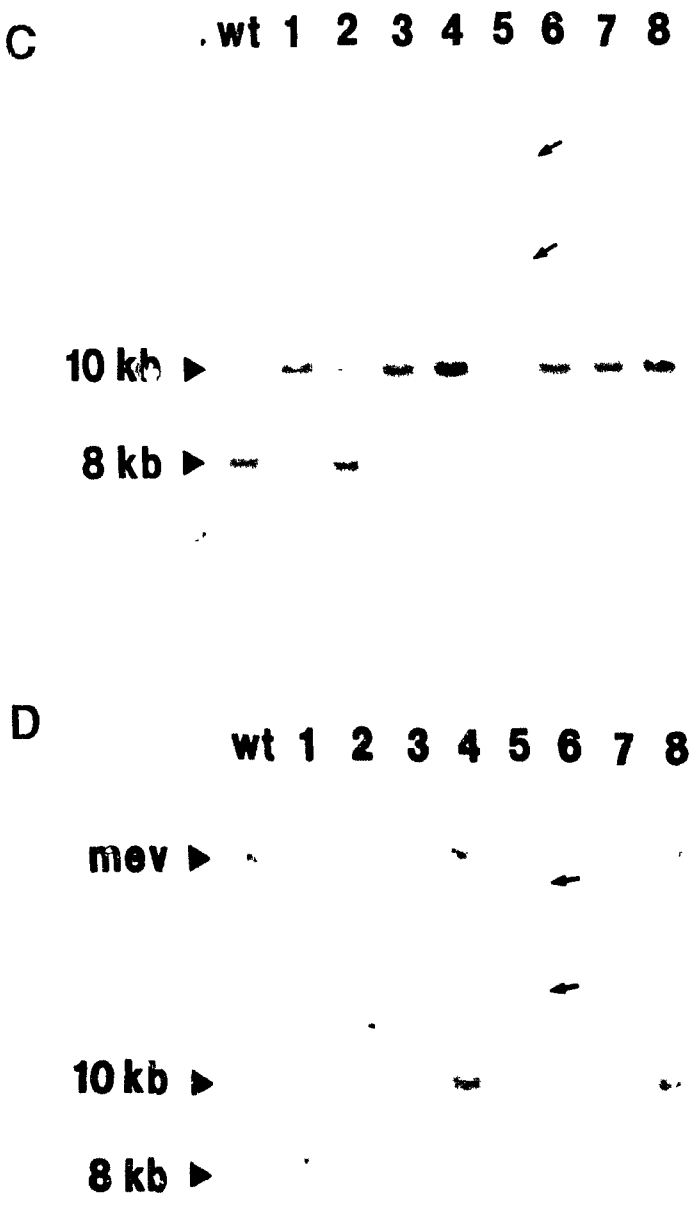


Figure 13. Schematic representation of the events leading to the replacement of the chromosomal *trpB* by the *mev* marker on linear or circular pmevΔtrp. (⬇) represents EcoRI sites. P stands for PstI, and C for ClaI.

Figure 14. Characterization of mevinolin resistant strains resulting from transformation of *Hf. volcanii* WFD11 with linear or circular pmevΔtrp. Construction of pmevΔtrp, and its use in gene replacement experiments, are summarized in Figure 13. EcoRI digestion of DNA samples isolated from WFD11 (wt) and 16 mevinolin resistant transformants (8 from transformation with linearized pmevΔtrp shown in **A** and **B**, and 8 from circular pmevΔtrp transformation shown in **C** and **D**) were analysed by Southern hybridization. **A** and **C** were probed with the 8 kb *trpB*-containing EcoRI fragment. **B** and **D** (blots identical to those used in **A** and **C**, respectively) were probed with the 3.5 kb mevinolin resistance marker (*mev*) from pWL102. Dots in **B** indicate the position of hybridization signals visible after longer exposure. Transformant 5 in **C** and **D** may have pmevΔtrp integrated at the *mev* locus. The 2 fragments, indicated by thin arrows, hybridize to both the *mev* and the *trp* probe, and futhermore, the signal for the wild-type *mev*-containing EcoRI fragment is missing from transformant 5.


A
wt 1 2 3 4 5 6 7 8
10 kb
8 kb
B
wt 1 2 3 4 5 6 7 8
mev
10 kb
8 kb

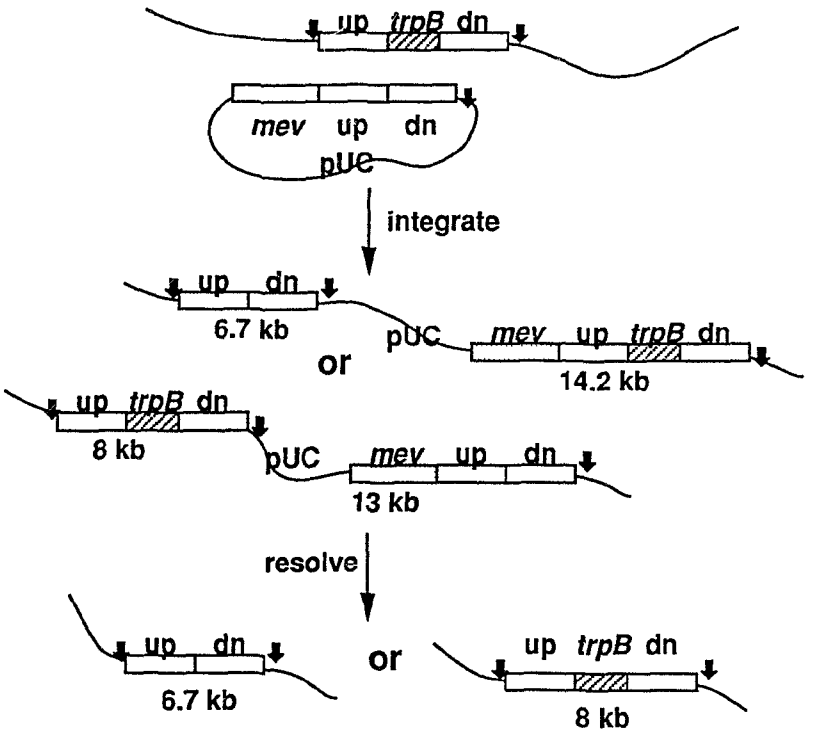

C
wt 1 2 3 4 5 6 7 8
10 kb ▶
8 kb ▶
D
wt 1 2 3 4 5 6 7 8
mev ▶
10 kb ▶
8 kb ▶

encompassing the *trp* locus should subsequently resolve via a second crossover event, either to give back the parental arrangement or to complete the replacement of *trpB* with *mev*^{r} (figure 13). Mevinolin resistant transformants, recovered from transformation experiments using circular pmevΔtrp, were allowed to grow in rich medium without mevinolin. DNA samples isolated from such transformants exhibit hybridization patterns consistent with those expected from the size of the resolution products (figure 14C and D).

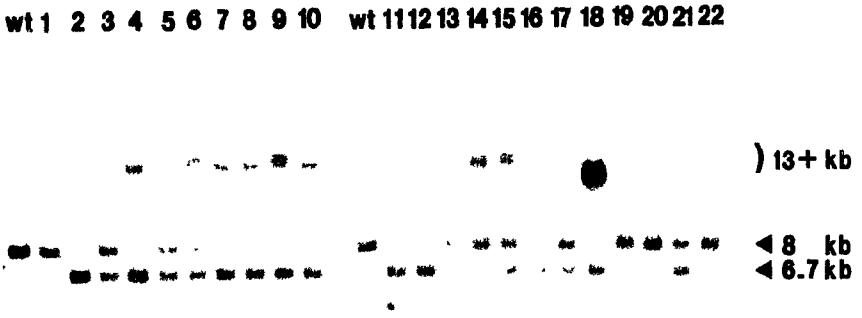
Deletion of *trpB* gene from chromosome. M. Mevarech suggested that the integration and resolution process could be applied to delete specific segments of DNA from the chromosome. The strategy is summarized in figure 15A. *Hf. volcanii* WFD11 was transformed with a non-replicatable circular plasmid construct (containing the *mev*^{r} marker, *trpB* upstream and *trpB* downstream sequences, in this order). Mevinolin resistant transformants were picked onto rich agar medium and allowed to grow without selection. DNA samples prepared from 22 of such transformants were digested with EcoRI, and compared to digested DNA from strain WFD11 using the 8 kbp EcoRI *trpB* containing fragment as hybridization probe. Integration of the construct by recombination between the plasmid-borne and the chromosomal *trpB* upstream regions results in 2 EcoRI fragments of 6.7 and 14.2 kbp in size. Integration at the *trpB* downstream region results in EcoRI fragments of 8 and 13 kbp. Resolution of integrated intermediates (via homologous recombination between *trpB* flanking sequences) could give either an 8 kbp fragment identical to that on the wild-type chromosome or a 6.7 kbp fragment representing the deletion of *trpB* from the chromosome (figure 15). The 22 transformants analysed show the same hybridization pattern, except that the relative intensities of hybridization signals are different for each mutant -- differences presumably represent varying degrees of resolution. The same strategy is being applied to generate a strain

Figure 15. Deletion of the *trpB* gene from the *Hf. volcanii* chromosome. (**A**) schematically describes the predicted sequence of events after introducing a non-replicatable circular plasmid construct (containing the *mev*^{r} marker, *trpB* upstream and *trpB* downstream sequences, in this order), into *Hf. volcanii* WFD11 spheroplasts. Integration of the construct by recombination between the plasmid-borne and the chromosomal *trpB* upstream (up) regions results in 2 EcoRI fragments of 6.7 and 14.2 kbp in size. Integration at the *trpB* downstream (dn) region results in EcoRI fragments of 8 and 13 kbp. Resolution of integrated intermediates (via homologous recombination between *trpB* flanking sequences) could give either an 8 kbp fragment identical to that on the wild-type chromosome or a 6.7 kbp fragment representing the deletion of *trpB* from the chromosome. Small vertical arrow (⬇) denotes EcoRI sites. (**B**) shows the result of Southern analysis of the resultant mevinolin resistant transformants. Mevinolin resistant transformants were picked onto rich agar medium and allowed to grow without selection. DNA samples prepared from 22 of such transformants were digested with EcoRI, and compared to digested DNA from strain WFD11 using the 8 kbp EcoRI *trpB* containing fragment as hybridization probe.

A


up trpB dn
mev up dn
pUC
integrate
up dn
6.7 kb
pUC
mev up trpB dn
14.2 kb
or
up trpB dn
8 kb
pUC
mev up dn
13 kb
resolve
up dn
6.7 kb
or
up trpB dn
8 kb

B


wt 1 2 3 4 5 6 7 8 9 10 wt 11 12 13 14 15 16 17 18 19 20 21 22
) 13+ kb
◀ 8 kb
◀ 6.7 kb

lacking *trpCBA* genes. Precise deletion of the *trpCBA* sequence is necessary to avoid disruption of neighbouring genes.

II. Tryptophan biosynthetic genes of *Hf. volcanii*

I used tryptophan (*trp*) biosynthesis as a model system to test and to refine the genetic tools developed for *Hf. volcanii*. The *trp* system was chosen because of the absolute conservation in the biochemical pathway and the tremendous variability in gene organization and regulation. All organisms capable of synthesizing tryptophan use seven identical enzymatic functions in the same biochemical pathway (figure 16), and the genes encoding these functions have been extensively characterized in a variety of organisms, both eukaryotic and eubacterial (Yanofsky 1984; Hutter and Niederberger 1986; Yanofsky and Crawford 1987). Crawford, in his recent reviews (Crawford 1989; Crawford and Milkman 1990), listed and thoroughly compared a collection of completed gene or protein sequences for each component polypeptide (E,G,D,C,F,B,A) catalyzing the five steps of the *trp* pathway. Among archaebacteria, the *Methanococcus voltae trpB*, *trpA* and the C-terminus of *trpF* coding regions have been sequenced using a DNA fragment cloned by virtue of its ability to complement *E. coli* auxotrophs (Sibold and Henriquet 1988).

Trp genes are arranged very differently in various organisms (Crawford 1989; Yanofsky 1984). Many linkage and fusion patterns exist. *TrpA* and *trpB*, which encode the two subunits of tryptophan synthase, are fused in the fungal *trp(AB)*, but remain as separate genes and linked to *trpF* in most eubacteria and *Methanococcus*, in a *trp*F·B·A transcription unit (dots indicate cotranscription, genes enclosed in parentheses are fused). Fusion of *trp(EG)*, *(GD)*, *(GC)*, *(CF)* and *(GCF)* have also been observed. In *Escherichia coli*, the *trp* genes exist in a single operon, *trpE·(GD)·(CF)·B·A*. In contrast,

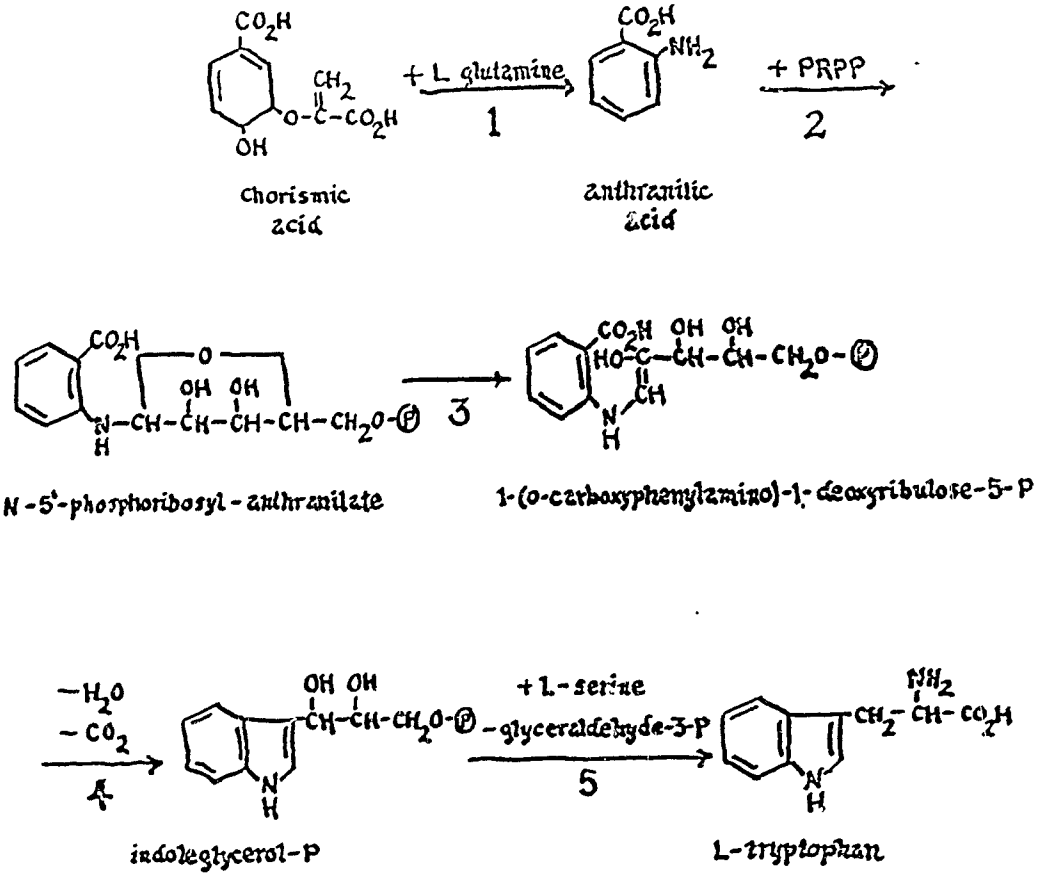


Figure 16. Biochemical pathway for tryptophan biosynthesis. This figure is taken directly from the 1968 Ph.D. dissertation of W.Ford Doolittle, with his permission. Step 1 is catalyzed by anthranilate synthase (TrpE, TrpG); step 2 requires 5-phosphoribosyl-1-pyrophosphate (PRPP), this step is catalyzed by anthranilate phosphoribosyl transferase (TrpD); step 3 by phosphoribosyl anthranilate isomerase (TrpF); step4 by indoleglycerol phosphate synthetase (TrpC) and step 5 by tryptophan synthase (TrpB + TrpA).

Pseudomonas aeruginosa trp genes are organized in four unlinked transcriptional units with no gene fusion, *trpE*, *trpG·D·C*, *trpF* and *trpB·A*. With *Hf. volcanii*, direct genetic methods allow the generation and the mapping of Trp mutations to two unlinked clusters, and the isolation of DNA fragments containing genes for all seven functions in tryptophan biosynthesis (see below). DNA sequencing confirms the functional assignments and reveals an organization of genes within these clusters unlike that in eukaryotes, eubacteria or *Methanococcus*.

1. Isolation of *trp* genes

Isolation and identification of mutants. Auxotrophic mutants of *Hf. volcanii* WFD11 were generated by treatment with EMS. Conditions to achieve >99% killing were established by varying the duration of of EMS treatment (see *Materials and Methods*). Incubation with 10 µl of liquid EMS per ml of cell suspension for 2 hours at 37°C reliably generates one auxotroph out of about 500 surviving colonies analyzed. Specific nutritional requirements were identified by streaking on plates supplemented with various combinations of amino acids and nucleotides. Mutants utilizing tryptophan, leucine, histidine, arginine, asparagine or glutamine were obtained.

Twenty-nine tryptophan-utilizing mutants were identified in three different mutant hunts (including a mutagenesis experiment performed by Annalee Cohen of our laboratory, which produced over 400 auxotrophs; see genetic mapping section below). These Trp$^-$ mutants were further categorized by their ability to grow on media containing anthranilic acid, indole or tryptophan (Table 3).

Cloning of *trp* genes. In initial experiments, genes of tryptophan biosynthesis were cloned by complementation of Trp auxotrophs with shotgun libraries of wild-type DNA in vectors derived from pHV2. Initially, two auxotrophs that would grow only on tryptophan, WFD35 and WFD36, were rescued with Eco RI-digested wild-type DNA

Table 3. Auxotrophic strains, their nutritional requirement, and the plasmids or cosmids which transform them to prototrophy.

WFD strain	allele number	use	plasmid clone(s)	cosmid(s) identified**
35	*trp*-101	T	pT9-42	452
36	*trp*-102	T	pT11-E60	452
37	*trp*-103	A,I	pT1323-n	A159
39	*trp*-105	T	pDWT1	452
41	*trp*-106	T	pDWT1	452
131*	*trp*-107	T	pDWT1	452
135	*trp*-108	I	pDWT1	452
163	*trp*-109	I	-------	A159
167	*trp*-110	I	-------	A159
178	*trp*-111	I	pT178-n	A159
186	*trp*-112	I	pT186-n	A159
192	*trp*-113	I	pT192-n	A159
204	*trp*-114	I	pT204-3	A159
207	*trp*-115	I	pT207-n	A159
216	*trp*-116	I	pT216-n	A159
240	*trp*-117	I	pT240-n	A159
246	*trp*-118	I	pT246-5	488,A159
258	*trp*-119	I	pT258-n	A159
261*	*trp*-120	I	pT261-n	488, A159
284	*trp*-121	I	-------	A159
287*	*trp*-122	T	pDWT1	452
292	*trp*-123	I	pT292-n	A159
300*	*trp*-124	I	pT300-n	488,A159
347	*trp*-125	T	pDWT1	452
369	*trp*-131	T	pDWT1	452
531*	*trp*-127	I	pT31-n	488, A159
543	*trp*-128	T	pDWT1	452
564	*trp*-129	I	pT64-n	A159
567	*trp*-130	I	pT67-n	A159

The indole-utilizing WFD135 and all the trp(T)-requiring mutants were complemented by pDWT1, whil . the rest of the mutants which grow on indole (I) or anthranilic acid (A) were rescued by pT240-1, pT186-4 and pT1323-15.

* Strains initially transformed with cosmid pools.

** All strains were tested for transformation with cosmids A159 and 452.

cloned into compatible sites in the ISH51 sequences of pHV51. (pHV51 is the variant of pHV2 containing a copy of the insertion sequence ISH51 described above.) Plasmid DNAs isolated from a number of independent Trp^+ transformants of either recipient contained a common 8 kbp insert. Two such plasmids (pT9E42, from transformation of WFD35, and pT11E60, from transformation of WFD36) could be used to transform either strain to prototrophy (6 to 8 thousand times more efficiently than linear wild-type DNA), and indeed all Trp^- strains which can grow only on tryptophan could be rescued with these plasmids. Plasmid pT11E60 was made into a hybrid vector, pDWT1, maintainable in *E. coli*, by insertion of the pBR322 derivative pAT153 into its unique HindIII site. pDWT1 is able to rescue one of the twenty indole- or anthranilate-utilizing mutants (WFD135 in Table 3).

