CHARACTERIZATION OF THE RIBOSOMAL RNA OPERONS OF *HALOARCULA MARISMORTUI*

by

SHANTHINI MYLVAGANAM

B.Sc. (Hons.), University of Jaffna, Sri Lanka, 1985
M. Sc., Bowling Green State University, U. S. A., 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES
(Department of Biochemistry and Molecular Biology)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
March 1996

In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of BIOCHEMISTRY

The University of British Columbia
Vancouver, Canada

Date July 23, 1996

ABSTRACT

The genome of *Haloarcula marismortui* contains two ribosomal RNA operons, designated as *rrnA* and *rrnB* (and possibly a third operon designated as *rrnC*) of which the characterization of the *rrnA* and *rrnB* operons are presented. Characterization of the *rrnA* and *rrnB* operons involved the analysis of primary and secondary structures and *in vivo* studies of the primary transcripts and processing intermediates. It was found that the gene orders of the *rrnA* and *rrnB* operons were 5'–16S rRNA–tRNA^{Ala}–23S rRNA–5S rRNA–tRNA^{Cys}–3' and 5'–16S rRNA–23S rRNA–5SrRNA–3', respectively. Computing the substitution rates for the entire *rrnA* and *rrnB* operons demonstrated that the major differences are localized in the non coding regions, that is the regions including the 5'-flanking of the 16S rRNA, 16S-23S rRNA spacer and the 3'-flanking of the 5S rRNA gene. The percentage similarities between the 16S, 23S and 5S rRNAs of *rrnA* and *rrnB* are 95%, 98.7% and 98.3%, respectively. A pairwise sequence comparison between the 23S rRNA sequence of the *rrnC* operon (Brombach *et al.*, 1989) and the other two operons, *rrnA* and *rrnB*, revealed that the sequence similarities are 98.8% and 99.6%, respectively. The 5S rRNA sequence from the *rrnC* operon is identical to the *rrnA* sequence.

The nucleotide substitutions within the 16S rRNA genes of *rrnA* and *rrnB* operons are concentrated in three separate domains 58-321, 508-823 and 986-1158. About 60% of the substitutions are concentrated within the 508-823 domain and are compensatory, affecting both components of the nucleotide base pairs within defined rRNA helices. Using nuclease S1 protection assays, it was shown that the 16S rRNAs from the *rrnA* and *rrnB* operons are expressed and present in intact 70S ribosomes. A comparison of the 23S rRNAs from the *rrnA* and *rrnB* operons showed that the substitutions are located within the variable regions of domains I, III, IV and VI of the universal secondary structures of 23S rRNAs. The 5S rRNA sequences of the two operons differ at two nucleotide positions in the helix IV of the universal secondary structure for the 5S rRNA.

The 5'-flanking regions of the *rrnA* contains four tandem promoters whereas the *rrnB* operon contains a single tandem promoter and a second promoter-like sequence. An internal promoter sequence was present within the 16S-23S spacer regions of all three operons from *Ha. marismortui*. Putative secondary structures of the primary transcripts from the *rrnA* operon showed that the 16S and 23S rRNAs are surrounded by inverted repeat structures containing the "bulge-helix-bulge" motif which is recognized by a processing endonuclease. In the case of the *rrnB* operon, the inverted repeat structure surrounding the 23S rRNA is identical to that of the *rrnA* operon and processing follows the same pathway. However, the inverted repeat structure surrounding 16S rRNA from *rrnB* does not contains the "bulge-helix-bulge" motif and its processing follows a distinct pathway. The 16S rRNA processing of the *rrnB* occurs at a single position within the 5'-flanking region and at three positions within the 16S-23S spacer region. The nucleotides present in these cleavage sites and their surrounding regions showed no sequence conservation. The 5S rRNA processing occurs close to or at its 5'- and 3'-ends. Two apparent termination sites were detected for the *rrnA* operon and they were found to be located downstream of the $tRNA^{Cys}$ RNA.