The more recently developed shuttle vector pWL102 was used in similar cloning experiments with those mutants which could grow on indole or anthranilic acid. Wild-type DNA partially-digested with MspI or HinPI was separately fractionated on sucrose gradients. DNA fragments 4 to 7 kbp in size were purified and ligated into the unique Cla I site of pWL102. The shuttle vector DNA was prepared from *Hf. volcanii*, to protect ligated DNA from restriction when transformed into *Hf. volcanii* auxotrophs. Recombinant plasmids isolated from prototrophic, mevinolin-resistant transformants shared some common restriction sites and could be used to define a 7 kbp region in which these mutations lie (figure 17A). In fact, three plasmids (pT1323-15, pT240-1 and pT186-4) which contain no extra sequences ouside this predicted region, could transform 19 of the 20 indole- or anthranilate-utilizing mutants to Trp^+ (all except WFD135 which is complemented by pDWT1 above), placing the gene for at least one component of the anthranilate synthase, and at least one other *trp* gene.

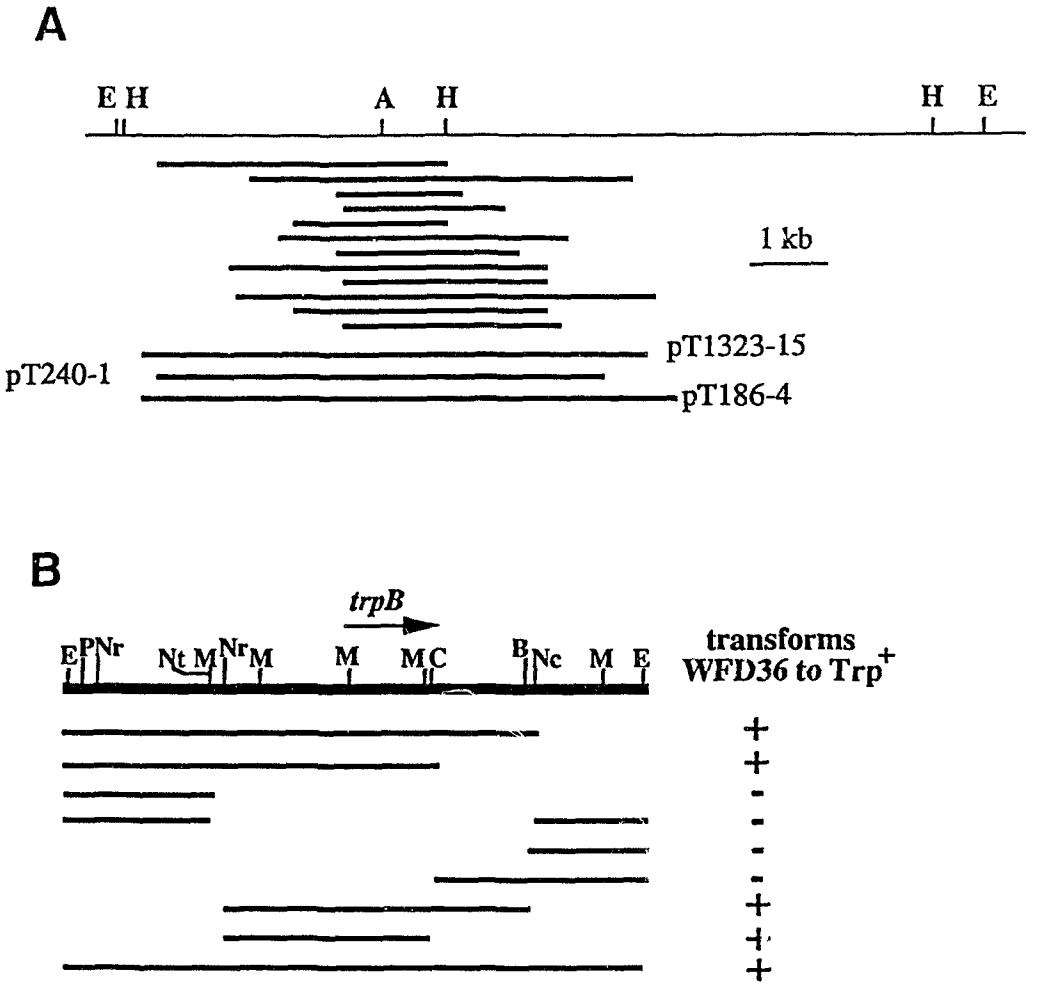


Figure 17. Delimitation of *trp* genes. In (A), the restriction maps of plasmid clones (isolated via transformation of individual mutants with a library of wild-type DNA cloned into pWL102) were aligned to determine the minimal region required for complementation. DNA inserts (MspI or HinPI partial digests) cloned in plasmids pT1323-15, pT240-1 and pT186-4 could complement 1 anthranilate-utilizing and 18 indole-utilizing mutants. The remaining 1 indole-utilizing and 9 tryptophan-requiring mutants could be rescued by pDWT1. (B) Deletion analysis of the 8 kb insert in pDWT1 localizes the *trpB* gene, since mutants with an absolute requirement for tryptophan are defective in the β subunit of tryptophan synthase which is encoded by *trpB*.

2. Analyses of *trp* genes

***trpCBA* gene cluster.** The ability to transform *Hf. volcanii* auxotrophs with small DNA fragments allows more precise localization of genes within plasmid or cosmid clones. The EcoRI fragment which was cloned into pDWT1 is able to rescue all mutants absolutely requiring tryptophan and the indole-utilizing WFD135, thus placing the *trpB* gene and at least one other *trp* gene within this fragment (Table 3). Using subfragments of this 8 kb insert to transform WFD36 allowed the identification of a 2.8 kbp *trpB*-containing region (figure 17B). This region also transforms WFD135 to Trp^+.

The nucleotide sequence of this and neighbouring regions shows three open reading frames (ORFs) transcribed in the same direction (figure 18). These ORFs were named *trpC* (756 bp), *trpB* (1272 bp) and *trpA* (834 bp), based on their amino acid sequence similarities to eubacterial and yeast homologs (figure 19 and Table 4).

The deduced amino acid sequence of *Hf. volcanii* TrpB is about 45% identical to homologs in eubacteria, methanogen and yeast (Table 4). The 423 residue *Hf. volcanii* TrpB sequence, although significantly larger than that of *E. coli* (397 a.a.), aligns coherently with eubacterial tryptophan synthase β subunit sequences (Yanofsky et al. 1981) and with the β chain sequence derived from part of the yeast *TRP5* gene (Zalkin and Yanofsky 1982) (figure 19). Extra residues at the C-terminus account for some of the size difference between the eubacterial and the *Hf. volcanii* TrpB. Three insertions (of 2, 6 and 8 residues at positions 40, 263 and 367 of the *E. coli* sequence) introduced to optimize the sequence alignment, are found in regions predicted by the *Salmonella* three-dimensional structure (Hyde et al. 1988) to be random coils. Half of these 16 inserted residues are acidic amino acids -- not surprising, since most proteins in halophilic archaebacteria are acidic (Eisenberg and Wachtel 1987). The nine extra residues found in *M. voltae* TrpB (at position 245 of *E. coli*) are not present in *Hf. volcanii*.

```
....G GTC GGT CGG GAC GGT CAG TCG ATG GGG ACC GAC GAG AGG TCC GTC TGG CCC  52
.... val gly arg asp gly gln ser met gly thr asp glu arg ser val trp pro

CGG TCG ATG GCG GCC AAC ACG GCG TCG AGG TGG GCG TGG TCC GAC TCG GAG TCC GCC GGC  112
arg ser met ala ala asn thr ala ser arg trp ala trp ser asp ser glu ser ala gly

AGC GAG TCC GGA AAC GAG ACG TCG ATG ACC TTC CCG CTG ACC ACG CCG ACC TGC ACG GGG  172
ser glu ser gly asn glu thr ser met thr phe pro leu thr thr pro thr cys thr gly

CGG CCG AGC GTG GAC GAT TCG CGC GCG TAGATTCCCGAGACGGGCATACGTTCGGCTCCCGTCGCCGGAC  242
arg pro ser val asp asp ser arg ala AMB

GCTTGACCCTTCCGGCGACGGCGTGCGGTTGGGGTGACTGCGCTCGGTTCGGGTGGCCGATACCTCGCCGTGCGGCTCTC  321

GCCCGCGCGCACGCGAGAGACGGTGAGTGTCCCCGCCGACGTAAGCCTTATGTACAGATACGTGCGTTAGTGTACATCA  400
trpC->
ATG AAC GCT AGT GGA GAC GAA TTA GCT CCC GAC GTG CGC GCC ATC TTG GAG GCC GCG CGG  460
Met asn ala ser gly asp glu leu ala pro asp val arg ala ile leu glu ala ala arg

GAG CGA CCC GGC GGG GAG ACG CGC GTG TCG GTC GAC GCG CGG TCG TTC CCC GAG GCC GTG  520
glu arg pro gly gly glu thr arg val ser val asp ala arg ser phe pro glu ala val

GCC GAG ACG GAG GCC GCG GGT CGG GTC CCC GTC ATC GCC GAG GTG AAG CCC ACG AGT CCG  580
ala glu thr glu ala ala gly arg val pro val ile ala glu val lys pro thr ser pro

ACG ACC GAG GGC GTC CGC GAG GAC GAC CCG GTC GAA CTG GCC CGC GAG ATG GTC GCC GGC  640
thr thr glu gly val arg glu asp asp pro val glu leu ala arg glu met val ala gly

GGC GCG ACG GCG CTG TCG GTC CTC ACC GAA CCC GAG CAC TTC GGA GGC TCC GCC GAG TCG  700
gly ala thr ala leu ser val leu thr glu pro glu his phe gly gly ser ala glu ser

CTG CGG CGC ATC CGC GAG GCC GTC GAC GTG CCC GTG CTC CGC AAG GAC TTC ATC ATG AAC  760
leu arg arg ile arg glu ala val asp val pro val leu arg lys asp phe ile met asn

GAG GCC CAA CTG GAT GTC GTC CAG TCC GAC CTC GTG CTC CTC ATC GCG CGG TTC GTC GGC  820
glu ala gln leu asp val val gln ser asp leu val leu leu ile ala arg phe val gly

GAG GAC CTC CCG GCC CTC GTC GAG GCG GCC CGC GAC CGC GGC TTC CAG CCG CTC GTG GAG  880
glu asp leu pro ala leu val glu ala ala arg asp arg gly phe gln pro leu val glu

GTC CAT ACG CGC GAG GAA CTC ACG GCG GCG CTC GCG GCC GGT GCC GAC ATC GTC GGC ATC  940
val his thr arg glu glu leu thr ala ala leu ala ala gly ala asp ile val gly ile

AAC AAC CGC GAC CTC GGG AAG CTG GAA GTC GAC CTC GGC ACG TTC GAG GAA CTC GCG CCC  1000
asn asn arg asp leu gly lys leu glu val asp leu gly thr phe glu glu leu ala pro

GAA GCG CCC GAG GAC GTG CTT CTC GTG GCC GAA AGC GGC GTG CAG ACG GTC GAC GAC GCG  1060
glu ala pro glu asp val leu leu val ala glu ser gly val gln thr val asp asp ala

CGG CGG ATG CGC GAG GCC GGT GCC GAC GCC CTC CTC GTC GGG ACC GCC ATC ATG GAC GGC  1120
arg arg met arg glu ala gly ala asp ala leu leu val gly thr ala ile met asp gly
                                        trpB->
GAC GTG CGA CAG AAC ACG GAG ACA CTC ACA CAA TG  AGC GCA GAC GGC AAA TTC GGC GAC  1179
asp val arg gln asn thr glu thr leu thr gln CPA
                                          Met   ser ala asp gly lys phe gly asp

TAC GGC GGA CAG TAC GTT CCC GAG GCA CTC ATG CCG GCC ATC GAG GAA CTG ACC GAC GCG  1239
tyr gly gly gln tyr val pro glu ala leu met pro ala ile glu glu leu thr asp ala

TAC GAG CGG TAC GTC CTC GAC AAC GAA GAC GGC TTC ATG GAC GAC TTC CGC GCG CGA CTG  1299
tyr glu arg tyr val leu asp asn glu asp gly phe met asp asp phe arg ala arg leu

CGG GAC TTC GGC GGG CGA CCG ACG CCC CTC CAG CGC GCG GAC CGA CTC TCG GAG CGC TAC  1359
arg asp phe gly gly arg pro thr pro leu gln arg ala asp arg leu ser glu arg tyr
```

continues...

Figure 18. DNA sequence of the *trpCBA* region and the amino acid sequence of open reading frames predicted by this sequence. Arrows indicate repeated sequences.

Figure 18 continues...

```
GAC CGC GAG GTC TAC CTC AAG CGC GAG GAC CTC CTC CAC GGC GGC GCG CAC AAG CTC AAC  1419
asp arg glu val tyr leu lys arg glu asp leu leu his gly gly ala his lys leu asn

AAC GCG CTC GGG CAG GTC TTA CTG GCG AAG TAC ATG GGC AAA GAG CGC ATC ATC GCC GAG  1479
asn ala leu gly gln val leu leu ala lys tyr met gly lys glu arg ile ile ala glu

ACC GGC GCG GGC CAG CAC GGC ACC GCC ACG GCG ATG GCC TGC GCG CAC CTC GAT ATG CCC  1539
thr gly ala gly gln his gly thr ala thr ala met ala cys ala his leu asp met pro


TGC GAG ATT TAC ATG GGC GAG CGC GAC ATC AAC CGC CAG CGC CCC AAC GTC TTC CGG ATG  1599
cys glu ile tyr met gly glu arg asp ile asn arg gln arg pro asn val phe arg met

AAG CTC AAC GGC TCC GAG GTG AAC CCC GTC ACC GTC GGC CGC GGC ACG CTC AAG GAG GCC  1659
lys leu asn gly ser glu val asn pro val thr val gly arg gly thr leu lys glu ala

ATC TCC GAG ACG ATG CGC GAC TGG GAG ACC AAC GTC GAG GAC ACC CAC TAC GTC ATC GGC  1719
ile ser glu thr met arg asp trp glu thr asn val glu asp thr his tyr val ile gly

TCC GTC GTC GGC CCG CAC CCG TTC CCG AGC ATG GTT CGG GAC TTC CAG TCG GTC ATC TCC  1779
ser val val gly pro his pro phe pro ser met val arg asp phe gln ser val ile ser

GAG GAG GCC CGC ACG CAG GCC AGA GAG AAA CTC GGC CGG CTC CCC GAC GCC GTC GTC GCC  1839
glu glu ala arg thr gln ala arg glu lys leu gly arg leu pro asp ala val val ala

TGC GCG GGC GGC GGC TCG AAC ACG ATG GGC GCG TTC GCC GAG TTC GTC GAC GAC GAG GAG  1899
cys ala gly gly gly ser asn thr met gly ala phe ala glu phe val asp asp glu glu

ACC GCG CTC TAC GCC GTC GAG GCC GGC GGC TCG ACG CTC GAA GTG GAC GAG GAA GCC GGC  1959
thr ala leu tyr ala val glu ala gly gly ser thr leu glu val asp glu glu ala gly

GTC GCG CCC AAC TCG GCG TCG CTC ACG ACC GGG TCG GAG GGC ATT CTC CAC GGC GCG CGC  2019
val ala pro asn ser ala ser leu thr thr gly ser glu gly ile leu his gly ala arg

ACC CGG CTC TTG CAG GAC CGC GAC GGC CAG ATT ATG GAG TCG CAC TCG GTG TCG TCC GGC  2079
thr arg leu leu gln asp arg asp gly gln ile met glu ser his ser val ser ser gly

CTC GAC TAC GCC GGC GTC GGC CCC GAA CTC GCA CAC CTC GTG GAC ACC GGC CGC GTC ACC  2139
leu asp tyr ala gly val gly pro glu leu ala his leu val asp thr gly arg val thr

GCC GTC AAC GTC GAC GAC GAC GCG GCG CTG ACC GCG TTC CAC CGG CTC TCG CAG ATG GAG  2199
ala val asn val asp asp asp ala ala leu thr ala phe his arg leu ser gln met glu

GGC ATC ATC CCC GCC CTG GAG TCG GCC CAC GCG TTC GGC TAC CTC GAA GTC TGG TCG TCT  2259
gly ile ile pro ala leu glu ser ala his ala phe gly tyr leu glu val trp ser ser

GGT CCC GAC GCG CCC GAC GCC GAG AAC GCG GAC GAC CTC GGC GAG TAC GTC GTG GTC AAC  2319
gly pro asp ala pro asp ala glu asn ala asp asp leu gly glu tyr val val val asn

GTT TCC GGT CGC GGC GAC AAA GAC CTC GAA TCG GCC ATC GAG GAG ACC TAC GAG CGC GAC  2379
val ser gly arg gly asp lys asp leu glu ser ala ile glu glu thr tyr glu arg asp
                                                        trpA->
ATC GAT ATC GCG CCG AAC ATG GAC GAG TTC ACG GGG GGC CTG TGA  TG TCG CTC GAA GAC  2438
ile asp ile ala pro asn met asp glu phe thr gly gly leu OPA
                                                      Met   ser leu glu asp

GCC TTC TCC GAC GGC CCG GCG TTC GTC CCC TAT CTT GCC GCC GGC GAC CCC GAC TAC GAG  2498
ala phe ser asp gly pro ala phe val pro tyr leu ala ala gly asp pro asp tyr glu

TCC TCG CTC GAA TAC GTC GAG GCG CTC GAA CGC GGC GGC GCG GAC GTC ATC GAA CTC GGA  2558
ser ser leu glu tyr val glu ala leu glu arg gly gly ala asp val ile glu leu gly

CTT CCG TTC TCG GAG CCC ATC GCC GAG GGG CCG ACC ATC CAG AAC GCG GTT GTT CGG TCG  2618
leu pro phe ser glu pro ile ala glu gly pro thr ile gln asn ala val val arg ser
```

continues...

Figure 18 continues...

```
CTC GAA GGC GGC ATG ACG CCG ACG CGC TTT TTC GAG TTC GTC GAG GAC CTC GAC GTG TCG  2678
leu glu gly gly met thr pro thr arg phe phe glu phe val glu asp leu asp val ser

GTG CCG CTG GTC TGT ATG AGG TAC TAC AAC CTC ATC TAT CGC TAC GGC GAT GAA CCC GGA  2738
val pro leu val cys met arg tyr tyr asn leu ile tyr arg tyr gly asp glu pro gly

CCG CGG CCG TTC GTC GAG AAG GCG GCG GAA GTC GGT ATC GAG GGC TTC GTC GTC CCC GAC  2798
pro arg pro phe val glu lys ala ala glu val gly ile glu gly phe val val pro asp

CTG CCG GCC GAG GAG GCC GGC CCG CTC CGC GAG GCC TGC GAC GAG TTC GGC CTC GAC CTC  2958
leu pro ala glu glu ala gly pro leu arg glu ala cys asp glu phe gly leu asp leu

GTG TTC ATC GTC GCG CCG ACG ACC CGC GGC GAG CGC CTC GAC CGA ATC ATG GAA CAG GTC  2918
val phe ile val ala pro thr thr arg gly glu arg leu asp arg ile met glu gln val

TCG GGC TAC GTC TAC GTG CAG GCG CGC CTC GGC ACG ACG GGC GCG CAG TCG AGC GTC TCC  2978
ser gly tyr val tyr val gln ala arg leu gly thr thr gly ala gln ser ser val ser

GAC CAG ACC GAC TCG TCG CTC GAA CGA CTC ACC GAC TAC GAC GTG CCC AAG GCG GTC GGC  3038
asp gln thr asp ser ser leu glu arg leu thr asp tyr asp val pro lys ala val gly

TTC GGA ATC AGC GAC GGC GAC CAC GCC GAG CGC ATC GTC GCC AGC GGT GCC GAC GGC ATC  3098
phe gly ile ser asp gly asp his ala glu arg ile val ala ser gly ala asp gly ile

ATC GTC GGC AGC GCG CTC GTC GAC ATC GTC GCC GAG GGC CAC GAG AAC GGC GAC GAC GCC  3158
ile val gly ser ala leu val asp ile val ala glu gly his glu asn gly asp asp ala

GAA ACG GTC GCG GAC CGA CTC GAA ACG CTC GCC CGC GAA CTC GAA GAC GGT GCG GTA GCG  3218
glu thr val ala asp arg leu glu thr leu ala arg glu leu glu asp gly ala val ala

GGC GCG TCG CAA CGC CCA CCG CAT CCG GAA CGC ACA TAA CTG CTC TGC CAC CCT TTC TCA  3278
gly ala ser gln arg pro pro his pro glu arg thr OCH
                  --------->   <----------
TAC ATG AAC ACA GAC GTC GGA CTC TCC GCA CGA CTC GAC CGC ATT TCC ACA GAC GGG CGA  3338

TAC CTC ATC GTC CCG ATG GAC CAC GGC ATC ACC CTC GGC CCG GTC ACG GGC CTC GTC GAT  3398
                  Met asp his gly ile thr leu gly pro val thr gly leu val asp

ATC GAA TCC ACT ATC GAC GGC ATC ACG CGC GGC GGC CGT GAC GCC GTC CTC ACG CAC AAG  3458
ile glu ser thr ile asp gly ile thr arg gly gly arg asp ala val leu thr his lys

GGG ACC GCG CCG CGC GTC CAC CCG AAC AAA AAC GGC AAA GGC TAC ATC GTC CAC GTA AAC  3518
gly thr ala pro arg val his pro asn lys asn gly lys gly tyr ile val his val asn

GGC TCG ACG GAC ATC GGC CCG GAC GAA AAC CAC AAG CGA CTC  ....                    3560
gly ser thr asp ile gly pro asp glu asn asp lys arg leu  ....
```

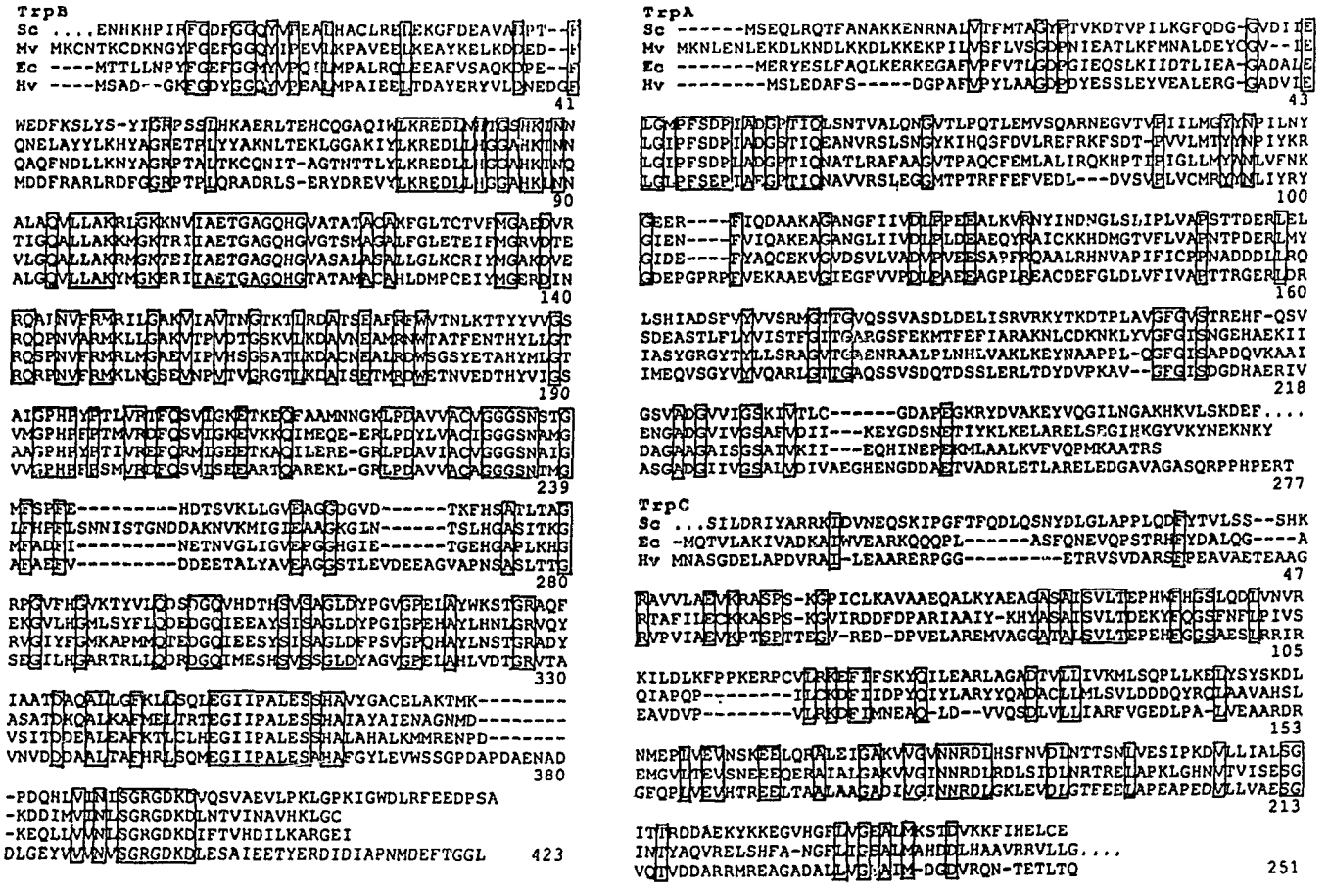


Figure 19. Alignment of amino acid sequences for TrpB, TrpA and TrpC from Sc (yeast), Mv (*M. voltae*), Ec (*E. coli*) and Hv (*Hf. volcanii*). Sequences were aligned by the MULTALIN method of Corpet (Corpet 1988). Very few adjustments were required to conform the alignments to those reported by other groups (Sibold and Henriquet 1988; Essar et al. 1990a). Conserved residues are boxed. Number at the end of each block refers to the position of the *Hf. volcanii* sequence.

Table 4. Conservation of TrpC, B and A sequences between *Hf. volcanii* and other organisms.

organisms compared	identical amino acids (%) in						
	TrpD	TrpF	TrpE	TrpG	TrpC	TrpB	TrpA
E. coli	36	26	30	32	32	45	32
S. cerevisiae	25	29	25	32	36	45	33
P. aeruginosa	41			37	38	49	37
M. voltae	---	---	---	---	---	47	40

The catalytic residue Lys-87 (Hyde and Miles 1990) and the Gly-Gly-Gly-Ser-Asn stretch at position 232-236 involved in co-enzyme binding (Crawford 1989; Hyde et al. 1988) are conserved in *Hf. volcanii*, along with other residues involved in TrpB function (His-82, His-86 and Cys-230). The invariant tryptophan residue of unknown function at position 177 is also found.

The 834 bp ORF downstream from *trpB* encodes a 277 residue peptide resembling the *E. coli* TrpA sequence (figure 19). As in *E. coli*, the start codon (AUG) of *trpA* overlaps the stop codon (UGA) of *trpB* by one nucleotide (UGAUG). In contrast, the *trpB* and *trpA* gene in the methanogen are separated by a 37 bp AT-rich region (Sibold and Henriquet 1988).

Overall sequence conservation for TrpA is significantly lower than for TrpB (Table 4). The *Hf. volanii* TrpA is about 30-40% identical to all TrpA sequences we have compared. Important residues (Phe-22, Glu-49, Tyr-175, Thr-183, Gly-211, Gly-234 and Ser-235) defined by missense mutations and second site reversion in *E. coli* (Crawford 1989; Hyde and Miles 1990; Yanofsky and Crawford 1972) are all found in the *Hf. volcanii* sequence, except Tyr-16 substitutes for Phe-22. Leu-177 is replaced by Asn-172 in *Hf. volcanii*. Isoleucine and valine substitutions at this position have previously been reported in other organisms (Sibold and Henriquet 1988, Zalkin and Yanofsky 1982). The Gly-Phe-Gly-Ile stretch which contacts the substrate in *E. coli* is located, by sequence alignment, at position 205-208 of the *Hf. volcanii* TrpA.

In the α subunit of the *Salmonella* tryptophan synthase, Glu-49 and Asp-60 serve as proton donors/acceptors in the cleavage of indoleglycerol phosphate. Substituting Asp-60 by Asn, Ala, or Tyr destroys α chain activity, but glutamic acid replacement retains partial activity (Hyde and Miles 1990). In fact, glutamic acid (Glu-54) is the residue found at the corresponding position in *Hf. volcanii*.

Immediately upstream of *trpB* is a 756 bp ORF which overlaps *trpB* by 4 nucleotides (AUGA, figure 19). The deduced peptide (251 a.a.) has stretches of

conserved sequences common to the indoleglycerol phosphate synthase (IGPS) domain of yeast (Zalkin et al. 1984) and eubacterial (Yanofsky et al. 1981; Essar et al. 1990a, b) TrpC sequences (figure 19 and Table 4). Residues which may contribute to catalytic activity of *E. coli* IGPS are mostly conserved in *Hf. volcanii.*

The sequences GTGAT (at the *trpBA* junction) and GGAG (near the end of *trpC*, 11 bp upstream of *trpB*) are both complementary to the 3'-terminus of the small subunit ribosomal RNA of *Hf. volcanii* (3'-UCCUCCACUA...), and may serve as ribosome-binding sites.

Codon usage of the *trpC*, *trpB* and *trpA* genes is very much biased to codons with C or G in the third position, reflecting the high G+C content of these genes (68 to 71 %) -- this is consistent with the >65% G+C estimated for halobacterial genomes (Moore and McCarthy 1969a, b).

***trpDFEG* gene cluster.** DNA sequencing of the region which complements anthranilate- and indole-utilizing mutants (defined in figure 17A and table 3) revealed four open reading frames that could encode a 337 residue TrpD, a 211 residue TrpF, a 523 residue TrpE and a 204 residue TrpG-like protein (figure 20). Although the gene structure-function relationship of eubacterial and eukaryotic *trpD*, *trpF*, *trpE* and *trpG* have not been characterized to the same extent as *trpB* and *trpA*, a large number of gene sequences are available for comparison.

The first gene of the cluster, *trpD*, encodes phosphoribosyl (PR) transferase. Variable regions occur throughout the alignment of many TrpD sequences (Crawford 1989; Essar et al. 1990a). All but a few conserved positions in the alignment by Crawford and co-workers are found in the *Hf. volcanii* sequence. Alignment of TrpD

```
                                                                AAGCTTTCGAATCGGACTACGTGCG  25
GCGGATTCGGGTCACAGCGCGGTCCCGCCGTCCCGAGGCGGTGCGATTCGACCCGCGCGCCGACCAGTCGATGCCGGCG  104
ATTTTCGCCAGCCCGACACGCTTAAGCCGATGTACGAATTTGTACATCGTAACCCGAAGAAGTACATCGGTGATTACCA  183
trpD->
ATG CAG GAT TAT ATC GAA CGT GTG ACC GGC GGT GCG GAC CTG ACT GTC GAG GAG GCG CGC  243
Met gln asp tyr ile glu arg val thr gly gly ala asp leu thr val glu glu ala arg

CGA CCG CCG CGC GGG CGG TCT TCG GAG GAC GCG ACC GAG GCG CAA ATC GGC GCG CTC CTC  303
arg pro pro arg gly arg ser ser glu asp ala thr glu ala gln ile gly ala leu leu

GCC GCG CTC CGG GCG AAA GGC GAG ACC GAG GCC GAA ATC GCG GGG TTC GCG CAG GGG ATG  363
ala ala leu arg ala lys gly glu thr glu ala glu ile ala gly phe ala gln gly met

CGC GAC GCC GCC CTG ATC CAT CGA GCC CGA CCC GCG GCC CGC TCG TCC GAT ACT GCT GGC  423
arg asp ala ala leu ile his arg ala arg pro ala ala arg ser ser asp thr ala gly

ACC GGC GGC GAC GAC TAC AAC ACC ATC AAC GTC CTC GAc CCC ACG ACG CGC TCG TCC GCG  483
thr gly gly asp asp tyr asn thr ile asn val leu asp pro thr thr arg ser ser ala

GCC GCG CCG GGC GCg gCG GCG GTG GCC AAG CAC GGC AAC TAC TCC GTC TCT TCG TCT TCT  543
ala ala pro gly ala ala ala val ala lys his gly asn tyr ser val ser ser ser ser

GGG AGC GCC GAC GTG CTG GAG GTC GCC GGC GTG AAC GTC GAG GCC GAA CCC GAG TCC GTC  603
gly ser ala asp val leu glu val ala gly val asn val glu ala glu pro glu ser val

GAG GCG TGC ATC GAG GAC AAC GGC GTC GGC TTC ATG CTC GCG CCC GTG TTC CAC CCC GCG  663
glu ala cys ile glu asp asn gly val gly phe met leu ala pro val phe his pro ala

ATG AAG GCC GTC ATC GGC CCG CGG AAG GAA CTC GGC ATG CGG ACG GTC TTC AAC GTC CTC  723
met lys ala val ile gly pro arg lys glu leu gly met arg thr val phe asn val leu

GGC CCG TTG ACG AAC CCC GCC GGG GCC GAC GCG CAG GTC CTC GGC GTC TAC GAC GCC GAC  783
gly pro leu thr asn pro ala gly ala asp ala gln val leu gly val tyr asp ala asp

CTC GTC CCC GTC ATC GCC GAG TCG CTG TCG CAC ATG CCC GTC GAG CGC GCC CTC GTC GTC  843
leu val pro val ile ala glu ser leu ser his met pro val glu arg ala leu val val

CAC GGC TCC GGC ATG GAC GAA ATC GCG CTC CAC GAC CGG ACC ACG GTC GCC GAA ATC GAC  903
his gly ser gly met asp glu ile ala leu his asp arg thr thr val ala glu ile asp

GGC GAC GAA ATC ACC GAG TAC ACC CTC ACG CCC GCC GAC CTC GGC CTC GAA CGG GCC CCC  963
gly asp glu ile thr glu tyr thr leu thr pro ala asp leu gly leu glu arg ala pro

ATC GAG GCG GTC GCC GGC GGC ACG CCG CAG GAG AAC GCC GAC GAC CTC GAA GGC ATC CTC 1023
ile glu ala val ala gly gly thr pro gln glu asn ala asp asp leu glu gly ile leu

ACC GGC GAC GTG ACC GGG CCG AAG CGC GAC CTC ATC CTC GCC AAC GCC GGC GCG GCA ATG 1083
thr gly asp val thr gly pro lys arg asp leu ile leu ala asn ala gly ala ala met

TAC GTC GCC GGC CTC GCG GAC TCG CTG GAG GGG GGC GTC GAA GTC GCC CGC GAC GCC ATC 1143
tyr val ala gly leu ala asp ser leu glu gly gly val glu val ala arg asp ala ile
                                                          trpF->
GAC TCC GGG GCC GCG AAA GCC AAA CAC GAC GCG CTT CCG GGA GGC GTG AGA TG  ACC CGC 1202
asp ser gly ala ala lys ala lys his asp ala leu pro gly gly val arg OPA
                                                        Met   thr arg

GTC AAA GTG TGC GGC GTC ACC GAC GAA ACC GAC CTC GCC GCC GTC GAC GCG GCC GGT GCC 1262
val lys val cys gly val thr asp glu thr asp leu ala ala val asp ala ala gly ala

GAC GCG GTC GGT GCC ATC TGC GAC GTG CCC GTG GAC ACG CCC CGC GAG ATT CCC CGC GAG 1322
asp ala val gly ala ile cys asp val pro val asp thr pro arg glu ile pro arg glu

CGC GCC CGC GAA CTG TTC GCC GCC GCG CCG CCG TTC CTC ACG ACC ACC CTC GTG ACG ATG 1382
arg ala arg glu leu phe ala ala ala pro pro phe leu thr thr thr leu val thr met
```

continues...

Figure 20. DNA sequence of the *trpDFEG* region. Deduced amino acid sequence of TrpD, TrpF, TrpE and TrpG are shown. Direct repeats immediately downstream of *trpG* are indicated by arrows.

Figure 20 continues...

```
CCC GAC TCC GTC GAC CAC GCC CGA GAC CTC GCC CGC GAG GTC GGC CCG GAC GTG CTC CAA  1442
pro asp ser val asp his ala arg asp leu ala arg glu val gly pro asp val leu gln

CTC CAC GGC GAC TTC GCG GCC GAC GAC CTC GAC TCG CTC CGC GCG ACG GGC GAC GGC GTC  1502
leu his gly asp phe ala ala asp asp leu asp ser leu arg ala thr gly val gly val

GTC CCG GTC GTC GAC GCG ACC GAC CTC GCC CGC GCA CGC GAC CTC GCG CCC GTC GTC GAC  1562
val pro val val asp ala thr asp leu ala arg ala arg asp leu ala pro val val asp

GCT ATC CTC GTC GAC ACG CCC TCC GAC TCG GGC GCG GGC GGC ACC GGC GAG ACC CAC GAC  1622
ala ile leu val asp thr pro ser asp ser gly ala gly gly thr gly glu thr his asp

TGG GAC GCC TCG CGC GAC CTC GTC GCG GCG GTC GAC GCG CCC GTA ATT CTC GCG GGC GGC  1682
trp asp ala ser arg asp leu val ala ala val asp ala pro val ile leu ala gly gly

CTG ACG CCA GAC AAC GTC GTC GAG GCG GTT CGA ACC GTC GAA CCT TAC GGC GTC GAC GTC  1742
leu thr pro asp asn val val glu ala val arg thr val glu pro tyr gly val asp val

GCG AGC GGC GTC GAG GCC TCC GGC GGG GTC AAG GAC CAC GAC GCG GTC CGC GCG TTC GTC  1802
ala ser gly val glu ala ser gly gly val lys asp his asp ala val arg ala phe val
                                                                    trpE->
GCC GCG GCG AAG ACG GCT CGC GGA GCG GTC GAC GAC CAC GAG GAG GTC GTC GCG TG  ACG  1861
ala ala ala lys thr ala arg gly ala val asp asp his glu glu val val ala OPA
                                                                      val   thr

GCC CCC GAC ACC GAC CGC GAG GAG TTC GTC TCC CTC GCC GGC GAC GCG GAC GGA CCG GTC  1921
ala pro asp thr asp arg glu glu phe val ser leu ala gly asp ala asp gly pro val

GTG ACC CAC CTC GTG GCC GAC CTC GAC GTG TCC GTG GAC CCG CTG GCC GCG TAC ACG ACG  1981
val thr his leu val ala asp leu asp val ser val asp pro leu ala ala tyr thr thr

CTC GCG GAC CGC AGC GAC TAC GGC TTC CTC CTG GAG AGC GCC GAG AAG GTC TCC TCG AGC  2041
leu ala asp arg ser asp tyr gly phe leu leu glu ser ala glu lys val ser ser ser

AAC CCG CAG GCG CGT TCT CCG CGC CCG CGC ACG GCC GCC GAC TCC CAC GCG CGC TTC TCG  2101
asn pro gln ala arg ser pro arg pro arg thr ala ala asp ser his ala arg phe ser

TTC GTC GGC TAC GAC CCC GAA GCG GTC GTG ACG GTC GGT CCC GAC GGC GTC GAC GTG ACC  2161
phe val gly tyr asp pro glu ala val val thr val gly pro asp gly val asp val thr

GAC CTC GGC GGC CCC GCC GCG GAG TTC GTC GGC GCG GGC GAC GGC GAC GTG CTC GAT TCC  2221
asp leu gly gly pro ala ala glu phe val gly ala gly asp gly asp val leu asp ser

CTG CGC GGC GCG CTC CCC GAC CTC CCG CGC GTC AAC TTC CCC GAG ACG GAC GCA GAC GTC  2281
leu arg gly ala leu pro asp leu pro arg val asn phe pro glu thr asp ala asp val

ACC GGC GGG CTG GTC GGC TTC CTC GCC TAC GAG GCC GTC TAC GAC CTC TGG CTC GAC GAG  2341
thr gly gly leu val gly phe leu ala tyr glu ala val tyr asp leu trp leu asp glu

GTC GGC CGC GAG CGC CCC GAC ACG GAC GAC CCG GAC GCC GAG TTC GTC CTC ACG ACC CGG  2401
val gly arg glu arg pro asp thr asp asp pro asp ala glu phe val leu thr thr arg

ACC CTC TCG TTC GAC CAC CGC GAG GAC GCC GTC CGC CTC GTC TGT ACC CCC GTC GTC TCG  2461
thr leu ser phe asp his arg glu asp ala val arg leu val cys thr pro val val ser

CCC GAC GAC GAC CCC GGC GAG GTG TAC GAC GGC GTC GTC GCC GAG GCC GAG CGC GTC GCG  2521
pro asp asp asp pro gly glu val tyr asp gly val val ala glu ala glu arg val ala

GAA AAG CTC CGC GCG GCG GAC GAC CCC GCC CCC GGC GGC TTC GAG CGG ACC GGC GAG GAC  2581
glu lys leu arg ala ala asp asp pro ala pro gly gly phe glu arg thr gly glu asp

GCC GGC TCC CGC GAG GAG TAC GAG GCC GCG GTC AGG AAG ACG AAA GAA CAC GTC CGC GAC  2641
ala gly ser arg glu glu tyr glu ala ala val arg lys thr lys glu his val arg asp

GGC GAC ATC TAT CAG GGC GTC ATC TCG CGC ACC CGG AAG CTC CGC GGG CAG GTC GAC CCG  2701
gly asp ile tyr gln gly val ile ser arg thr arg lys leu arg gly gln val asp pro

GTC GGT CTG TAC GCC TCG CTG CGC GAG GTG AAC CCC TCG CCG TAC ATG TTC CTG CTC CGG  2761
val gly leu tyr ala ser leu arg glu val asn pro ser pro tyr met phe leu leu arg
```

continues...