Phylogenetic analysis of the 508-823 region of the 16S rRNA, the entire 16S rRNA, 23S rRNA genes and the combination of 16S-23S-5S rRNA genes were performed using the PAUP program. The phylograms from the 16S rRNA, 23S rRNA and the 16S-23S-5S rRNA sequences indicated that the *rrnA* and *rrnB* group together and the rate of divergence of the *rrnB* operon is higher than that of the *rrnA* operon. However, comparison of the 508-823 regions showed that the two operons do not group together and that *rrnB* is evolving slower than *rrnA*. Based on the comparisons made between the *rrnA* and *rrnB* operons, it is obvious that the *rrnB* operon is different from the *rrnA* operon in its gene order, rRNA sequences and 16S rRNA processing and also probably evolving at a different rate than the *rrnA* operon.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

A	Adenosine
A260	Absorbance at 260 nm
AMV	Avian myeloblastosis virus
ATP	Adenosine 5'-triphosphate
bp	Base pair
BPB	Bromophenol blue
BSA	Bovine serum albumin
C	Cytosine
dATP	2'-deoxyadenosine - 5' - triphosphate
dCTP	2'-deodycytidine - 5' - triphosphate
ddATP	2', 3' - dideoxyadenosine - 5' - triphosphate
ddCTP	2', 3' - dideoxycytidine - 5' -t riphosphate
ddGTP	2', 3' - dideoxyguanosine - 5' - triphosphate
ddTTP	2', 3' - dideoxythymidine - 5' triphosphate
dNTP	2', 3' - deoxyribonucleotide - 5' - triphosphate (dATP. dCTP, dGTP and dTTP)
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediamine tetraacetic acid
FDM	Formamide Dye Mix
G	Guanosine
GTP	Guanosine - 5' - triphosphate
IPTG	Isopropyl - β - D - thiogalactopyranoside
kbp	Kilobase pair
kd	Kilodaltons
LSU	Large subunit ribosomal RNA
mRNA	Messenger RNA

NTP	Ribonucleotide - 5' triphosphate (ATP, CTP, GTP and TTP)
PAGE	Polyacrylamide gel electrophoresis
PAUP	Phylogenetic analysis using Parsimony
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RNase	Ribonuclease
RNP	Ribonuclear protein
S	Svedberg unit of sedimentation coefficient
SDS	Sodium dodecyl sulphate
SSC	Standard saline citrate
SSU	Small subunit ribosomal RNA
T	Thymidine
Tris	Trihydroxymethylaminomethane
tRNA	Transfer RNA
x-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside

ACKNOWLEDGEMENTS

I would like to thank Dr. Patrick Dennis for providing me the opportunity to work in his lab, for his guidance and encouragement throughout this study. I am also indebted to Drs. Peter Candido and Rosemary Redfield for their advice and support. I would also like to thank Dr. George Mackie for the critical reading of this thesis and comments.

A special thanks goes to Deidre de Jong-Wong for all her technical assistance, for proof reading my thesis several times without any hesitation and all her help, support and friendship.

I thank all my colleagues, Phulgun Joshi, Lawrence Shimmin, Janet Chow, Janet Yee, Luc Bissonett, Peter Durovic, Simon Potter, Daiqing Liao and Josephile Yau for their help, cooperation and friendship in the lab. I am also grateful to Dr. Lingwood and his students from Hospital for Sick Children for their help during the final preparation of my thesis

Most of all, I am very grateful to Yogam Aunty, Yoga, Jeyanthy, Devini and the children, Amanda, Sandra and Melinda for taking care of Meena, for their love and unconditional support during the past few years.

DEDICATION

To Meena, Saeyon, Myl and My Family

CHAPTER 1

General Introduction

1.1 The Evolutionary Record

It is widely acknowledged by scientists from various disciplines that our planet earth came into existence about 4.5 billion years ago (Walter, 1983). Furthermore, an evaluation of microbial fossil records suggests that the photosynthetic eubacteria emerged 3 to 4 billion years ago (Ernst, 1983). Recently, eleven taxa of cellularly preserved filamentous microbes have been discovered in a bedded chert unit of the early archaean apex basalt of Northwestern Western Australia (Schopf, 1993), which are among the oldest fossil records known. This prokaryotic assemblage establishes that the trichomic cyanobacterium-like microorganisms were extant and morphologically diverse as early as 3.5 billion years ago. The records further suggest that oxygen-producing photoautotrophy may have already evolved and been present in this early stage of biotic history (Schopf, 1993). Therefore on the basis of these findings it is suggested that the emergence of life on our planet occurred about 3.5 billion years ago or even earlier.

Geological fossil records provide valuable information in reconstructing early evolutionary events. However, for biologists, the components of the cell also reveal the evolutionary past, particularly the amino acid sequences of its proteins and the nucleotide sequence of its nucleic acids, DNA and RNA. These living records are potentially richer and more extensive than the fossil records. Also, they reach beyond the time of the oldest fossil records known, back to the period when the common ancestor of all life forms existed.