Figure 20 continues...

```
CAC GGC GAC CGG CGC GTC GTC GGC GCG AGC CCC GAA ACG CTC GTC TCC GTC AGG GGC GAC  2821
his gly asp arg arg val val gly ala ser pro glu thr leu val ser val arg gly asp

CGC GTC GTC GTC AAC CCC ATC GCG GGG ACG TGC CAG CGC GGG TCC GGC CCG GTT GAG GAC  2881
arg val val val asn pro ile ala gly thr cys gln arg gly ser gly pro val glu asp

CGC CGC CTC GCC GGC GAA CTC CTC GCG GAC GCG AAA GAG CGC GCC GAG CAC ACG ATG CTC  2941
arg arg leu ala gly glu leu leu ala asp ala lys glu arg ala glu his thr met leu

GTC GAC CTC GGC CGC AAC GAC GTG GGC CGG GTC TCG ACG CCC GGG AGC GTC CGC GTC GAG  3001
val asp leu gly arg asn asp val arg arg val ser thr pro gly ser val arg val glu

GAC TTC ATG AGC ATC ATC AAG TAC AGC CAC GTC CAG CAC ATC GAA TCG ACC GTC TCG GGG  3061
asp phe met ser ile ile lys tyr ser his val gln his ile glu ser thr val ser gly

ACG CTC GAC GCC GAC GCC GAC GCC TTC GAC GCC ACG CGA GCG ACG TTC CCC GCG GGG ACG  3121
thr leu asp ala asp ala asp ala phe asp ala thr arg ala thr phe pro ala gly thr

CTC ACC GGC GCG CCG AAG GTC CGC GCG ATG GAG ATT ATC GAC GAC CTC GAA GCC GAG CCC  3181
leu thr gly ala pro lys val arg ala met glu ile ile asp asp leu glu ala glu pro

CGT GGC GTC TAC GGC GGC GGC GTC GGC TAC TAC TCG TGG ACC GGC GAC GCC GAC GTG GCA  3241
arg gly val tyr gly gly gly val gly tyr tyr ser trp thr gly asp ala asp val ala

ATC GTC ATC CGA ACC GCG ACG GTC GAC TCC GGC GGC GCG GAC GAC GCC ATC ACC GTC CGC  3301
ile val ile arg thr ala thr val asp ser gly gly ala asp asp ala ile thr val arg

GCG GGT GCC GGC ATC GTC GCC GAC TCC GAC CCG ACC GCG GAG TAC GAG GAG ACG GAA CAG  3361
ala gly ala gly ile val ala asp ser asp pro thr ala glu tyr glu glu thr glu gln

AAG ATG GGC GGG GTG CTC GAC GCC GTC CGC CGC ATC GAG TAC GGG ACC GAG GAG GCG TCG  3421
lys met gly gly val leu asp ala val arg arg ile glu tyr gly thr glu glu ala ser
  trp?->
CAA TG  ATT CGG CTC GTC GTC GTC GAC GAC AAC TTC GAC TCC TTC ACG TAC AAC CTC GTG GAG  3480
gln OPA
  Met   ile arg leu val val val asp asn phe asp ser phe thr tyr asn leu val glu

TAC TTC TCC GAG CAG ACC GTC GAG GGC GAA CCG CTC GAC ATC GAG GTG CGC AAG ACC ACC  3540
tyr phe ser glu gln thr val glu gly glu pro leu asp ile glu val arg lys thr thr

GCC TCG CTC GAC GAG ATA CGC GAC CTC GAC CCC GAC GCT ATC GTC ATC TCG CCG GGG CCT  3600
ala ser leu asp glu ile arg asp leu asp pro asp ala ile val ile ser pro gly pro

GGC CAC CCG AAA AAC GAC CGC GAC GTG GGC GTG ACC AAC GAC GTT CTC ACG GAG CTG TCG  3660
gly his pro lys asn asp arg asp val gly val thr asn asp val leu thr glu leu ser

ACC GAG ATT CCC ACC CTC GGC GTC TGT CTC GGC CTC GAA GCC GCG GTG TAC GCC TAC GGC  3720
thr glu ile pro thr leu gly val cys leu gly leu glu ala ala val tyr ala tyr gly

GGC ACC ATC GGC CAC GCG CCG GAC GCG ATT CAC GGG AAG GCG TTC CCC GTC GAC CAC GAC  3780
gly thr ile gly his ala pro asp ala ile his gly lys ala phe pro val asp his asp

GGC GCG GGC GTC TTC GCC GGC CTC GAA GAC GGC TTC CCG GCC GGG CGC TAC CAC TCG CTC  3840
gly ala gly val phe ala gly leu glu asp gly phe pro ala gly arg tyr his ser leu

GTC GCC ACG GAC GTC CCC GAC TGC TTC GAC GTC TCT GCG ACG ACC GAC CAC GAC GGC GAG  3900
val ala thr asp val pro asp cys phe asp val ser ala thr thr asp his asp gly glu

GCG CTG GTG ATG GGC GTC CGC CAC CGC GAC TAC CCA ATC GAG TGC GTC CAG TTC CAC CCC  3960
ala leu val met gly val arg his arg asp tyr pro ile glu cys val gln phe his pro

GAG AGC GTG CTC ACG GGG TCG GGA CAC GGC GTC GTC AGA AAC TTC CTC ACG GCG GTC GGC  4020
glu ser val leu thr gly ser gly his gly val val arg asn phe leu thr ala val gly

GGC TTC GAC GTG GCC TGA GTC GGG TCG GGT CGG GTC AGA CGC GAC GGG CGA GAC GTG CGC  4080
gly phe asp val ala OPA

ACTCCGTGTCGGCCGCGGCGACGGCCGCGCGCTCGGATTACAGGACGTTGATGCCGAAAAACGAGAGCAACAGCAGCGC  4172
GCCCACGACGATGGCGGCGATAGTGACCAGTCGCCACGCGATTTTGAGGAACAGCCGGCCGACGAGGATGACGACGGCG
ACGGCCGCGAGCACCACGAGCATCTGACCGACGCACGGCCCCGAGAGCAGGCCGCCGAGCTGTAACGGGGTGAGCGCGA
GGGGAGTCAGTACCATACCCCGATATGACATCCGTAGAGCGGATAAGCTT                              4380
```

sequences from yeast (Furter et al. 1986), *E. coli* (Yanofsky et al. 1981) and *Hf. volcanii* is presented in figure 21.

Hf. volcanii ORF *trpF*, which encodes N-phosphoribosylanthranilate (PRA) isomerase, overlaps *trpD* by 4 bp (AUGA). The sequence of the 221 residue TrpF protein conforms to the alignment of 10 eubacterial and eukaryotic sequences proposed by Priestle et al. (1987) according to predicted protein structures. With 26% identical residues with *E. coli* and 29% with yeast, *trpF* is the least conserved of the Trp enzymes (figure 21 and Table 4). Residues of functional importance (Priestle et al. 1987, Crawford 1989) are largely conserved in the *Hf. volcanii* TrpF.

ORF *trpE*, which encodes the α subunit of anthranilate synthase, does not begin with a methionine. Overlapping the stop codon of the upstream *trpF* is a codon for valine (GUGA). Valine start has been reported in archaebacterial protein-coding genes before (Gropp et al. 1989). Sequence conservation among eubacterial and eukaryotic TrpE proteins is concentrated in the C-terminal domain; the N-terminal half of the protein is usually variable (Crawford 1989). *Hf. volcanii* TrpE is no exception: conservation within the first 250 residues is minimal (figure 21). Interestingly, the segment between position 52 and 55 (leu-leu-glu-ser), identified as a site for mutations affecting feedback inhibition by tryptophan in *Brevibacterium lactofermentum* (Matsui et al. 1987) and in *Salmonella typhimurium* (cited as R. Bauerle personal communication, by Crawford 1989), is conserved in the *Hf. volcanii* sequence.

The high degree of sequence conservation in the C-terminal half of TrpE proteins reflects the functional importance of this region. This is consistent with the locations of five inactivating missense mutations occuring between position 388 and 526 of the *Salmomella* TrpE (cited as R. Bauerle personal communication, by Crawford 1989).

TrpG, the β subunit of anthranilate synthase, is the smallest of the seven *Hf. volcanii* Trp sequences. The *Hf. volcanii* sequence fits with the alignment of eubacterial

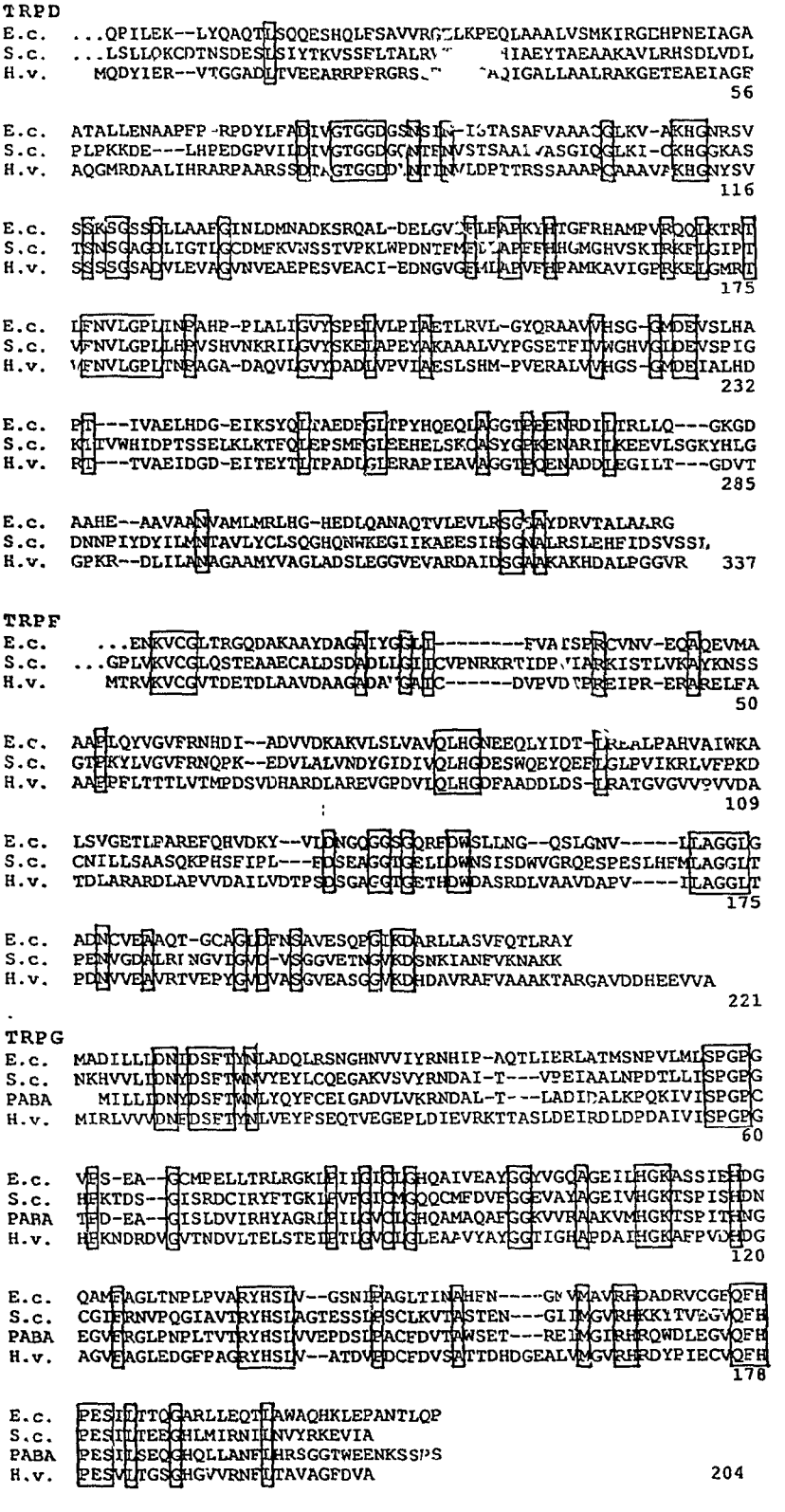


Figure 21 continues...

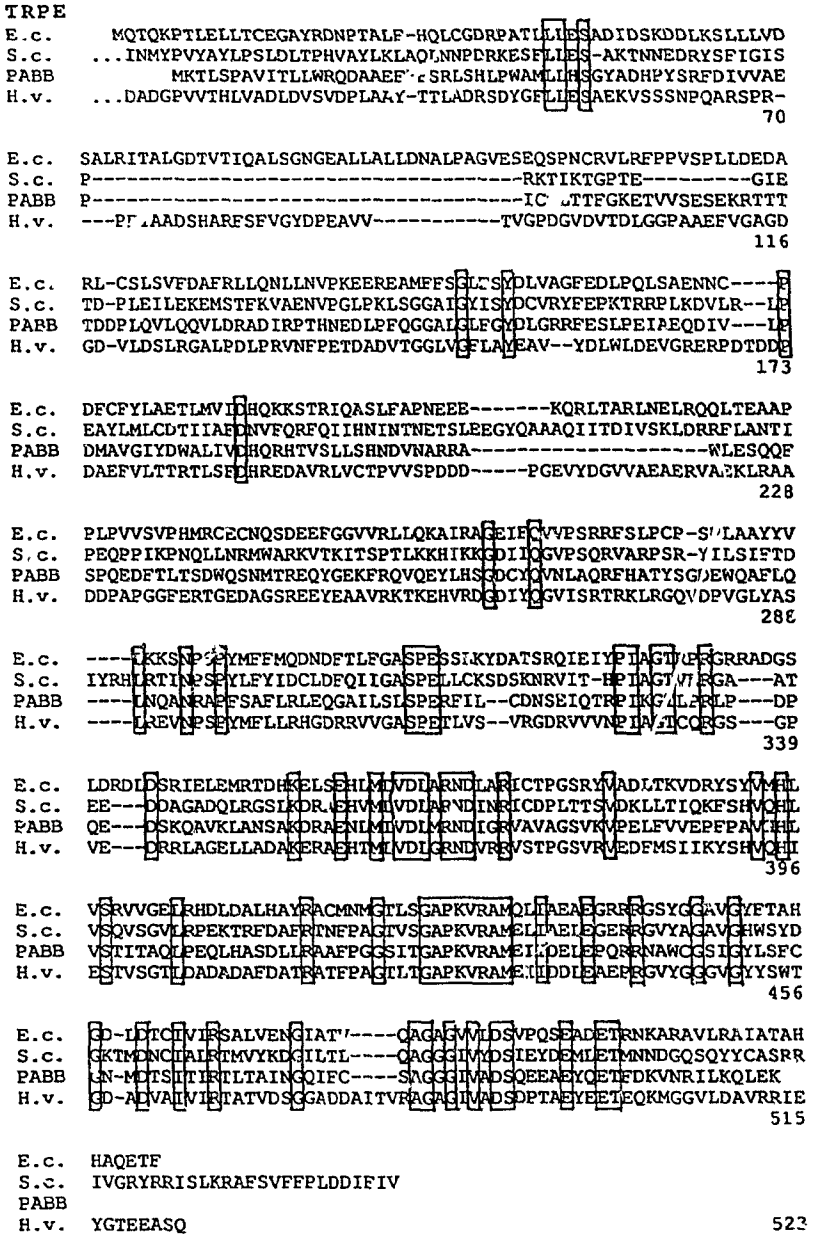


Figure 21. Comparison of *Hf. volcanii* TrpD, TrpF, TrpE and TrpG amino acid sequences with homologous sequences in *S. cerevisiae*, *P. aeruginosa* and *E. coli* and with *E. coli* PabA and PabB. Sequences were aligned by the MULTALIN method (Corpet 1988). Conserved residues are boxed. The number at the end of each block refers to the position of the *H. volcanii* sequence.

and eukaryotic TrpG sequences by Crawford and coworkers (Essar et al. 1990a). Among the invariant positions identified, the catalytic residue Cys-84, which has been shown to form a covalent glutaminyl catalytic intermediate (Tso et al. 1980; Zalkin et al. 1984), is found at position 88 of the *Hf. volcanii* TrpG.

The subunits of *Hf. volcanii* anthranilate synthase (TrpE and TrpG) show similarities to the products of *E. coli pabA* and *pabB* genes, the subunits of *para*-aminobenzoate synthase of the folate pathway (figure 21). These genes are believed to be products of ancient duplications of *trpE* and *trpG* (Crawford 1989; Crawford and Milkman 1990).

All four genes in the *trpDFEG* cluster are preceded by good ribosome-binding sequences: GGUGAU for *trpD*, GGAGG for *trpF*, AGGAGGU for *trpE* and AGGAGG for *trpG*. The sequence of the 3'-terminus of the *Hf. volcanii* small subunit ribosomal RNA is , as mentioned before, 3'-UCCUCCACUA....

Codon usage of the genes in the *trpDFEG* cluster is similar to that of the *trpCBA* cluster; that is, it shows an extreme bias towards codons with C or G in the third position. Codons ending with C or G account for over 90% of the alanine, glycine and valine residues in the two gene clusters. (Codons for each of these amino acid residues are variable only in their third position, ending in all 4 different bases.)

All seven Trp enzymes of *Hf. volcanii* contain 17 to 20% acidic residues (aspartic acid and glutamic acid residues). This is consistent with the finding that most proteins in halophilic archaebacteria are acidic (Eisenberg and Wachtel 1987).

3. Expression and regulation of *trp* genes

The fact that *trpC, B* and *A* overlap by their stop and start codons suggests that these genes are transcribed as a unit from the same promoter. Similarly, *trpD, F, E* and *G* are probably transcribed together. Northern analyses are required to verify operon structure. Incidentally, ribosome-binding sequences precede all the *Hf. volcanii trp* genes

except for *trpC*. Examination of the sequence upstream and downstream of the two gene clusters revealed obvious secondary structures which may represent signals for transcription initiation and termination (figure 22A and B).

About 40 bp in front of the translation start (AUG) of *trpD* is the sequence ATTTGTA, which resembles the promoter consensus. The distance between this sequence and the transcription start site has not been determined.

Approximately 30 bp upstream from the start codon (AUG) of *trpC* is the putative halobacterial promoter sequence TTATGTA, followed by two nearly perfect direct repeats 10 bp apart (see figure 22A). The sequence within 100 bp further upstream can potentially be folded into stable secondary structures (figure22). Remarkably, the sequence between 76 and 95 bp from the AUG can form alternate stable stem and loop structures with the adjacent sequences immediately upstream or downstrem (from position 127 to 112 and from 67 to 53 upstream of the AUG). Site-specific mutagenesis experiments are required to assess any role of these sequences in gene expression or its control.

Many predictions regarding regulation of gene expression could be made, based on the primary sequence of the *trp* genes and their flanking regions. Experimentally, I have just begun to analyze the regulation of these *trp* gene clusters. At this point, several qualitative conclusions can be drawn from the limited data.

The expression of *trp* genes is regulated in *Hf. volcanii*. Primer extension assays show a signal absent only in cells supplied with tryptophan (figure 23). Interestingly, the segment (leu-leu-glu-ser) in the eubacterial α subunit of anthranilate synthase, essential for feedback inhibition by tryptophan, is conserved in *Hf. volcanii* (refer to section on *trpE* above).