Current speculations on the nature of the early genetic information focuses on RNA because of its unique physical, chemical and genetic property. The inherent template properties of the RNA enables it to self replicate using a minus strand intermediate (Gilbert, 1986; Weiner and Maizels, 1987). Furthermore, the discovery of catalytic RNA as illustrated

by the self-splicing group I and II introns, the putative peptidyl transferase activity of the 23S RNA of the ribosome (Noller *et al.*, 1992) and the catalytic activity of the RNA component of RNase P (Darr *et al.*, 1992) suggest the possibility that early catalytic and later information-driven protein synthesis could have been carried out exclusively by RNA. It has been shown recently that a genetically engineered ribozyme is able to function effectively as catalyst and template in self-copying reactions (Green and Szostack, 1992). The realization that RNA can serve as genome and catalyst initiated extensive discussion on the role of RNA in the origin of life (Pace and Marsh, 1985; Sharp, 1985; Lewin, 1986) which in turn led to the coining of the phrase "the RNA world" (Gilbert, 1986).

1.2 The RNA World and Beyond

How protein synthesis would have evolved in an RNA world was addressed in the "genomic tag model" of Weiner and Maizels (1987, 1991). In this model, it was proposed that ancient linear RNA genomes possessed 3'-terminal tRNA-like structures which they called genomic tags. Like the 3'-terminal tRNA-like motifs of contemporary bacterial and plant RNA viruses (Rao *et al.*, 1989) and possibly animal picornaviruses (Pilipenko *et al.*, 1992), the genomic tag also would have served in important roles such as providing an initiation site for replication and functioning as a simple telomere.

Transfer RNAs would have been derived from the 3'-terminal tRNA-like structures which tagged RNA genomes for replication by a replicase made of RNA. Activation of these tRNA-like structures with an amino acid could have been selected to facilitate replication. For example, a variant RNA replicase may have given rise to the first tRNA synthetase that could transfer an activated amino acid to a 3'-terminal tRNA-like structure (Weiner and Maizels, 1987). The demonstration of catalysis of a reaction at a carbon center by an RNA enzyme suggests that the RNA world could have expanded to include reactions of amino acids and other non-nucleic acid substrates prior to the involvement of proteins (Piccirilli *et al.*, 1992). With tRNAs, mRNAs, and ribozymes functioning as tRNA synthetases and peptidyl transferases, it is possible that information-driven protein synthesis could have been carried

out entirely by RNA.

The first ribosomes, also referred to as protoribosomes, may have been peptide specific (Watson *et al.*, 1987). The dependence of peptide-specific protoribosomes on built-in mRNAs would have significantly limited the variety of peptides that could be made. This limitation would then have provided the driving force for the evolution of a true ribosome capable of using separate mRNA molecules as templates for protein synthesis. The first ribosomal proteins were likely to be small peptides which interacted with rRNA to stabilize the structure and function of the rRNA (Wiener and Maizels, 1987; 1991). Therefore, the crude translation machinery that gave birth to the RNP world, a world made of both RNA and protein, must have been radically simpler than the ribosome of today.

Early in the evolution of the RNP world, very sophisticated catalytic reactions had to be carried out by RNA, where rudimentary proteins might have assisted the RNA catalysts. As the RNP world became capable of translating longer mRNAs accurately, proteins became more important and began to assume some of the independent activities of RNA. Once the RNP world reached a certain level of sophistication, living systems became so complex that RNA was no longer a suitable genetic material. It is reasoned that RNA is far more vulnerable to "spontaneous" hydrolysis than DNA; the 2'-hydroxyl group of RNA could be involved in the catalytic cleavage of the adjacent phosphoester bond (Watson *et al.*, 1987). This catalytic reaction is accelerated by the extreme conditions that might have been unavoidable on the primitive conditions of earth (e.g. high pH, high temperatures, and the presence of certain divalent cations from transition metals).

The actual conversion from RNA to DNA genomes would have been a complicated process. However, there is a simple and attractive scenario for this conversion. Once the RNP world reached a certain level of complexity, the production of dNDPs by ribonucleoside diphosphate reductase might have enabled the preexisting replication enzymes to copy the RNA into DNA genomes and then to transcribe these DNA genomes back into functional RNA molecules. It is also possible that the enzymatic activities of ribonucleoside reductase

(to synthesize DNA precursors), reverse transcriptase (to transcribe the preexisting RNA genomes into DNA), DNA polymerase (to replicate new DNA genomes), and DNA-dependent RNA polymerase (to transcribe DNA segments back into functional RNA molecules), present in modern organisms, evolved simultaneously. In the DNA world, right-handed double helical DNA serves as the genetic material and some ribonucleoprotein complexes have been retained as important cellular components, such as the ribosome and RNase P. The progenote, the common ancestor of all modern life, is likely to have evolved in this DNA world. Figure **1.1** shows a possible scheme for early evolution (Darnell and Doolittle, 1986).