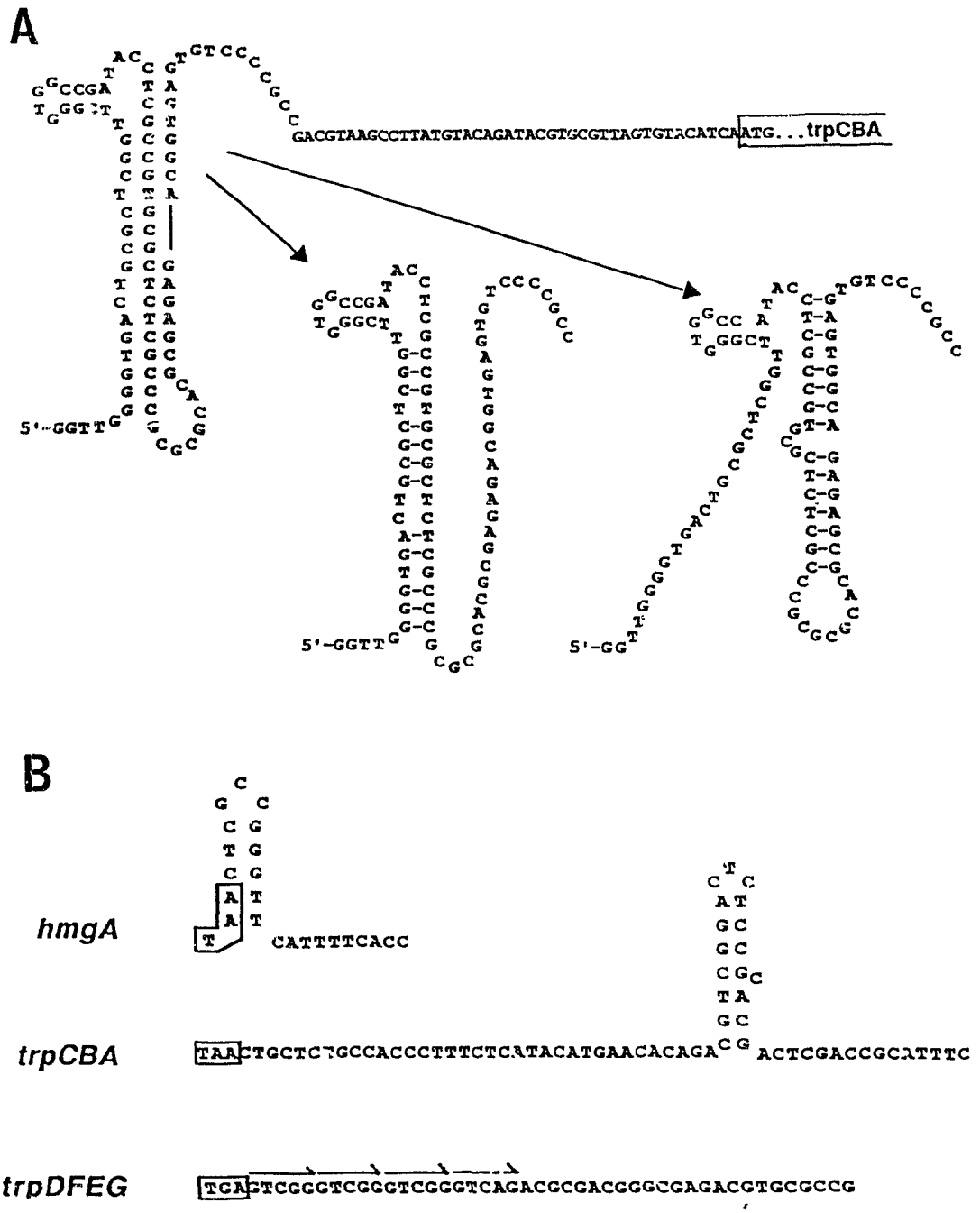


Figure 22. Putative secondary structures flanking *hmgA*, *trpCBA* and *trpDFEG*. Sequence between 53 and 127 bp upstream of the translation start codon (AUG) of *trpCBA* can be folded into two alternative secondary structures (A). Putative terminator structures are shown in (B).

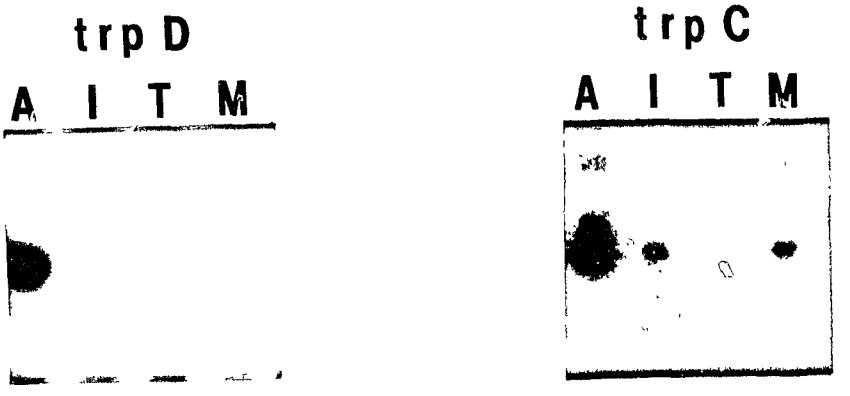


Figure 23. Transcription of *trp* genes in response to tryptophan and tryptophan biosynthetic intermediates. Levels of *trpD* and *trp C* transcripts, under various growth conditions, were estimated by primer extension assays. RNA samples (15 μg), isolated from *Hf. volcanii* cultures grown on minimal liquid medium (M), or on medium supplemented with 25 μg/ml of anthranilic acid (A) or 10 μg/ml of indole (I) or 25 μg/ml of L-tryptophan (T), were subjected to primer extension assays, using oligonucleotides complementary to sequences near the 5'-terminus of the *trpD* or of the*trpC* mRNA.

Transcription of the two *trp* gene clusters seems to be coordinately activated by anthranilic acid (figure 23). Data on hand cannot distingish whether anthranilic acid directly activates one or both gene clusters, or indirectly regulates gene expression via activator proteins. Also, one cannot overlook the possibility of the activation of *trpCBA* expression by a gene product of the *trpDFEG* cluster. In *Pseudomonas aeruginosa*, an activator protein regulates *trpBA* expression in response to the level of indoleglycerol phosphate, the substrate of the tryptophan synthase reaction (Chang et al. 1989). Indole did not shut off transcription of either cluster (figure 23).

A variety of tryptophan analogs, commonly used in studies on the regulation of trp biosynthesis in eubacterial and eukaryotic systems, were tested on *Hf. volcanii*. Certain analogs did inhibit growth of *Hf. volcanii*. (They include 4-methyltryptophan, 5-methyltryptophan, 6-methyltryptophan, 7-methyltryptophan, 5-fluorotryptophan, 6-fluorotryptophan, 7-azatryptophan and indole acrylic acid.) Frequencies of resistant mutations have been determined for several of these analogs.

4. Evolutionary considerations

The *Hf. volcanii* TrpB and TrpA amino acid sequences are as different from all sequences we have compared (including that of the methanogen) as the eubacterial sequences from the eukaryotic ones (Table 4). The same observation applies to TrpC, TrpD, TrpF, TrpE and TrpG, except *Methanococcus* sequences are not available for comparison. The rate of amino acid substitution outside the invariable positions is too high to allow the use of these sequences for assessing phylogenetic relatedness of organisms. However, the existence of conserved regions in all of the seven Trp enzymes indicates the presence of these polypeptides before the divergence of archaebacteria, eubacteria and eukaryotes.

All seven Trp enzymes of *Hf. volcanii* are encoded by separate genes, arranged in two gene clusters: *trpDFEG* (figure 20) and *trpCBA* (figure 18). Both *trp* gene clusters

are unusual in their organization. These arrangements have little in common with the eukaryotic and the eubacterial arrangements observed to date, execpt that the E,G order and the B,A order are found in eubacteria (Table 5).

In all known cases, *trpE* , which encodes the first enzyme of the tryptophan biosynthetic pathway, is either solitary or is the first gene in a polycistronic unit (Crawford 1989; Crawford and Milkman 1990). The significance of this gene arrangement for gene regulation is unknown. *Hf. volcanii* provides the only example of *trpE* in the middle of a gene cluster.

As in eubacteria, the two domains of *Hf. volcanii* tryptophan synthase are encoded by transcriptionally linked but separate genes. The *trpB* and *A* genes are preceded by *trpC*-- an arrangement unique to *Hf. volcanii*. Methanococcal and eubacterial *trpB* and *trpA* genes either follow *trpF* or exist as a unit alone (Sibold and Herniquet 1988; Yanofsky 1984). The *trpC* gene is either fused to or transcriptionally linked to *trpF*, *trpG* or *trpD* (Yanofsky 1984; Crawford 1989; Crawford and Milkman 1990).

The difference in gene arrangement between *Hf. volcanii* (C·B·A) and *M. voltae* (F·B·A) is perhaps not surprising, since no unified arrangements are common to all eukaryotes or all eubacteria (except for the linkage of *trpB* and *A* in prokaryotes). Diversity is apparent even within the γ subdivision of purple eubacteria (Table 5): *E. coli* has a single *trp* operon, but the *trp* genes of *Pseudomonas* exist in 4 transcriptional units and respond to 3 distinct regulatory mechanisms (Essar et al. 1990a, b).

Nevertheless, *trp* genes tend to be fused or transcriptionally linked. Unless linkage provides tremendous selective advantage, it is difficult to envision seven separate genes repeatedly seeking out one another to form fusion units and operons -- the

Table 5. Organization of *trp* genes in various organisms.

EUBACTERIA								
Purple bacteria:								
α subdivision								
Rhizobium and relatives			E-----G		D	C	F B A	
Zymomonas mobilis					D	C	F B A	
Caulobacter crescentus			E		D	C	F B A	
β subdivision								
Pseudomonas acidovorans			E	G	D	C	F B A	
γ subdivision								
Escherichia coli and relatives			E-----G		D	C-----F	B A	
Serratia and relatives			E	G	D	C-----F	B A	
Pseudomonas aeruginosa and relatives			E	G	D	C	F B A	
Acinetobacter calcoaceticus			E	G	D	C	F B A	
Gram-positive eubacteria:								
low (G+C) subdivision								
Bacillus subtilis and relatives			E		D	C	F B A	G
Lactobacillus casei					D	C	F B A	
high (G+C) subdivision								
Brevibacterium lactofermentum			E	G	D	C-----F	B A	
Spirochetes:								
Spirochaeta aurantia			E					
Leptospira biflexa			E	G				
ARCHAEBACTERIA								
Methanococcus voltae							F B A	
Haloferax volcanii	D	F	E	G		C	B A	
EUKARYOTES								
Saccharomyces cerevisiae			E	G----------C		F	A-----B	D
Neurospora crassa			E	G----------C-----F			A-----B	D
Aspergillus nidulans				G----------C-----F			A-----B	
Coprinus cinereus							A-----B	
Penicillium chrysogenum				G----------C-----F				
Arabidopsis thaliana							B	

Fused genes are jointed by dashes; transcriptional units are underlined.

ancestral set of *trp* genes must have been linked, and subsquently [illegible] shuffled, fused and rearranged in ways only restricted by the needs of the organism.

III. Genetic mapping

In preparing a bottom-up restriction map for *Hf. volcanii*, a minimally overlapping set of 154 cosmids covering >95% of the genome was assembled in this laboratory (Charlebois et al. 1989). Although cosmids cannot replicate autonomously in *Hf. volcanii*, cosmid DNA (individually or in pools) can be used to transform *Hf. volcanii* auxotrophs to prototrophy (Conover and Doolittle, 1990), presumably through homologous recombination. Such transformation allows us to map mutations, and to localize genes on the 3.8 million base pair genome of *Hf. volcanii* to within 1%.

Genetic mapping of *trp* genes. The 29 Trp auxotrophs were used to develop strategies for locating genes within the developing genome map. A pooling scheme (in which each of the 154 cosmids is represented in two separate pools) allowed the deduction, from a single transformation experiment with 25 cosmid DNA pools and 2 control samples, of the individual cosmid which bears the wild-type allele for any auxotrophic mutation. Of the five Trp$^-$ mutants initially tested, two mapped unequivocally to cosmid 452 -- the two cosmid pools which transformed WFD131 and WFD287 to prototrophy have only cosmid 452 in common. The remaining three mapped to cosmids 488 and A159. These latter two cosmids overlap by 33 kbp, and the mutations must lie within the overlap. We then transformed each of our 29 auxotrophs with cosmid 452 and A159. All mutants which grow only on tryptophan, as well as the indole-utilizing WFD135, could be transformed with cosmid 452, while the rest, which utilized indole or anthranilate, mapped to cosmid A159. These locations are separated by a minimum of 120 kbp according to the physical mapping data which has linked the cosmid set of 154 into a few large map fragments (R. Charlebois, L. Schalkwyk, J. Hofman, and

W.F.Doolittle, unpublished data). Similar mapping experiments have placed guanosine loci on cosmids near cosmids 488/A159 and tyrosine loci on a cosmid close to cosmid 452.

Plasmid clones bearing *trp* genes can be labeled and used as probes against dot blot filters bearing DNA from all 154 cosmids of the minimal set in an ordered array. This provides an independent method of assigning loci on the physical map and establishing linkage between markers. All cosmid assignments made by transformation could be confirmed by this procedure. Mutant WFD246, which reverted at too high a frequency to be mapped by transformation with cosmid DNA (which is subject to restriction in *Hf. volcanii*) could only be mapped in this way. All 29 Trp^- mutations so far obtained can be mapped to one of two unlinked regions of the *Hf. volcanii* genome, using these independent physical and genetic methods (Table 3).

Mapping auxotrophic mutations. Annalee Cohen of our laboratory has been actively generating auxotrophic mutants using the EMS mutagenesis procedure described above. She has identified the nutritional requirements of 254 amino acid requiring- or nucleotide requiring- mutants. Using the strategies developed for mapping the *trp* genes we (Dr. Cohen and I) proceeded to map these mutations either directly by transforming with pooled cosmid clones from the minimal set, or indirectly with plasmid libraries, which requires subsequent hybridization of clones to DNA dot blots of the cosmid set.

Out of the two hundred mutants transformed, we were able to unambiguously place 135 mutations. For the rest, mapping was hampered either because the wild-type locus fell in one of the gaps which the cosmid set does not cover (evident when genomic DNA from wild-type cells, but not DNA from cosmid pools, successfully transforms the mutant to prototrophy) or by high reversion frequencies (so that successful transformation is masked). Cosmid DNA cannot replicate autonomously in *Hf. volcanii,* requiring recombination events for transformation. In addition, cosmid DNA is subjected to host restriction, reducing the transformation efficiency to roughly 10^2 to 10^3 transformants per

μg of DNA per 10^8 regenerants. Each of the mapped mutations was verified by transforming with individual cosmids and proper controls were included in these confirmatory transformation experiments to recheck the initial assignment of nutritional requirements. Table 6 summarizes the mapped loci, the result of over 4000 transformation experiments.

Table 6. Auxotrophic mutations mapped by cosmid transformation experiments.

WFD strain	allele	requires	positive pools	transformation with selected cosmids	cosmid identified
283	*ser*-101	serine	10, 25		499
276	*ilv*-105	val+ile		110,499,501	501
512	*ilv*-111	val+ile		110,499,501	501
514	*ilv*-112	val+ile	10, 24		501
286	*thr*-101	threonine		A99,126,496	126
318	*thr*-102	threonine	6, 17		126
322	*thr*-103	threonine	6, 17		126
504	*thr*-105	threonine		A99,126,496	126
161	*met*-102	methionine	10, 24		501
285	*met*-103	methionine	13, 16		A78 (high reversion)
580	*gln*-102	glutamine	1, 18		460
595	*gln*-103	glutamine	1, 18		460
126	*aro*-101	early aromatic	12, 19		G411
137	*tyr*-102	phe+tyr	13, 20		410
155	*tyr*-103	phe+tyr	12, 19		G411
177	*tyr*-104	phe+tyr	12, 19		G411
596	*tyr*-117	phe+tyr	10,24		G283
35	*trp*-101	tryptophan		452,A159	452
36	*trp*-102	tryptophan		452,A159	452
37	*trp*-103	anth,indole		452, A159	A159
39	*trp*-105	tryptophan		452,A159	452
41	*trp*-106	tryptophan		452,A159	452
131	*trp*-107	tryptophan	13,22	452,A159	452
135	*trp*-108	indole		452,A159	452
163	*trp*-109	indole		452,A159	A159
167	*trp*-110	indole		452,A159	A159
178	*trp*-111	indole		452,A159	A159
186	*trp*-112	indole		452,A159	A159
192	*trp*-113	indole		452,A159	A159
204	*trp*-114	indole		452,A159	A159
207	*trp*-115	indole		452,A159	A159
216	*trp*-116	indole		452.A159	A159
240	*trp*-117	indole		452,A159	A159
246	*trp*-118	indole		452,A159	488,A159 (also by hyb.)
258	*trp*-119	indole		452,A159	A159
261	*trp*-120	indole	4,19,20	452,A159	488, A159
284	*trp*-121	indole		452,A159	A159
287	*trp*-122	tryptophan	13,22	452,A159	452
292	*trp*-123	indole		452,A159	A159
300	*trp*-124	indole	4,19,20	452,A159	488,A159
347	*trp*-125	tryptophan		452,A159	452
369	*trp*-126	tryptophan		452,A159	452
531	*trp*-127	indole	4,19,20	452,A159	488, A159
543	*trp*-128	tryptophan		452,A159	452
564	*trp*-129	indole		452,A159	A159
567	*trp*-130	indole		452,A159	A159

Table 6 continues...

WFD strain	allele	requires	positive pools	transformation with selected cosmids	cosmid identified
140	*lys*-101	lysine	7, 18		H680
144	*lys*-102	lysine		H3,H680,530	H680,530
157	*lys*-103	lysine		H3,H680,530	H680
254	*lys*-105	lysine		H3,H680,530	H680
270	*lys*-107	lysine		H3,H680,530	H680
309	*lys*-108	lysine		H3,H680,530	H680,530
327	*lys*-109	lysine		H3,H680,530	H680,530
329	*lys*-110	lysine		H3,H680,530	H680
364	*lys*-111	lysine		H3,H680,530	H680,530
368	*lys*-112	lysine		H3,H680,530	H680,530
383	*lys*-114	lysine		H3,H680,530	H680
400	*lys*-115	lysine	7,18		H680
586	*lys*-116	lysine		H3,H680,530	H680,530
105	*his*-102	histidine	15,24		H10
112	*his*-103	histidine		32,G171,G411,499,501	499
213	*his*-105	histidine		32,G171,G411,499,501	499
217	*his*-106	histidine		32,G171,G411,499,501	32,G171
220	*his*-107	histidine	15,24		H10
255	*his*-109	histidine		32,G171,G411,499,501	G171,G411
256	*his*-110	histidine		32,G171,G411,499,501	G171,G411
259	*his*-111	histidine		32,G171,G411,499,501	499
268	*his*-112	histidine	5,19		347
274	*his*-113	histidine		32,G171,G411,499,501	499
545	*his*-123	histidine		32,G171,G411,499,501	499
556	*his*-124	histidine		32,G171,G411,499,501	G171,G411
572	*his*-125	histidine	10, 25		499
597	*his*-126	histidine		32,G171,G411,499,501	499
L7	*leu*-101	leucine		pL7-N (cloned)	56 by hyb.
L12	*leu*-102	leucine		pL7-N (cloned)	56 by hyb.
L18	*leu*-103	leucine		pL18-N (cloned)	56 by hyb.
130	*leu*-104	leucine		56, 535	56
136	*leu*-105	leucine		56, 535	56
209	*leu*-107	leucine		56, 535	56
223	*leu*-109	leucine		56, 535	56
227	*leu*-110	leucine		56, 535	56
234	*leu*-112	leucine		56, 535	56
241	*leu*-113	leucine		56, 535	56
272	*leu*-114	leucine		56, 535	56
314	*leu*-115	leucine		56, 535	56
357	*leu*-117	leucine		56, 535	56
406	*leu*-119	leucine	4, 22		56
541	*leu*-122	leucine		56, 535	56 (high background)
569	*leu*-123	leucine	4, 22		56
587	*leu*-124	leucine		56, 535	56

Table 6 continues...

WFD strain	allele	requires	positive pools	transformation with selected cosmids	cosmid identified	
108	*gua*-101	guanosine		97, 326	97	
146	*gua*-104	guanosine		97, 326	97	
185	*gua*-112	guanosine	4, 16		97	
199	*gua*-113	guanosine		97,326	97	
219	*gua*-115	guanosine		97, 326	97	
235	*gua*-116	guanosine	4, 16		97	
245	*gua*-119	guanosine	9,22,24		101,C138	
281	*gua*-125	guanosine		97, 326	97	
312	*gua*-127	guanosine		97, 326	97	
344	*gua*-131	guanosine		97, 326	97	
378	*gua*-132	guanosine	5,6,17		G329,497	
415	*gua*-138	guanosine		97, 326	97	
517	*gua*-140	guanosine		97, 326	97	
526	*gua*-141	guanosine		97, 326	97	
539	*gua*-142	guanosine		97, 326	97	
532	*pur*-128	adenine	7,19		530	
169	*pyr*-101	uracil	10, 22		266	haze
243	*pyr*-102	uracil	10, 22		266	low numbers
299	*pyr*-103	uracil	1,16		478	
337	*pyr*-104	uracil	4,24		166	
143	*pyr*-106	uracil+arg	4, 21		531	
304	*pyr*-107	uracil+arg	12, 23,24		37,128	
536	*pyr*-109	uracil+arg	12,23,24,25		128,H37,208	
551	*pyr*-110	uracil+arg	1,16		478	
552	*pyr*-111	uracil+arg	12,24		H37	
T9R12	*arg*-101 (+*trp*-101)	arg+trp		arg clone	21	
127	*arg*-102	arginine		21,247,545	247, 21	
133	*arg*-103	arginine		21,247,545	247, 21	
138	*arg*-104	arginine		21,247,545	247, 21	
174	*arg*-106	arginine		21,247,545	247, 21	
197	*arg*-107	arginine		21,247,545	247, 21	
321	*arg*-108	arginine		21,247,545	247, 21	
345	*arg*-109	arginine		21,247,545	21	
360	*arg*-110	arginine		21,247,545	(247), 21	
362	*arg*-111	arginine		21,247,545	21	
375	*arg*-112	arginine		21,247,545	247, 21	
388	*arg*-113	arginine		21,247,545	247, 21	
407	*arg*-114	arginine		21,247,545	(247), 21	
506	*arg*-115	arginine		21,247,545	247, 21	
511	*arg*-116	arginine		21,247,545	247, 21	
547	*arg*-117	arginine		21,247,545	(21)	
577	*arg*-118	arginine		21,247,545	247, 21	
591	*arg*-119	arginine		21,247,545	21	
599	*arg*-120	arginine		21,247,545	247,21	

Linkage of alleles. Organization of the mapped alleles on the *Hf. volcanii* chromosome will become clear when the physical map of the chromosome is completed and made available. However, the transformation mapping data have already provided much linkage information. All 17 *leu* loci are linked. The 29 Trp mutations map to two unlinked clusters. Histidine mutations, so far, have been mapped to five separate cosmids. There are still more His auxotrophs to be mapped. All 13 Lys mutations map together. The 19 Arg mutations are rescued by two overlapping cosmids. Fine scale transformation experiments and DNA sequencing revealed several contiguous *arg* genes (K. Conover, personal communication). Pyrimidine mutants map to several locations. A uracil-requiring and a uracil plus arginine-requiring mutant are transformed to prototrophy by the same cosmid, suggesting the linkage of at least two *pyr* genes. With the purine auxotrophs, guanosine-requiring mutants are rescued by at least three different cosmids; only one of the many adenine auxotrophs was mapped. Mutants requiring phenylalanine plus tyrosine map to three different cosmids. One of these three cosmids also transforms a strain defective early in the aromatic pathway. Two separate cosmids could rescue methionine auxotrophs. One of these cosmids (cosmid 501) also transforms mutants requiring valine plus isoleucine to prototrophy. Cosmid 501 overlaps cosmid 499 which contains a serine allele. The relative locations of these and the other alleles listed (in Table 6) will provide valuable information on the organization of an archaebacterial chromosome.