After the discovery of split protein genes, it became clear that modern exons sometimes encode relatively large domains of protein structure (Gilbert, 1978). This observation gave rise to the hypothesis that the rate of acquisition of complexity (i.e., the creation of new genes with new functions) may be enhanced in the DNA world by exon shuffling. Such novel exon combinations can produce functional mRNAs because the junctions of the acquired units coincides precisely with the borders between exons and introns (Doolittle, 1985; Patthy, 1985). Other means of creating new genetic information such as gene duplication, overlapping genes, alternative splicing, and gene sharing are also important in evolution (Haldane, 1932; Ohno, 1970; Anderson *et al.*, 1981; Perlman and Butow, 1989; Piatigorsky *et al.*, 1988).

1.3 Gene Duplication and Formation of Gene Families

The importance of gene duplication in evolution was first noted by Haldane (1932) and Muller (1935). They suggested that a redundant duplicate of a gene, freed from its functional constraints, acquired divergent mutations and eventually emerged as a new gene with a distinct function. Although there are several other means by which a new gene can arise, gene duplication plays a major role in the evolution of new genes in contemporary biological systems (Ohno 1970; Anderson *et al.*, 1981; Perlman and Butow, 1989; Piatigorsky *et al*, 1988).

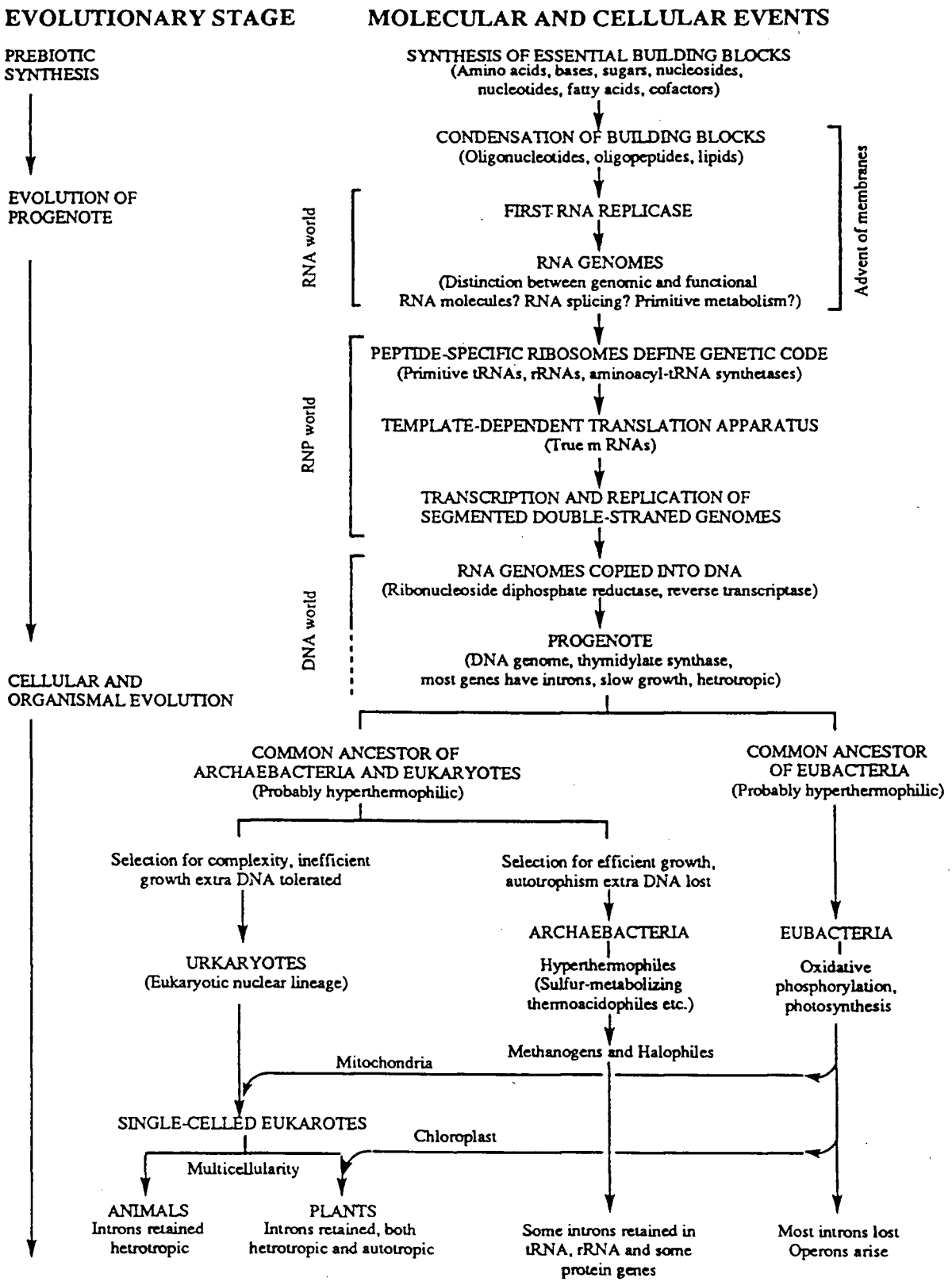

Figure 1.1 A possible scheme for early evolution (modified from Darnell and Doolittle, 1986)

Two genes are said to be paralogous if they are derived from a duplication event within a genome, but orthologous if they are derived from replication followed by a speciation event. A complete gene duplication produces two identical copies. How they will diverge varies from case to case. The copies may, for instance, retain the original function, enabling the organism to produce large quantities of the RNA or protein gene products. Alternatively, one of the copies may become a functionless pseudogene. Genetic novelties or new genes may emerge if one of the duplicates retains the original function whereas the other accumulates molecular changes such that, in time, it performs a rather different task. For example, the globin superfamily has experienced all of the evolutionary pathways that can occur in families consisting of repeated sequences. The evolutionary pathways are; (i) retention of original function (i.e. two identical copies of the α-gene in humans), (ii) acquisition of new function (i.e. the α- and β genes), and (iii) loss of function (i.e. the pseudo α-gene or pseudo β-gene).

Paralogous multigene families with invariant sequences like those encoding rRNA or histones are special cases of gene duplication, in which functional uniformity rather than diversity of gene copies is needed in the organisms (Ohta, 1991). For example, in *E. coli*, removal of one or two of the seven rRNA operons reduces slightly the doubling time of the organism and also causes increased expression of the remaining intact copies (Condon *et al.*, 1993). Histone proteins are among the slowest evolving proteins known and their sequences are conserved across species. For example, histone 3 interacts directly with either the DNA or other core histones in the formation of nucleosomes. It is reasonable to assume that there are very few possible substitutions that can occur without impeding the function of this protein. In addition, histone 3 must retain its strict compactness and its high alkalinity which are necessary for interaction with the acidic DNA molecule. As a consequence, histone 3 is resistant to molecular changes.

In the case of ribosomal RNA genes, they are essentially invariant within a species but between species they show sequence uniformity only within the functionally essential region of the rRNAs. Variable regions show different mutational rates across species. The primary nucleotide sequence of rRNA is essential for maintaining secondary and higher order

structural features and for participation in mRNA recognition, tRNA binding (in A, P, and E sites), ribosomal subunit association, proofreading, factor binding, antibiotic interactions, translational fidelity and suppression of frameshifting, termination and the peptidyl transferase function (for reviews, see Brimacombe *et al.*, 1990; Noller 1991). Changes in rRNA gene sequences that are fixed at all loci are usually considered to be nearly neutral or balanced by compensatory changes elsewhere in the molecule (for reviews see Noller *et al.*, 1990; Brimacombe *et al.*, 1990).

The phenomenon of compensatory mutation is a special case of molecular coevolution (or covariation, Dover and Flavell, 1984; Gerbi, 1985). Here a disrupting mutation in one strand of a stem is compensated by a complementary mutation in the other strand which preserves base pairing at that particular position in the rRNA secondary structure (Figure **1.2**). The non–canonical G•U and AoG base pairs are likely to be the intermediate stages in the transitional and transvertional compensatory mutation pathways respectively. Mutations which cause unpaired bases may form unstable structures. Many secondary structural elements are capable of withstanding the introduction of bulges or noncanonical base pairs, either because they are not destabilized or they are stabilized by other interactions such as RNA-protein or additional RNA-RNA interactions (Hancock and Dover 1988; Brimacombe *et al.*, 1990).