IV. Conclusions

The significance of archaebacteria has been recognized for over a decade. Genetic analysis of this unique assemblage of organisms is long overdue. Using the transformation technology developed in our laboratory (Cline and Doolittle 1987), I built vectors for genetic transfer and adapted a variety of genetic methods (commonly used in *E. coli* and yeast) to study *Hf. volcanii* at the DNA level.

Shuttle vectors, marked by mevinolin resistance in *Hf. volcanii* and ampicillin resistance in *E. coli*, efficiently transform *Hf. volcanii* (typically, 10^8 resistant transformants per μg plasmid DNA prepared from *Hf. volcanii*, and 10^4 per μg of DNA made from *E. coli*). These vectors, and ones bearing a cloned insert, could be isolated from *Hf. volcanii*, and used directly to transform *E. coli* to ampicillin resistance. *In vivo* rearrangement of the shuttle vectors has not been observed -- even after shuttling from *E.coli* to *Hf. volcanii* and back to *E.coli* they are stably maintained in either host. DNA libraries in pWL102 have been instrumental in the isolation of biosynthetic genes, such as the *trp* genes. Furthermore, a cloned DNA fragment containing the *trpCBA* genes (on the plasmid pDWT1) was altered *in vitro*, prepared in *E. coli* and reintroduced into auxotrophic strains of *Hf. volcanii* for functional analysis. These applications demonstrate that the vectors and transformation are functional.

The nature of mevinolin resistance has been investigated. Gene amplification, as well as an up-promoter mutation, lead to overexpression of the HMG CoA reductase gene. These mechanisms allow *Hf. volcanii* to grow in the presence of mevinolin, an inhibitor of HMG CoA reductase. The promoter mutation represents the first genetic evidence of an archaebacterial promoter. The amino acid sequence of *Hf. volcanii* reductase is very similar to eukaryotic HMG CoA reductases, but has little in common with the homolog in *Pseudomonas mevalonii*. Alignment of the archaebacterial, eukaryotic and the eubacterial reductase sequences defines two regions of sequence conservation; conservation may

reflect functional importance. In addition, the *Hf. volcanii* mevinolin resistance gene may provide a useful marker for the development of genetic systems in other archaebacteria.

Transformation experiments using non-replicating constructs containing the mevinolin resistance marker and the *trpB* flanking sequences have shown that strategies for gene replacement and methods for deleting specific segments of the chromosome are applicable to *Hf. volcanii*.

The shuttle vector/transformation system permitted the isolation of two clusters of unusually arranged *trp* genes from *Hf. volcanii*. Genes for all 7 enzymatic functions for tryptophan biosynthesis have been identified by deletion analysis and DNA sequencing. Products of these *Hf. volcanii trp* genes are equally different from their eubacterial and eukaryotic homologs. A preliminary examination of *trp* gene expression suggests coordinated regulation of the two gene clusters

Strategies for genetic mapping were first tested on 29 Trp auxotrophs and then applied to map over a hundred more auxotrophic mutations. Distribution of the mapped alleles has given a preview of the overall organization of the archaebacterial chromosome, and an emerging genetic map. Linkage of functionally related genes is evident. In addition, DNA sequencing revealed gene organization at the operon level -- overlapping genes implies an eubacterial-like operon structure.

In summary, the advances in genetic methods for *Hf. volcanii* have promoted this organism to a model host for archaebacterial genetics. The accumulating genetic tools, information and experience will facilitate the genetic dissection of specific biological systems, as well as the determination of genetic organization in archaebacteria.

REFERENCES

Alberts, A. W., Chen, J., Kuron, G., Hunt, V., Huff, J., Hoffman, C., Rothrock, J., Lopez, M., Joshua, H., Harris, E., Patchett, A., Monaghan, R., Currie, S., Stapley, E., Albers-Schonberg, G., Hensens, O., Hirshfield, J., Hoogsteen, K., Liesch, J., and Springer, J. 1980. Mevinolin: A highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme A reductase and a cholesterol-lowering agent. Proc. Natl. Acad. Sci. USA 77, 3957-3961.

Auer, J., Lechner, K. and Bock, A. 1989a. Gene organization and structure of two transcriptional units from *Methanococcus* coding for ribosomal proteins and elongation factors. Can. J. Microbiol. 35, 200-204.

Auer, J., Spicker, G. and Bock, A. 1989b. Organization and structure of the *Methanococcus* transcriptional unit homologous to the *Escherichia coli* "spectinomycin operon". J. Mol. Biol. 209, 21-36.

Baier, G., Piendl, W., Redl, B. and Stoffler, G. 1990. Structure, organization and evolution of the L1 equivalent ribosomal protein gene of the archaebacterium *Methanococcus vannielii.* Nucl. Acids Res. 18, 719-724.

Basson, M.E., Thorsness, M., Finer-Moore, J., Stroud, R.M. and Rine, J. 1988. Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductase, the rate-limiting enzyme of sterol biosynthesis. Mol. Cell. Biol. 8, 3797-3808.

Beach, M.J. and Rodwell, V.W. 1989. Cloning, sequencing and overexpression of *mvaA*, which encodes *Pseudomonas mevalonii* 3-hydroxy-3-methylglutaryl coenzyme A reductase. J. Bacteriol. 171, 2994-3001.

Beckler, G.S. and Reeve, J.N. 1986. Conservation of primary sequence in the *hisI* gene of the archaebacterium, *Methanococcus vannielii*, the eubacterium *Escherichia coli* and the eukaryote *Saccharomyces cerevisiae.* Mol. Gen. Genet. 204, 133-140.

Belay, N., Sparling, R. and Daniels, L. 1984. Dinitrogen fixation by a thermophilic methanogenic bacterium. Nature 312, 286-288.

Bertani, G. and Baresi, L. 1987. Genetic transformation in the methanogen *Methanococcus voltae* PS. J. Bacteriol. 169, 2730-2738.

Blanck, A. and Oesterhelt, D. 1987. The halo-opsin gene. II. Sequence, primary structure of halorhodopsin and comparison with bacteriorhodopsin. EMBO J. 6, 265-273.

Betlach, M., Friedman, J., Boyer, H. and Pfeifer, F. 1984. Characterization of a halobacterial gene affecting bacterio-opsin gene expression. Nucl. Acids Res. 12, 7949-7959.

Betlach, M., Friedman, J., Pfeifer, F. and Boyer, H.W. 1983. Bacterio-opsin mutants of *Halobacterium halobium.* Proc. Natl. Acad. Sci. USA 80, 1416-1420.

Betlach, M., Leong, D.M., Boyer, H.W. 1986. Bacterio-opsin gene expression in *Halobacterium halobium.* Syst. Appl. Microbiol. 7, 83-89.

Betlach, M.C., Shand, R.F. and Leong, D.M. 1989. Regulation of the bacterio-opsin gene of a halophilic archaebacterium. Can. J. Microbiol. 35, 134-140.

Blanck, A., Oesterhelt, D., Ferrando, E., Schegk, E.S. and Lottspeich, F. 1989. Primary structure of sensory rhodopsin I, a prokaryotic photoreceptor. EMBO J. 8, 3963-3971.

Blaseio, U. and Pfeifer, F. 1990. Transformation of *Halobacterium halobium*: development of vectors and investigation of gas vesicle synthesis. Proc. Natl. Acad. Sci. USA 87, 6772-6776.

Bokranz, M., Baumner, G., Allmansberger, R., Ankel-fuchs, D. and Klein, A. 1988. Cloning and characterization of methyl coenzyme M reductase genes from *Methanobacterium thermoautotrophicum*. J. Bacteriol. 170, 568-577.

Bollschweiler, C., Kuhn, R. and Klein, A. 1985. Non-repetitive AT-rich sequences are found in intergenic regions of *Methanococcus voltae* DNA. EMBO J. 4, 805-809.

Bowen, T.L. and Whitman, W.B. 1987. Incorporation of exogenous purines and pyrimidies by *Methanococcus voltae* and isolation of analog-resistant mutants. Appl. Environ. Microbiol. 53, 1822-1826.

Brown, J. and Reeve, J.N. 1985. Polyadenylated, noncapped RNA from the archaebacterium *Methanococcus vannielii*. J. Bacteriol. 162, 909-917.

Brown, J. and Reeve, J.N. 1986. Polyadenylated RNA isolated from archaebacterium *Halobacterium halobium*. J. Bacteriol. 166, 686-688.

Brown, J. W., Daniels, C. J. and Reeve, J. N. 1989. Gene structure, organization, and expression in archaebacteria. CRC Crit. Rev. Microbiol. 16, 287-338.

Brown, J.W., Thomm, M., Beckler, G.S., Frey, G., Stetter, K.O. and Reeve, J.N. 1988. An RNA polymerase binding site and transcription initiation of the *hisA* gene of *Methanococcus vannielii*. Nucleic Acids Res. 16, 135-150.

Brown, M.S. and Goldstein, J.L. 1980. Multivalent feedback regulation of HMG CoA reductase, a control mechanism coordinating isoprenoid synthesis and cell growth. J. Lipid Res. 21, 505-517.

Bucher, P. and Trifonov, E.N. 1986. Compilation and analysis of eukaryotic POL II promoter sequences. Nucl. Acids Res. 14, 10009-10026.

Cabrera, J.A., Bold, J., Shields, P.E., Havel, C.M. and Watson, J.A. 1986. Isoprenoid synthesis in *Halobacterium halobium*: modulation of 3-hydroxy-3-methylglutaryl coenzyme A concentration in response to mevalonate availability. J. Biol. Chem. 261, 3578-3583.

Chang, M., Hadero, A. and Crawford, I.P. 1989. Sequence of the *Pseudomonas aeruginosa trpI* activator gene and relatedness of *trpI* to other prokaryotic regulatory genes. J. Bacteriol. 171, 172-183.

Chant, J. and Dennis, P.P. 1986. Archaebacteria: transcription and processing of ribosomal RNA sequences in *Halobacterium cutirubrum*. EMBO J. 5, 1091-1097.

Charlebois, R. L., Hofman, J. D., Schalkwyk, L. C., Lam, W. L. and Doolittle, W. F. 1989. Genome mapping in halobacteria. Can. J. Microbiol. 35, 21-29.

Charlebois, R. L., Lam, W. L., Cline, S. W.and Doolittle, W. F. 1987. Characterization of pHV2 from *Halobacterium volcanii* and its use in demonstrating transformation of an archaebacterium. Proc. Natl. Acad. Sci. USA 84, 8530-8534.

Charlebois, R.L. and Doolittle, W.F. 1989. Transposable elements and genome structure in Halobacteria. In *Mobile DNA* (Berg, D.E. and Howe, M.H., ed), pp. 297-307, American Society for Microbiology, Washington, D.C.

Chin, D.J., Gil, G., Russell, D.W., Liscum, L., Luskey, K.L., Basu, S.K., Okayama, H., Berg, P., Goldstein, J.L. and Brown, M.S. 1984. Nucleotide sequence of 3-hydroxy-3-methylglutaryl coenzyme A reductase, a glycoprotein of endoplasmic reticulum. Nature 308, 613-617.

Chomczynski, P. and Sacchi, N. 1987. Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162, 156-159.

Chung, C.T. and Miller, R.H. 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucl. Acids Res. 16, 3580.

Cline, S. W. and Doolittle, W. F. 1987. Efficient transfection of the archaebacterium *Halobacterium halobium*. J. Bacteriol. 169, 1341-1344.

Cline, S. W., Lam, W. L., Charlebois, R. L., Schalkwyk, L. C. and Doolittle, W. F. 1989a. Transformation methods for halophilic archaebacteria. Can. J. Microbiol. 35, 148-152.

Cline, S., Schalkwyk, L.C. and Doolittle, W.F. 1989b. Transformation of the archaebacterium *Halobacterium volcanii* with genomic DNA. J. Bacteriol. 171, 4987-4991.

Conover, R.K. and Doolittle, W.F. 1990. Characterization of a gene involved in histidine biosynthesis in *Halobacterium* (*Haloferax*) *volcanii*: isolation and rapid mapping by transformation of an auxotroph with cosmid DNA. J. Bacteriol. 172, 3244-3249.

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering Nucl. Acids Res. 16, 10881-10890.

Cozzarelli, N.R., Gerrard, S.P., Schlissel, M., Brown, D.D. and Bogenhagen, D.F. 1983. Purified RNA polymerase III accurately and efficiently terminates transcription of 5S RNA genes. Cell 34, 829-835.

Cram, D.S., Sherf, B.A., Libby, R.T., Mattaliano, R.J., Ramachandran, K.L. and Reeves, J.N. 1987. Structure and expression of the genes mcr BDCGA, which encode the subunits of component C of methyl coenzyme M reductase in *Methanococcus vannielii*. Proc. Natl. Acad. Sci. USA 84, 3992-3996.

Crawford, I.P. 1989. Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. 43, 567-600.

Crawford, I.P. and Milkman, R. 1990. Orthologous and paralogous divergence, reticulate evolution, and lateral gene transfer in bacterial *trp* genes. *In* Evolution at the Molecular level. R.K. Selander, A.G. Clark and T.S. Whittman ed., Sinauer Assoc. Inc. Publisher. Massachusetts.

Cue, D., Beckler, G.S., Reeve, J.N., and Konisky, J. 1985. Structure and sequence divergence of two Archaebacterial genes. Proc. Natl. Acad. Sci. USA. 82, 4207-4211.

Culham, D.E. and Nazar, R.N. 1989. Multiple termination sites in an unlinked 5S rRNA operon in the archaebacterium *Thermococcus celer* . Mol. Gen. Genet. 216, 412-416.

Daniels, C.J., Douglas, S.E., McKee, A.H.Z. and Doolittle, W.F. 1986. Archaebacterial tRNA genes: structure and intron processing. In *Microbiology 1986* (Leive, L., ed) pp.349-355, ASM Publications, Wash. D.C.

Daniels, C.J., Gupta, R. and Doolittle, W.F. 1985. Transcription and excision of a large intron in the $tRNA^{Trp}$ gene of an archaebacterium, *Halobacterium volcanii*. J. Biol. Chem. 260, 3132-3134.

Daniels, L.L. and Wais, A.C. 1984. Restriction and modification of halophage S45 in *Halobacterium*. Curr. Microbiol. 10, 133-136.

Darr, S.C., Pace, B. and Pace, N.R. 1990. Characterization of ribonuclease P from the archaebacterium *Sulfolobus solfataricus*. J. Biol. Chem. 265, 12927-12932.

DasSarma, S. 1989. Mechanisms of genetic variability in *Halobacterium halobium*: the purple membrane and gas vesicle mutations. Can. J. Microbiol. 35, 65-72.

DasSarma, S., Damerval, T., Jones, J.G., and Tandeau De Marsac, N. 1987. A plasmid-encoded gas vesicle protein gene in a halophilic archaebacterium. Mol. Microbiol. 1, 365-370.

DasSarma, S., Halladay, J.T., Jones, J.G., Donovan, J., Giannasca, P.J. and Tandeau De Marsac, N. 1988. High-frequency mutations in a plasmid-encoded gas vesicle gene in *Halobacterium halobium*. Proc. Natl. Acad. Sci. USA 85, 6861-6865.

DasSarma, S., RajBhandary and Khorana, H.G. 1983. High-frequency spontaneous mutation in the bacterio-opsin gene in *Halobacterium halobium* is mediated by transposable elements. Proc. Natl. Acad. Sci. USA 80, 2201-2205.

DasSarma, S., RajBhandary and Khorana, H.G. 1984. Bacterio-opsin mRNA in wild-type and bacterio-opsin-deficient *Halobacterium halobium* strains. Proc. Natl. Acad. Sci. USA 81, 125-129.

Datta, P.K., Hawkins, L.K. and Gupta, R. 1989. Presence of an intron in elongator methionine-tRNA of *Halobacterium volcanii*. Can. J. Microbiol. 35, 189-193.

De Rosa, M. and Gambacorta, A. 1986. Lipid biogenesis in archaebacteria. System. Appl. Microbiol. 7, 278-285.

De Rosa, M., Gambacorta, A. and Gliozzi, A. 1986. Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. Microbiol. Rev. 50, 70-80.

DeLa Tour, C.B., Portemer, C., Nadal, M., Stetter, K.O., Forterre, P. and Duguet, M. 1990. Reverse gyrase, a hallmark of the hyperthermophilic archaebacteria. J. Bacteriol. 172, 6803-6808.

Dennis, P.P. 1986. Molecular biology of archaebacteria. J. Bacteriol. 168, 471-478.

Dugan, R.E. and Katiyar, S.S. 1986. Evidence for catalytic cysteine and histidine by chemical modification of β-hydroxy-β-methylglutaryl-coenzyme A reductase. Biochem. Biophys. Res. Comm. 141, 278-284.

Ebert,K. and Goebel, W. 1985. Conserved and varible regions in the chromosomal and extrachromosomal DNA of halobacteria. Mol. Gen. Genet. 200, 96-102.

Ebert, K., Hanke, C., Delius, H., Goebel, W. and Pfeifer, F. 1987. A new insertion element, ISH26, from *Halobacterium halobium*. Mol. Gen. Genet. 206, 81-87.

Ehrlich, M., Gama-Sosa, M.A., Carreira, L.H., Ljungdahl, L.G., Kuo, K.C. and Gehrke, C.W. 1985. DNA methylation in thermophilic bacteria: N^4-methylcytosine, 5-methylcytosine and N^6-methyladenine. Nucl. Acids Res. 13, 1399-1428.

Eisenberg, H. and Wachtel, E.J. 1987. Structural studies of halophilic proteins, ribosomes, and organelles of bacteria adapted to extreme salt concentrations. Annu. Rev. Biophys. Biophys. Chem.16, 69-92.

Ekiel, I., Smith, I. C. P. and Sprott, G. D. 1983. Biosynthetic pathways in *Methanospirillum hungatei* as determined by ^{13}C nuclear magnetic resonance. J. Bacteriol. 156, 316-326.

Ekiel, I., Sprott, G. D. and Smith, I. C. P. 1986. Mevalonic acid is partially synthesized from amino acids in *Halobacterium cutirubrum*: a ^{13}C nuclear magnetic resonance study. J. Bacteriol. 166, 559-564.

Essar, D.W., Eberly, L. and Crawford, I.P. 1990a. Evolutionary differences in chromosomal locations of four early genes of the tryptophan pathway in fluorescent Pseudomonads: DNA sequences and characterization of *Pseudomonas putida trpE* and *trpGDC*. J. Bacteriol. 172, 867-883.

Essar, D.W., Eberly, L., Han, C.-Y. and Crawford, I.P. 1990b. DNA sequences and characterization of four early genes of the tryptophan pathway in *Pseudomonas aeruginosa*. J. Bacteriol. 172, 867-883.

Feinberg, A.P. and Vogelstein, B. 1983. A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132, 6-13.

Feinberg, A.P. and Vogelstein, B. 1984. Addendum to "A technique for radio-labeling DNA restriction endonuclease fragments to high specific activity." Anal. Biochem. 137, 266-267.

Forterre, P., Elie, C., Sioud, M. and Hamal, A. 1989. Studies on DNA polymerases and topoisomerases in archaebacteria. Can J. Microbiol. 35, 228-232.