In general, it is thought that molecular interaction mechanisms, such as frequent unequal meiotic crossovers and gene conversions, which "correct" one gene sequence against another, are responsible for the concerted evolution of rRNA genes (concerted evolution). Recombinant DNA studies with yeast rDNA have shown that frequent unequal crossovers do, in fact, occur within the rRNA genes. When the rate of occurrence of such interactions is high, a uniform gene family (homogeneity) is expected, whereas when it is low, the family acquires variability (heterogeneity) (Ohta, 1991).

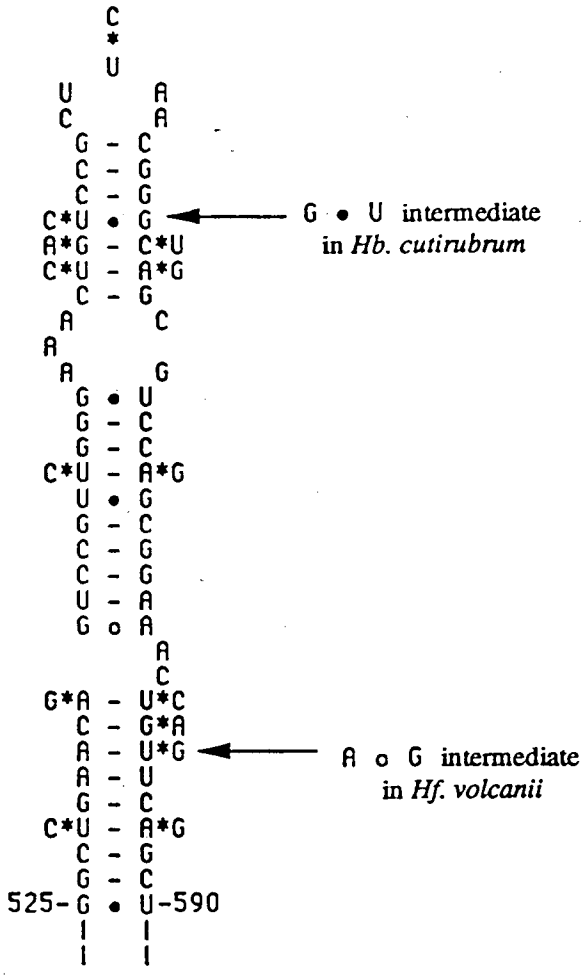

Figure **1.2** Compensatory base changes within rRNA secondary structures. A secondary structure within the 16S rRNA from *Hb. cutirubrum* is illustrated and the mutational differences observed in the corresponding positions from *Hf. volcanii* are indicated by asterisks (*). Here, a mutation in one strand of a stem is compensated by a mutation in the complementary strand which preserves base pairing at a particular position in the rRNA secondary structure. Canonical Watson-Crick base pairs (-) and non-canonical G•U and AoG base pairs are indicated. The G•U base pairs and AoG base pairs are considered to be the intermediate stages of the transitional and transversional compensatory mutations respectively.

1.4 Molecular Phylogeny and the Three Kingdom Classification

Ribosomal RNA sequences are the most useful chronometers for defining the genealogical relationships between organisms (Woese, 1987). They are universally distributed, functionally conserved and the divergence of their primary sequences exhibits relatively good clock-like behavior. Since different positions in their sequences change at very different rates, most phylogenetic relationships can be ascertained. These molecules are abundant, easy to isolate and sequence, and due to their large size, they usually provide a sufficient amount of data for statistically reliable analysis.

Phylogenetic trees based on the rRNA sequences show that the life on this planet can be divided into three domains; the eubacteria (or the bacteria), the archaebacteria (or the archaea) and the eukaryotes (or the eucarya) (Pace *et al.*, 1986; Woese and Olsen, 1986; Woese *et al.*, 1990; Pace, 1991). Each of these domains contains two or more kingdoms, for example, the eukaryotes contain Animalia, Plantae, Fungi and a number of others yet to be defined.

Figure **1.3** is a universal phylogenetic tree showing the relationships among the primary domains (Woese *et al.*, 1990). The root of the tree is seen to separate the eubacteria from the other two primary groups, making the archaea and eukaryotes specific (but distant) relatives. The position of the root was determined by comparing pairs of paralogous gene sequences (elongation factors Tu and G and the α– and β– subunits of ATPase) which are thought to have diverged by gene duplication before divergence of three primary domains (Iwabe *et al.*, 1989).

According to Woese *et al.*, (1990), the archaea comprise a monophyletic group which could be divided into two major phylogenetic lineages, euryarchaeota and crenarchaeota (Woese, 1987; Achenbach-Richter and Woese, 1988). The euryarchaeota is phenotypically heterogeneous and comprised of three methanogenic lineages; extreme halophiles, sulfur-reducing species (the genus *Archaeoglobus*), and two types of thermophiles (the genus *Thermoplasma* and the *Thermococcus-Pyrococcus* group, Woese, 1987; Achenbach-Richter *et al.*, 1987). Analysis of the small subunit rRNA sequences also indicates that

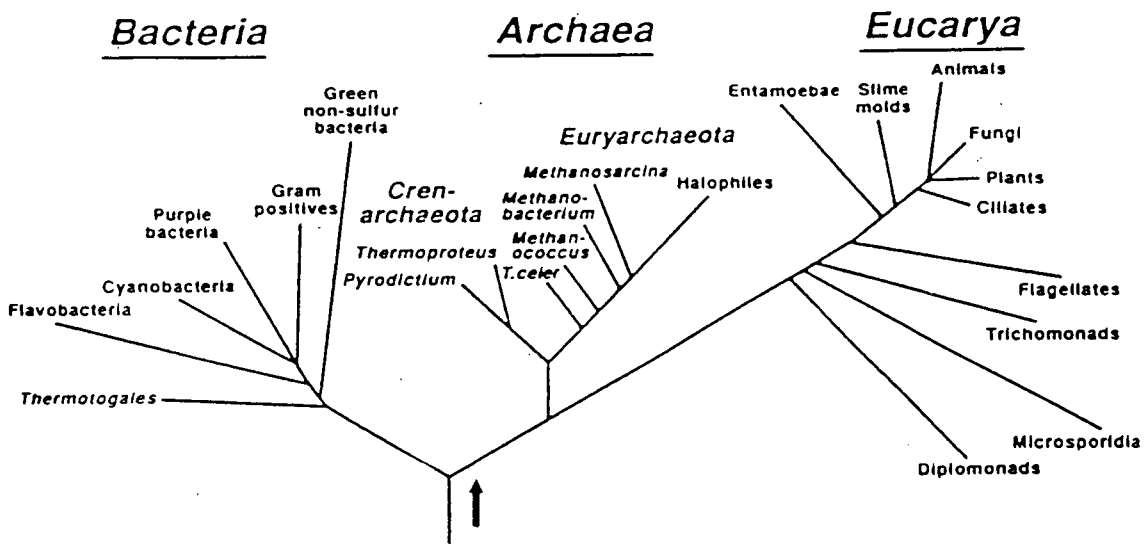

Figure **1.3** The universal phylogenetic tree. The tree is in rooted form and shows the three domains (adapted from Woese and Pace, 1993). Branching order and branch lengths are based upon 23 different small subunit (16S and 16S-like) rRNA sequence comparisons. The position of the root was determined by comparing sequences of pairs of paralogous genes that diverged from each other before the three primary lineages emerged from their common ancestral condition (Iwabe *et al.*, 1989). This rooting strategy (Schwartz and Dayhoff, 1978) uses one set of (aboriginally duplicated) genes as an out group for the other. The root of the universal tree is seen to lie between the bacteria on the one hand and the common lineage that the archaea and eukaryotes share on the other.

halophilic archaea have emerged from the methanogen group (Woese, 1987). The crenarchaeota comprise all of the sulphur-dependent extreme thermophiles. This classification has been reinforced by analysis of the sequences of RNA polymerase subunits (Puhler *et al.*, 1989b) and other biological molecules.

However, Lake (1988) has argued that archaea are paraphyletic and that they fall into three groups; extreme thermophiles or eocytes, methanogens, and extreme halophiles. The phylogenetic tree constructed by Lake from rRNA sequences of organisms is shown in Figure **1.4**. According to Lake's tree (eocyte tree), extreme thermophiles are more closely related to eukaryotes than to other archaeal groups whereas methanogens and halophiles