Forterre, P., Mirambeau, G., Jaxel, C., Nadal, M. and Duguet, M. 1985. High positive supercoiling *in vitro* catalysed by an ATP and polyethyleneglycol-stimulated topoisomerase from *Sulfolobus acidocaldarius*. EMBO J. 4, 2123-2128.

Forterre, P., Nadal, M., Elie, C., Mirambeau, G., Jaxel, C. and Duguet, M. 1986. Mechanisms of DNA synthesis and topoisomerization in archaebacteria- reverse gyration *in vitro* and *in vivo*. Syst. Appl. Microbiol. 7, 167-171.

Furter, R., Paravicini, G., Aebi, M., Braus, G., Prantl, F., Niederberger, P. and Hutter, R. 1986. The *TRP4* gene of *Saccharomyces cerevisiae*: isolation and structural analysis. Nucl. Acids Res. 14, 6357-6373.

Gehrmann, R., Henschen, A., Postulka, W. and Klink, F. 1986. Comparative studies on structure and function of archaebacterial elongation factors indicate the phylogenetic diversity of the urkingdom. System. Appl. Microbiol. 7, 115-122.

Geliebter, J. 1987. Dideoxynucleotide sequencing of RNA and uncloned DNA. Focus 9:1, 5-8.

Gerl, L. and Sumper, M. 1988. Halobacterial flagellins are encoded by a multigene family. J. Biol. Chem. 263, 13246-13251.

Gernhardt, P., Possot, O., Foglino, M., Sibold, L. and Klein, A. 1990. Construction of an integration vector for use in the archaebacterium *Methanococcus voltae* and expression of a eubacterial resistance gene. Mol. Gen. Genet. 221, 273-279.

Gertler, F.B., Chiu, C.Y., Richter-Mann, L. and Chin, D.J. 1988. Developmental and metabolic regulation of *Drosophila melanogaster* 3-hydroxy-3-methyl coenzyme A reductase. Mol. Cell. Biol. 8, 2713-2721.

Gogarten, J.P., Kibak, H., Dittrich, P., Taiz, L., Bowman, E.J., Bowman, B.J., Manolson, M.F., Poole, R.J., Date, T., Oshima, T., Konishi, J., Dena, K. and Yoshida, M. 1989. Evolution of the vacuolar H^+-ATPase: Implications for the origin of eukaryotes. Proc. Natl. Acad. Sci. USA 86, 6661-6665.

Goldstein, J.L. and Brown, M.S. 1990. Regulation of the mevalonate pathway. Nature 343, 425-430.

Green, G.R., Searcy, D.G. and DeLange, R.J. 1983. Histone-like protein in the archaebacterium *Sulfolobus acidocaldarius*. Biochem. Biophys. Acta 741, 251-257.

Gropp, F., Palm, P. and Zillig, W. 1989. Expression and regulation of *Halobacterium halobium* phage ϕH genes. Can. J. Microbiol. 35, 182-188.

Gropp, F., Reiter, W.D., Sentenac, A., Zillig, W., Schnabel, R., Thomm, M. and Stetter, K.O. 1986. Homologies of components of DNA-dependent RNA polymerase of archaebacteria, eukaryotes and eubacteria. Syst. Appl. Microbiol. 7, 95-101.

Gupta, R. 1984. *Halobacterium volcanii* tRNAs: identification of 41 tRNAs covering all amino acids and the sequences of 33 class I tRNAs. J. Biol. Chem. 259, 9461-9471.

Hackett, N.R. and DasSarma, S. 1989. Characterization of the small endogenous plasmid of *Halobacterium* strain SB3 and its use in transformation of *H. halobium*. Can. J. Microbiol. 35, 86-91.

Hamilton, P.T. and Reeve, J.N. 1985a. Sequence divergence of an archaebacterial gene cloned from a mesophilic and a thermophilic methanogen. J. Mol. Evol. 22, 351-360.

Hamilton, P.T. and Reeve, J.N. 1985b. Structure of genes and an insertion element in the methane producing archaebacterium *Methanobrevibacter smithii*. Mol. Gen. Genet. 200, 47-59.

Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28, 351-359.

Hofman, J.D., Schalkwyk, L.C. and Doolittle, W.F. 1986. ISH51: a large, degenerate family of insertion sequence-like elements in the genome of the archaebacterium, *Halobacterium volcanii*. Nucleic Acids Res. 14, 6983-7000.

Holmes, M.L. and Dyall-Smith, M.L. 1990. A plasmid vector with a selectable marker for halophilic archaebacteria. J. Bacteriol. 172, 756-761.

Holmes, M.L. and Dyall-Smith, M.L. 1991. Mutations in DNA gyrase result in novobiocin resistance in halophilic archaebacteria. J. Bacteriol. 173, 642-648.

Horne, M., Englert, C., and Pfeifer, F. 1988. Two genes encoding gas vacuole proteins in *Halobacterium halobium*. Mol. Gen. Genet. 213, 459-464.

Hudepohl, U., Reiter, W.D. and Zillig, W. 1990. *In vitro* transcription of two rRNA genes of the archaebacterium *Sulfolobus* sp. B12 indicates a factor requirement for specific initiation. Proc. Natl. Acad. Sci. USA 87, 5851-5855.

Huet, J., Schnabel, R., Sentenac and Zillig, W. 1983. Archaebacteria and eukaryotes possess DNA-dependent RNA polymerases of a common type. EMBO J. 2, 1291-1294.

Hui, I. and Dennis, P.P. 1985, Characterization of the ribosomal RNA gene clusters in *Halobacterium cutirubrum*. J. Biol. Chem. 260, 899-906.

Hutter, R., Niederberger, P. and DeMoss, J.A. 1986. Tryptophan biosynthetic genes in eukaryotic microorganisms. Annu. Rev. Microbiol. 40, 55-77.

Hyde, C. C., Ahmed, S. A., Padlan, E. A., Miles, E. W. and Davies, D. R. 1988. Three-dimensional strucyure of the tryptophan synthetase $\alpha 2\beta 2$ multienzyme complex from *Salmonella typhimurium*. J. Biol. Chem. 263: 17857-17871.

Hyde, C.C. and Miles, E. W. 1990. The tryptophan synthase multienzyme complex: exploring structure-function relationships with x-ray crystallography and mutagenesis. Biotechnology 8, 27-32.

Iwabe, N., Kuma, K.-I., Hasegawa,M., Osawa, S. and Miyata, T. 1989. Evolutionary relationship of archaebacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. Proc. Natl. Acad. Sci. USA 86, 9355-9359.

Janekovic, D., Wunderl, S., Holz, I., Zillig, W., Gierl, A. and Neumann, H. 1983. TTV1, TTV2 and TTV3, a fa nily of viruses of the extremely thermophilic, anaerobic, sulfur reducing archaebacterium *Thermoproteus tenax*. Mol Gen. Genet. 192, 39-45.

Jarsch, M. and Bock, A. 1985. Sequence of the 23S rRNA gene from the archaebacterium *Methanococcus vannielii* : evolutionary and functional implications. Mol. Gen. Genet. 200, 305-312.

Jones, W. J., Nagle, D.P. and Whitman, W.B. 1987. Methanogens and the diversity of archaebacteria. Microbiol. Rev. 51, 135-177.

Jordan, M., Meile, L. and Leisinger, T. 1989. Organization of *Methanobacterium thermoautotrophicum* bacteriophage ψM1 DNA. Mol. Gen. Genet. 220, 161-164.

Jordon-Starck, T.C. and Rodwell V.W. 1989a. *Pseudomonas mevalonii* 3-hydroxy-3-methylglutaryl coenzyme A reductase: characterization and chemical modification. J. Biol. Chem. 264, 17913-17918.

Jordon-Starck, T.C. and Rodwell V.W. 1989b. Role of cysteine residues in *Pseudomonas mevalonii* 3-hydroxy-3-methylglutaryl-Co A reductase: site-directed mutagenesis and characterization of the mutant enzymes. J. Biol. Chem. 264, 17919-17923.

Juez, G., Rodriguez-Valera, F., Herrero, N. and Mojica, F.J.M. 1990. Evidence for salt-associated restriction pattern modifications in the archaebacterium *Haloferax mediterranei*. J. Bacteriol. 172, 7278-7281.

Kagramanova, V.K., Derckacheva, N.I. and Mankin, A. S. 1989. Unusual nucleotide sequence heterogeneity of small multicopy pHSB plasmid from *Halobacterium* strain SB3, an archaebacterium. Can. J. Microbiol. 35, 160-163.

Kaine, B. P. 1987. Intron-containing tRNA genes of *Sulfolobus solfataricus*. J. Mol. Evol. 25, 248-254.

Kaine, B. P., Gupta, R. and Woese, C.R. 1983. Putative introns in tRNA genes of prokaryotes. Proc. Natl. Acad. Sci. USA 80, 3309-3312.

Kates, M., Wassef, M.K. and Kushner, D.J. 1968, Radioisotopic studies on the biosynthesis of the glyceryl diether lipids of *Halobacterium cutirubrum*. Can. J. Biochem. 46, 971-977.

Kikuchi, A., and Asai, K. 1984. Reverse gyrase- a topoisomerase which introduces positive superhelical turns into DNA. Nature 309, 677-681.

Kjems, J. and Garrett, R. A. 1985. An intron in the 23S ribosomal RNA gene of the archaebacterium *Desulfurococcus mobilis*. Nature 318, 675-677.

Kjems, J. and Garrett, R.A. 1987. Novel expression of the rRNA genes in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. EMBO J. 6, 3521-3530.

Kjems, J. and Garrett, R. A. 1988. Novel splicing mechanism for the ribosomal RNA intron in the archaebacterium *Desulfurococcus mobilis*. Cell 54, 693-703.

Kjems, J. and Garrett, R. A. 1990. Secondary structural elements exclusive to the sequences flanking ribosomal RNAs lend support to the monophyletic nature of the archaebacteria. J. Mol. Evol. 31, 25-32.

Kjems, J. and Garrett, R.A. 1991. Ribosomal RNA introns in archaea and evidence for RNA conformational changes associated with splicing. Proc. Natl. Acad. Sci. USA 88, 439-443.

Kjems, J., Jensen, J., Olesen, T. and Garrett. 1989. Comparison of transfer RNA and ribosomal RNA intron splicing in the extreme thermophile and archaebacterium *Desulfurococcus mobilis*. Can. J. Microbiol. 36, 210-214.

Kjems, J., Leffers, H., Garrett, R.A., Wich, G., Leinfelder, W. and Bock, A. 1987. Gene organization, transcription signals, and processing of the single ribosomal RNA operon of the archaebacterium *Thermoproteus tenax*. Nucl. Acids Res. 15, 4821-4835.

Klebe, R.J., Harriss, J.V., Sharp, Z.D. and Douglas, M.G. 1983. A general method for polyethylene-glycol-induced genetic transformation of bacteria and yeast. Gene 25, 333-341.

Klein, A. and Schnorr, M. 1984. Genome complexity of methanogenic bacteria. J. Bacteriol. 158, 628-631.

Knox, M.R. and Harris, J.E. 1988. Isolation and characterization of mutants of mesophilic methanogenic bacteria resistant to analogues of DNA bases and nucleosides. Arch. Microbiol. 149, 557-560.

Kopke, A.K.E. and Wittmann-Liebold, B. 1989. Comparative studies of ribosomal proteins and their genes from *Methanococcus vannielii* and other organisms. Can. J. Microbiol. 35, 11-20.

Lam, W.L. and Doolittle, W.F. 1989. Shuttle vectors for the archaebacterium *Halobacterium volcanii*. Proc. Natl. Acad. Sci. USA 86, 5478-5482.

Lam, W.L., Charlebois, R.L. and Doolittle, W.F. 1990a. Progress in molecular biology of the archaebacteria. In *Molecular Evolution*, UCLA Symposia on Molecular and Cellular Biology, New Series (Clegg, M.T. and O'Brien, J., ed.) Vol. 122, pp.265-272, Wiley-Liss Publisher, N.Y.

Lam, W.L., Cohen, A., Tsouluhas, D. and Doolittle, W.F. 1990b. Genes for tryptophan biosynthesis in the archaebacterium *Haloferax volcanii*. Proc. Natl. Acad. Sci. USA 87, 6614-6618.

Lanyi, J.K., Duschl, A., Hatfield, G.W., May, K. and Oesterhelt, D. 1990. The primary structure of a halorhodopsin from *Natronobacterium pharaonis*. J. Biol. Chem. 265, 1253-1260.

Larsen, N., Leffers, H., Kjems, J. and Garrett, R.A. 1986. Evolutionary divergence between the ribosomal RNA operons of *Halococcus morrhuae* and *Desulfurococcus mobilis*. System. Appl. Microbiol. 7, 49-57.

Lawrence, N., Wesolowski, D., Gold, H., Bartkiewicz, M., Guerrier-Takada, C., McClain, W.H. and Altman, S. 1987. Characteristics of ribonuclease P from various organisms. Cold Spring Harbor Symp. Quant. Biol. 52, 233-238.

Learned, R.M. and Fink, G.R. 1989. 3-hydroxy-3-methylglutaryl coenzyme A reductase from *Arabidopsis thaliana* is structurelly distinct from the yeast and animal enzymes. Proc. Natl. Acad. Sci. USA 86, 2779-2783.

Lechner, J. and Sumper, M. 1987. The primary structure of a procaryotic glycoprotein. J. Biol. Chem. 262, 9724-9729.

Leffers, H., Gropp, F., Lottspeich, F., Zillig, W. and Garrett, R.A. 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. J. Mol. Biol. 206, 1-17.

Leong, D., Boyer, H. and Betlach, M. 1988b. Transcription of genes involved in bacterio-opsin gene expression in mutants of a halophilic archaebacterium. J. Bacteriol. 170, 4910-4915.

Leong, D., Pfeifer, F., Boyer, H. and Betlach, M. 1988a. Characterization of a second gene involved in bacterio-opsin gene expression in a halophilic archaebacterium. J. Bacteriol. 170, 4903-4909.

Lobo, A.L. and Zinder, S.H. 1990. Nitrogenase in the archaebacterium *Methanosarcina barkeri* 227. J. Bacteriol. 172, 6789-6796.

Lunnen, K.D., Morgan, R.D., Timan, C.J., Krzycki, J.A., Reeve, J.N. and Wilson, G.G. 1989. Characterization and cloning of MwoI (GCN7GC), a new type-II restriction-modification system from *Methanobacterium wolfei*. Gene 77, 11-19.

Luskey, K.L. and Stevens, B. 1985. Human 3-hydroxy-3-methyl coenzyme A reductase. J. Biol. Chem. 260, 10271-10277.

Maniatis,T., Fritsch, E.F. and Sambrook, J. 1982. Molecular Cloning, A Laboratory Manual. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.)

Mankin, A., Teterina, N., Rubstov, P., Baratova, L., Kagramonova, V. 1984. Putative promoter regions of rRNA operon from archaebacterium *Halobacterium halobium*. Nucl. Acids Res. 12, 6537-6546.

Marck, C. 1988. 'DNA Strider': a 'C' program for the fast analysis of DNA and protein sequences on the Apple Macintosh family of computers. Nucl. Acids Res. 16, 1829-1836.

Martin, A. Yeats, S., Janekovic, D. Reiter, W.D., Aicher, W. and Zillig, W. 1984, SAV1, a temperate u.v. inducible DNA virus-like particle from the archaebacterium *Sulfolobus acidocaldarius* isolate B12. EMBO J. 3, 2165-2168.

Matsui, K., Miwa, K. and Sano, K. 1987. Two single-base-pair substitutions causing desensitization to tryptophan feedback inhibition of anthranilate synthase and enhanced expression of tryptophan genes of *Brevibacterium lactofermentum*. J. Bacteriol. 169, 5330-5332.

May, B.P. and Dennis, P.P. 1989. Evolution and regulation of the gene encoding superoxide dismutase from the archaebacterium *Halobacterium cutirubrum* . J. Biol. Chem. 264,12253-12258.

May, B.P., Tam, P. and Dennis, P.P. 1989. The expression of the superoxide dismutase gene in *Halobacterium cutirubrum* and *Halobacterium volcanii*. Can. J. Microbiol. 35, 171-175.

McCloskey, J.A. 1986. Nucleoside modification in archaebacterial transfer RNA. System. Appl. Microbiol. 7, 246-252.

McConnell, D.J., Searcy, D.G. and Sutcliffe, J.G. 1978. A restriction enzyme ThaI from the thermophilic mycoplasm *Thermoplasma acidophilum*. Nucl. Acids Res. 5, 1729-1739.

Meile, L., Jenal, U., Studer, D., Jordan, M. and Leisinger, T. 1989. Characterization of ψM1, a virulent phage of *Methanobacterium thermoautotrophicum* Marburg. Arch. Microbiol. 152, 105-110.

Messing, J. 1983. New M13 vectors for cloning.Methods Enzymol., 101, 20-78.

Mevarech, M. and Werczberger, R. 1985. Genetic transfer in *Halobacterium volcanii*. J.Bacteriol. 162, 461-462.

Mitchell, R.M., Loeblish, L.A., Klotz, L.C. and Loeblich, A.R. 1979. DNA organization of *Methanobacterium thermoautotrophicum*. Science 204, 1082-1084.

Moore, R.L. and McCarthy, B.J. 1969a. Characterization of the deoxynucleic acid of various strains of halophilic bacteria. J. Bacteriol. 99, 248-254.

Moore, R.L. and McCarthy, B.J. 1969b. Base sequence homology and renaturation studies of the deoxyribonucleic acid of extremely halophilic bacteria. J. Bacteriol. 99, 255-262.

Morris, C.J. and Reeve, J.N. 1988. Conservation of the structure in the human gene encoding argininosuccinate synthetase and the *argG* gene of the archaebacteria *Methanosarcina barkeri* MS and *Methanococcus vannielii*. J. Bacteriol. 170, 3125-3130.

Muller, B., Allmansberger, R. and Klein, A. 1985. Termination of a transcriptional unit comprising highly expressed genes in the archaebacterium *Methanococcus voltae*. Nucl. Acid. Res. 13, 6439-6445.

Murray, P.A. and Zinder, S.H. 1984. Nitrogen fixation by methanogenic archaebacterium. Nature 312, 284-286.

Nadal, M., Mirambeau, G., Forterre, P., Reiter, W.D. and Duguet, M. 1986. Positively supercoiled DNA in A virus-like particle of an archaebacterium. Nature 321, 256-258.

Nagle, D.P. 1989. Development of genetic systems in methanogenic archaebacteria. J. Indust. Microbiol. 30, 43-51.

Nakasu, S. and Kikuchi, A. 1985. Reverse gyrase: ATP-dependent typeI DNA topoisomerase from *Sulfolobus*. EMBO J. 4, 2705-2710.

Nieuwlandt, D.T. and Daniels, C.J. 1990. An expression vector for the archaebacterium *Haloferax volcanii*. J. Bacteriol. 172, 7104-7110.

Noll, K.M. 1989. Chromosome map of the thermophilic archaebacterium *Thermococcus celer*. J. Bacteriol. 171, 6720-6725.

Normand, P. and Bousquet, J. 1989. Phylogeny of nitrogenase sequences in *Frankia* and other nitrogen fixing microorganisms. J. Mol. Evol. 29, 436-447.

Oesterhelt, D. and Stoeckenius, W. 1973. Functions of a new photoreceptor membrane. Proc. Natl. Acad. Sci. USA 70, 2853-2857.

Oesterhelt, D. and Tittor, J. 1989. Two pumps, one principle: light-driven ion transport in *Halobacteria*. Trends Biochem. Sci. 14, 57-61.

Patterson, N.H. and Pauling, C. 1985. Evidence for two restriction-modification systems in *Halobacterium cutirubrum*. J. Bacteriol. 163, 783-784.

Pauling, C. 1982. Bacteriophages of *Halobacterium halobium*: isolation from fermented fish sauce and primary characterization. Can J. Microbiol. 28, 916-921.

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA 85, 2444-2448.

Pfeifer, F. 1986. Insertion elements and genome organization of *Halobacterium halobium*. Syst. Appl. Microbiol. 7, 36-40.

Pfeifer, F. and Betlach, M. 1985. Genome organization in *Halobacterium halobium*: a 70 kb island of more (AT) rich DNA in the chromosome. Mol. Gen. Genet. 198, 449-455.

Pfeifer, F. Betlach, M., Martienssen, R., Friedman, J. and Boyer, H.W. 1983. Transposable elements of *Halobacterium halobium*. Mol. Gen. Genet. 191, 182-188.

Pfeifer, F. and Blaseio, U. 1989. Insertion elements and deletion formation in a halophilic archaebacterium. J. Bacteriol. 171, 5135-5140.

Pfeifer, F. and Blaseio, U. 1990. Transposition burst of the ISH27 insertion element family in *Halobacterium halobium*. Nucl. Acid.Res. 18, 6921-6925.

Pfeifer, F., Blaseio, U. and Ghahraman, P. 1988. Dynamic plasmid populations in *Halobacterium halobium*. J. Bacteriol. 170, 3718-3724.

Pfeifer, F., Blaseio, U. and Horne, M. 1989. Genome structure of *Halobacterium halobium*: plasmid dynamics in gas vacuole deficient mutants. Can. J. Microbiol. 35, 96-100.

Pfeifer, F., Boyer, H. and Betlach, M. 1985. Restoration of bacterio-opsin gene expression in a revertant of *Halobacterium halobium*. J. Bacteriol. 164, 414-420.

Pfeifer, F., Friedman, J., Boyer, H. and Betlach, M. 1984. Characterization of insertions affecting the expression of the bacterio-opsin gene in *Halobacterium halobium*. Nucl. Acids Res. 12, 2489-2497.

Pfeifer, F., Weidinger,G. and Goebel,W. 1981. Genetic variability of *Halobacterium halobium*. J. Bacteriol. 145, 369-374.

Prangishvilli, D.A., Vashakidze, R.P., Chelidze, M.G. and Gabriadze, I.Y. 1985. A restriction endonuclease SauI from the thermophilic archaebacterium *Sulfolobus acidocaldarius*. FEBS Lett. 192, 57-60.

Priestle, J.P., Grutter, M.G., White, J.L., Vincent, M.G., Kania, M., Wilson, E., Jardetzky, T.S., Kirschner, K. and Jansonius, J.N. 1987. Three-dimensional structure of the bifunctional enzyme N-(5'-phosphoribosyl)anthranilate isomerase-indole-3-glycerol-phosphate synthase from *Escherichia coli*. Proc. Natl. Acad. Sci. USA 84, 5690-5694.

Puhler, G., Leffers, H., Gropp, F., Palm, P., Klenk, H.-P., Lottspeich, F., Garrett, R. and Zillig, W. 1989. Archaebacterial DNA-dependent RNA polymerases testify to the evolution of the eukaryotic nuclear genome. Proc. Natl. Acad. Sci. USA 86, 4569-4573.

Queen, C. and Korn, L.J. 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucl. Acids Res. 12, 581-599.

Rajkovic, A., Simonsen, J.N., Davis, R.E. and Rottman, F.M. 1989. Molecular cloning and sequencing analysis of 3-hydroxy-3-methylglutaryl coenzyme A reductase from the human parasite *Schistosoma mansoni*.Proc. Natl. Acad. Sci. USA 86, 8217-8221.

Ramirez, C., Shimmin, L.C., Newton, C.H., Matheson, A.T. and Dennis, P.P. 1989. Structure and evolution of the L11, L1, L10, and L12 equivalent ribosomal proteins in eubacteria, archaebacteria, and eucaryotes. Can. J. Microbiol. 35, 234-244.

Ree, H.K., Cao, K., Thurlow, D.L. and Zimmermann, R.A. 1989. The structure and organization of the 16S ribosomal RNA gene from the archaebacterium *Thermoplasma acidophilum*. Can. J. Microbiol. 35, 124-133.

Ree, H.K. and Zimmermann, R.A. 1990. Organization and expression of the 16S, 23S and 5S ribosomal RNA genes from the archaebacterium *Thermoplasma acidophilum*. Nucl. Acids Res. 18, 4471-4478.

Reiter, W.D., Hudepohl, U. and 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.

Reiter, W.D., Palm, P. and Yeats, S. 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucl. Acids Res. 17 1907-1914.

Reiter, W.D., Palm, P. and Zillig, W. 1988a. Transcription termination in the archaebacterium *Sulfolobus*: signal structures and linkage to transcription initiation. Nucl. Acids Res. 14, 2445-2459.

Reiter, W.D., Palm, P. and Zillig, W. 1988b. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic pol II promoters. Nucl. Acids Res. 16, 1-19.

Reiter, W.D., Palm, P., Henschen, A., Lottspeich, F., Zillig, W. and Grampp, B. 1987a. Identification and characterization of the genes encoding three structural proteins of the *Sulfolobus* virus-like particle SSV1. Mol. Gen. Genet. 205, 144-153.

Reiter, W.D., Palm, P., Yeats, S. and Zillig, W. 1987b. Gene expression in archaebacteria: physical mapping of constitutive and uv-indu ble transcripts from *Sulfolobus* virus like particle SSV1. Mol. Gen. Genet. 209, 270-27.

Reiter, W.D., Zillig, W. and Palm, P. 1988c. Archaebacterial viruses. Adv. Virus Res. 34, 143-188.

Reiter, W.D.and Palm, P. 1990. Identification and characterization of a defective SSV1 genome integrated into a tRNA gene in the archaebacterium *Sulfolobus* sp. B12. Mol. Gen. Genet. 221, 65-71.

Reznikoff, W.S., Siegele, D.A., Cowing, D.W. and Gross, C.A. 1985. The regulation of transcription initiation in bacteria. Annu. Rev. Genet. 19, 355-387.

Rohrmann, G.F., Cheney, R. and Pauling, C. 1983. Bacteriophages of *Halobacterium halobium*: virion DNAs and proteins. Can. J. Microbiol. 29, 627-629.

Roitelman, J. and Shechter, I. 1989. Studies on the catalytic site of rat liver HMG-CoA reductase: interaction with CoA-thioesters and inactivation by indoacetamide. J. Lipid Res. 30, 97-107.

Rosenberg, M. and Court, D. 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. 13, 319-353.

Rosenshine, I. Zusman, T., Werczberger, R. and Mevarech, M. 1987. Amplification of specific DNA sequences correlates with resistance of the archaebacterium *Halobacterium volcanii* to the dihydrofolate reductase inhibitors trimethoprim and methotrexate. Mol. Gen. Genet. 208, 518-522.

Rosenshine, I. and Mevarech, M. 1989. Isolation and partial characterization of plasmids found in three *Halobacterium volcanii* isolates. Can. J. Microbiol. 35, 92-95.

Rosenshine, I., Tchelet, R. and Mevarech, M. 1989. The mechanism of DNA transfer in the mating system of an archaebacterium. Science 245, 1387-1389.

Sambrook, J., Fritsch, E.F. and Maniatis,T. 1989. Molecular Cloning, A Laboratory Manual, 2nd edition. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.)

Sanangelantoni, A.M., Barbarini, D., Di Pasquale, G., Cammarano, P. and Tiboni, O. 1990.Cloning and nucleotide sequence of an archaebacterial glutamine synthetase gene: phylogenetic implications. Mol. Gen. Genet. 221, 187-194.

Sanger, F., Nicklen, S. and Coulson, A.R. 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463-5467.

Sapienza, C. and Doolittle, W.F. 1982. Unusual physical organization of the *Halobacterium* genome. Nature 295, 384-389.

Sapienza, C., Rose, M.R. and Doolittle, W.F. 1982. High-frequency genomic rearrangements involving archaebacterial repeat sequence elements. Nature 299, 182-185.

Scherer, P. 1989. Vanadium and molybdenum requirement for the fixation of molecular nitrogen by two *Methanosarcina* strains. Arch. Microbiol. 151, 44-48.

Schmid, K., Thomm, M., Laminet, A., Laue, F.G., Kessler, C., Stetter, K.O. and Schmitt, R. 1984. Three restriction endonucleases MaeI, MaeII and MaeIII from *Methanococcus aeolicus*. Nucl. Acids Res. 12, 2619-2628.

Schnabel, H. 1984. An immune strain of *Halobacterium halobium* carries the invertible L segment of ϕH as a plasmid. Proc. Natl. Acad. Sci. USA 81, 1017-1020.

Schnabel, H., Palm, P., Dick, K. and Grampp, B. 1984. Sequence analysis of the insertion element ISH1.8 and associated structural changes in the genome of phage ϕH of the archaebacterium *Halobacterium halobium*. EMBO J. 3, 1717-1722.

Schnabel, H., Schramm, E., Schnabel, R. and Zillig, W. 1982a. Structural variability in the genome of phage ϕH of *Halobacterium halobium*. Mol. Gen. Genet. 188, 370-377.

Schnabel, H. and Zillig, W. 1984. Circular structure of the genome of phage ϕH in a lysogenic *Halobacterium halobium*. Mol. Gen. Genet. 193, 422-426.

Schnabel, H., Zillig, W., Pfaffle, M., Schnabel, R., Michel, H. and Delius, H. 1982b. *Halobacterium halobium* phage ϕH. EMBO J. 1, 87-92.

Searcy, D.G. and Doyle, E.K. 1975 Characterization of *Thermoplasma acidophilum* deoxynucleic acid. Int. J. Syst. Bacteriol. 25, 286-289.

Shimmin, L.C. and Dennis, P.P. 1989. Characterization of the L11, L1, L10 and L12 equivalent ribosomal protein gene cluster of the halophilic archaebacterium *Halobacterium cutirubrum* . EMBO J. 8, 1225-1235.

Shimmin, L.C., Newton, C.H., Ramirez, C., Yee, J., Downing, W.L., Louie, A., Matheson, A.T. and Dennis, P.P. 1989a. Organization of genes encoding the L11, L1, L10, and L12 equivalent ribosomal proteins in eubacteria, archaebacteria, and eucaryotes. Can. J. Microbiol. 35, 164-170.

Shimmin, L.C., Ramirez, C., Matheson, A.T. and Dennis, P.P. 1989b. Sequence alignment and evolutionary comparison of the L10 equivalent and L12 equivalent ribosomal proteins from archaebacteria, eubacteria, and Eucaryotes. J. Mol. Evol. 29, 448-462.

Sibold, L. and Henriquet, M. 1988. Cloning of the *trp* genes from the archaebacterium *Methanococcus voltae*: nucleotide sequence of the *trpBA* genes. Mol. Gen. Genet. 214, 439-450.

Sibold, L., Pariot, D., Bhatnagar, L., Henriquet, M. and Aubert, J.P. 1985. Hybridization of DNA from methanogenic bacteria with nitrogenase structural genes (*nifHDK*). Mol. Gen. Genet. 200, 40-46.

Sibold, L. and Souillard, N. 1986. Primary structure and expression of a gene homologous to *nifH* (nitrogenase Fe protein) from the archaebacterium *Methanococcus voltae*. Mol. Gen. Genet. 203, 21-28.

Sibold, L. and Souillard, N. 1989. Primary structure, functional organization and expression of nitrogenase structural genes of the thermophilic archaebacterium *Methanococcus thermolithotrophicus*. Mol. Microbiol. 3, 541-551.

Simon, R.D. 1978. *Halobacterium* strain 5 contains a plasmid which is correlated with the presence of gas vacuoles. Nature 273, 314-317.

Simsek, M., DasSarma, S., RajBhandary, U.L., and Khorana, H.G. 1982. A transposable element from *Halobacterium halobium* which inactivates the bacteriorhodopsin gene. Proc. Natl. Acad. Sci. USA 79, 7268-7272.

Sioud, M., Baldacci, G., Forterre, P. and DeRecondo, A.M. 1988a. Novobiocin induces accumulation of a single strand of plasmid pGRB-1 in the archaebacterium *Halobacterium* GRB. Nucl. Acids Res. 16, 7833-7841.

Sioud, M., Possot, O., Elie, C., Sibold, L. and Forterre, P. 1988b. Coumarin and quinolone action in archaebacteria: evidence for the presence of a DNA gyrase-like enzyme. J. Bacteriol. 170, 946-953.

Slezarev, A.I. 1988. Positive supercoiling catalyzed in vitro by ATP dependent topoisomerase from *Desulfurococcus amylolyticus*. Eur. J. Biochem. 173, 395-399.

Sment, K.A. and Konisky, J. 1986. Regulated expression of *Methanococcus voltae hisA* gene. Syst. Appl. Microbiol. 7, 90-94.

Souillard, N., Magot, M., Possot, O. and Sibold, L. 1988. Nucleotide sequence of regions homologous to *nifH* (nitrogenase Fe protein) from the nitrogen-fixing archaebacteria *Methanococcus thermolithotrophicus* and *Methanobacterium ivanovii*: evolutionary implications. J. Mol. Evol. 27, 65-76.

Spiridonova, V.A., Akhmanova, A.S., Kagramanova, V.K., Kopke, A.K.E. and Mankin, A.S. 1989. Ribosomal protein gene cluster of *Halobacterium halobium*: nucleotide sequence of the gene coding for S3 and L29 equivalent ribosomal proteins. Can. J. Microbiol. 35, 153-159.

Spudich, E.N., Takahashi, T. and Spudich, J.L. 1989. Sensory rhodopsins I and II modulate a methylation/demethylation system in *Halobacterium halobium* phototaxis. Proc. Natl. Acad. Sci. USA 86, 7746-7750.

Spudich, J.L. and Bogomolni, R.A. 1988. Sensory rhodopsins of halobacteria. Annu. Rev. Biophys. Biophys. Chem. 17, 193-215.

Stein, D.B. and Searcy, D.G. 1978. Physiologically important stablization of DNA by a prokaryotic histone-like protein. Science 202, 219-221.

Stoeckenius, W. and Bogomolni, R. 1982. Bacteriorhodopsin and related pigments of halobacteria. Annu. Rev. Biochem. 52, 587-616.

Sudhof, T.C., Fried, V.A., Stone, V.A., Johnston, P.A. and Xie, X.-S. 1989. Human endomembrane H^+ pump strongly resembles the ATP-synthetase of Archaebacteria. Proc. Natl. Acad. Sci. USA 86, 6067-6071.

Takao, M., Kobayashi, T., Oikawa, A. and Yasui, A. 1989. Tandem arrangement of photolyase and superoxide dismutase genes in *Halobacterium halobium*. J. Bacteriol. 171, 6323-6329.

Thomas, M.S. and Nomura, M. 1987. Translational regulation of the L11 ribosomal protein operon of *Escherichia coli*: mutations that define the target site for repression by L1. Nucl. Acids Res. 15, 3085-3096.

Thomm, M. and Wich, G. 1988. An archaebacterial promoter element for stable RNA genes with homology to the TATA-box of higher eukaryotes. Nucleic Acids Res. 16, 151-163.

Thomm, M., Wich, G., Brown, J.W., Frey, G., Sherf, B.A. and Beckler, B.S. 1989. An archaebacterial sequence assigned by RNA polymerase binding experiments. Can. J. Microbiol. 35, 30-35.

Thomm, M. Sherf, B.A. and Reeve, J.N. 1988. RNA polymerase binding and transcription initiation sites upstream of the methyl reductase operon of *Methanococcus vannielii*. J. Bacteriol. 170, 1958-1961.

Thompson, L. and Daneils, C.J. 1988. A $tRNA^{Trp}$ intron endonuclease from *Halobacterium volcanii*. J. Biol. Chem. 263, 17951-17959.

Thompson, L., Brandon, L.D., Nieuwlandt, D.T. and Daneils, C.J. 1989. Transfer RNA intron processing in the halophilic archaebacteria. Can. J. Microbiol. 35, 36-42.

Torsvik, T. and Dundas, I.D. 1974. Bacteriophage of *Halobacterium salinarium*. Nature 248, 680-681.

Torsvik, T. and Dundas, I.D. 1980. Persisting phage infection in *Halobacterium salinarium* str.1. J. Gen. Virol. 47, 29-36.

Tso, J.Y. Hermodson, M.A. and Zalkin, H. 1980. Primary structure of *Serratia marcescens* anthranilate synthase component II. J. Biol. Chem. 255, 1451-1457.

Tu, J. and Zillig, W. 1982. Organization of rRNA structural genes in the archaebacterium *Thermoplasma acidophilum*. Nucl. Acids Res. 10, 7231-7245.

Twigg, A.J. and Sherratt, D. 1980. Trans-complementable copy-number mutants of plasmid ColE1. Nature 283, 216-218.

Vogelsang-Wenke, H. and Oesterhelt, D. 1988. Isolation of a halobacterial phage with a fully cytosine-methylated genome. Mol. Gen. Genet. 211, 407-414.

Wais, A.C., Kon, M. and MacDonald, R.E. 1975. Salt-dependent bacteriophage infecting *H. cutirubrum* and *H. halobium*. Nature 256, 314-315.

Weidinder, G., Klotz, G. and Goebel, W. 1979. A large plasmid from *Halobacterium halobium* carrying genetic information for gas vacuole formation. Plasmid 2, 377-386.

Weil, C.F., Beckler, G.S. and Reeve, J.N. 1987. Structure and organization of the *hisA* gene of the thermophilic archaebacterium *Methanococcus thermolithotrophicus*. J. Bacteriol. 169, 4857-4860.

Wich, G. Jarsch, M. and Bock, A. 1984. Apparent operon for a 5S ribosomal RNA gene and for tRNA genes in the archaebacterium *Methanococcus vannielii*. Mol. Gen. Genet. 196, 146-151.

Wich, G., Hummel, H., Jarsch, M. Bar, U and Bock, A. 1986a. Transcription signals for stable RNA genes in *Methanococcus*. Nucl. Acids Res. 14, 2459-2479.

Wich, G., Leinfelder, W. and Bock, A. 1987. Gene for stable RNA in the extreme thermophile *Thermoproteus tenax*: introns and transcription signals. EMBO J. 6, 523-528.

Wich, G., Sibold, L. and Bock, A. 1986b. Genes for tRNA and their putative expression signals in *Methanococcus*. System. Appl. Microbiol. 7, 18-25.

Woese, C. R. 1987. Bacterial evolution. Microbiol Rev 51, 221-271.

Woese, C. R. and Fox, G. F. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74, 5088-5092.

Woese, C.R. 1981. Archaebacteria. Sci. Amer. 244, 98-122.

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

Woese, C.R. and Plsen, G.J. 1986. Archaebacterial phylogeny: perspective on the urkingdoms. System. Appl. Microbiol. 7, 161-177.

Wood, A.G., Redborg, H., Cue, D.R., Whitman, W.B., and Konisky, J. 1983. Complementation of *argG* and *hisA* mutations of *Escherichia coli* by DNA cloned from the archaebacterium *Methanococcus voltae*, J. Bacteriol. 156, 19-29.

Woodward, H.D., Allen, J.M.C. and Lennarz, W.J. 1988. 3-hydroxy-3-methylglutaryl coenzyme A reductase of the sea urchin embryo. J. Biol. Chem. 263, 18411-18418.

Worrell, V.E. and Nagle D.P. 1990. Genetics and physiological characterization of the purine salvage pathway in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. J. Bacteriol. 172, 3328-3334.

Worrell, V.E., Nagle D.P., McCarthy, D. and Eisenbraun, A. 1988. Genetic transformation system in the archaebacterium *Methanobacterium thermoautotrophicum* Marburg. J. Bacteriol. 170, 653-656.

Xu, W.L. and Doolittle, W.F. 1983. Structure of the archaebacterial transposable element ISH50. Nucl. Acids Res. 11, 4195-4199.

Yaniagish, A. and Oshima, T. 1990. Circular chromosomal DNA in the sulfur-dependent archaebacterium *Sulfolobus acidocaldarius*. Nucl. Acids Res. 18, 1133-1136.

Yanofsky, C. 1984. Comparison of regulatory and structural regions of genes of tryptophan metabolism. Mol. Biol. Evol. 1, 143-161.

Yanofsky, C. and Crawford, I.P. 1972. Tryptophan synthetase. In *The Enzyme* vol. 7 (Boyer,P.D , ed) 3rd ed, pp. 1-31, Academic Press, New York.

Yanofsky, C. and Crawford, I.P. 1987. The tryptophan operon. In *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology* (Neidhardt, F.C., ed) Vol. 2, pp. 1453-1472, Amercian Society for Microbiology, Washington, D.C.

Yanofsky, C., Platt, T., Crawford, I. P. Nichols, B. P., Christie, G. E., Horowitz, H., M. VanCleemput and Wu, A. M. 1981. The complete nucleotide sequence of the tryptophan operon of Escheria coli. Nucleic Acids Res. 9: 6647-6668.

Zalkin, H., Paluh, J., van Cleemput, M., Moye, W.S. and Yanofsky, C. 1984. Nucleotide sequence of *Saccharomyces cerevisiae* genes *TRP2* and *TRP3* encoding bifunctional anthranilate synthase:Indole-3-glycerol phosphate synthetase. J. Biol. Chem. 259, 3985-3992.

Zeff, R.A. and Geliebter, J. 1987. Oligonucleotide probes for genomic DNA blots. BRL Focus 9:2, 1-2.

Zillig, W., Klenk H.P., Palm, P., Puhler, G. and Gropp, F. 1989. The phylogenetic relations of DNA-dependent RNA polymerases of archaebacteria, eukaryotes and eubacteria. Can. J. Microbiol. 35, 73-80.

Zillig, W., Palm, P., Reiter, W.-D., Gropp, F., Puhler,G. and Klenk, H. 1988. Comparative evaluation of gene expression in archaebacteria. Eur. J. Biochem. 173, 473-482.

Zillig, W., Yeats, S., Holz, I., Böck, A., Gropp, F., Rettenberger, M. and Lutz, S. 1985. Plasmid-related anaerobic autotrophy of the novel archaebacterium, *Sulfolobus ambivalens*. Nature 313, 789-791.