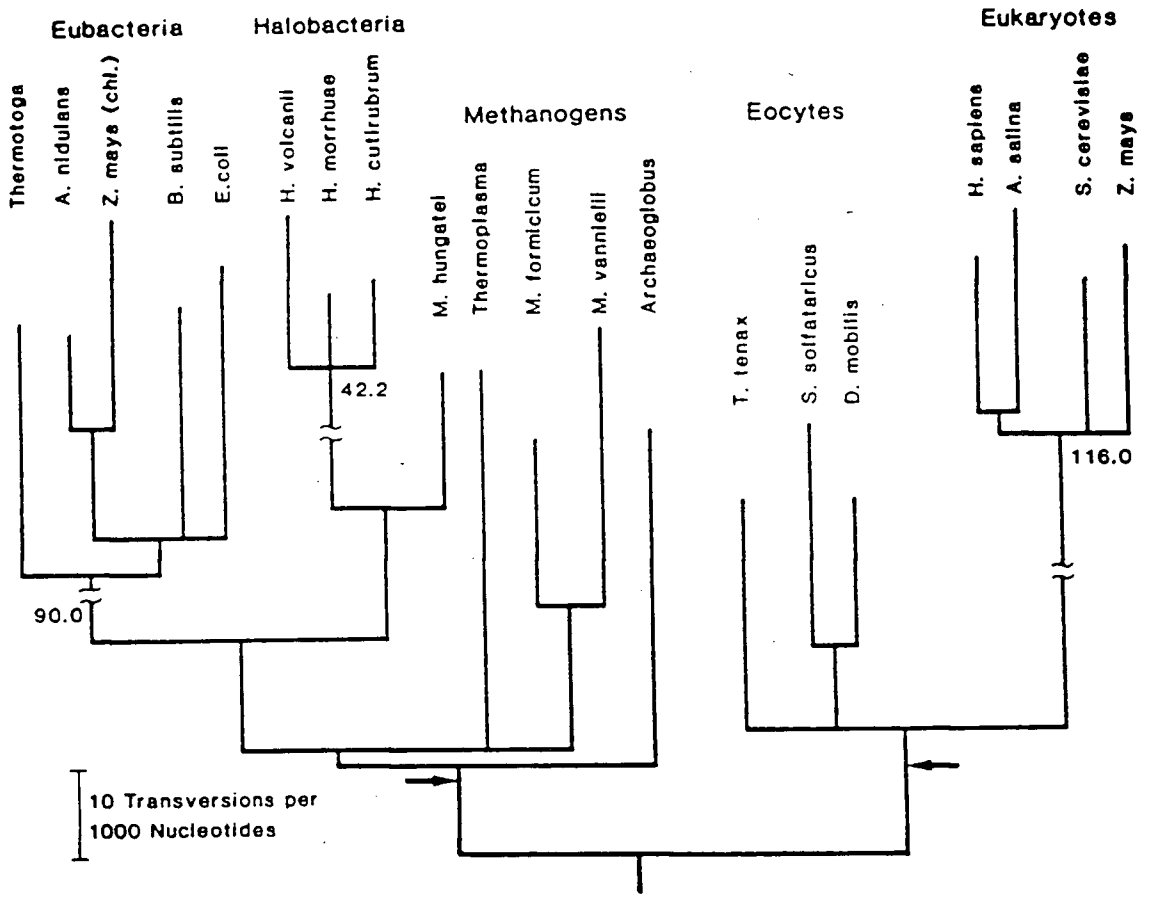

Figure **1.4** The rooted evolutionary tree that relates all known groups of extant organisms (adapted from Lake, 1988). The five groups are (from left to right) eubacteria, halobacteria, methanogens, eocytes and eukaryotes. The rate-independent technique of evolutionary parsimony (Lake, 1987a) in conjunction with the neighbourliness procedure (Fitch, 1981) were used to construct the unrooted tree. Branch lengths were determined by operator matrices (Lake, 1987b). Eubacteria, and eukaryotes (and halobacteria to a lesser extent) are "fast clock" organisms; hence, the length of the three long branches have been shortened to fit into the figure. Rooting was based on the simultaneous application of two criteria, parsimony (Fitch, 1977) and the relative rate test (Wilson *et al.*, 1977). The 5% significance level for the most parsimonious rooting is shown by the two horizontal arrows on either side of the root.

are closer to eubacteria. The root of the tree, based mainly on small-subunit rRNAs, was determined by evolutionary parsimony. The accuracy and reliability of various tree building procedures are continuously reevaluated (Woese and Olsen, 1986; Lake, 1988) and therefore it is difficult to determine, which one of the trees, if any, discussed above represents the correct phylogenetic tree.

1.5 Archaea

1.5.1 Characteristic Features

Although archaea share some properties with the eubacteria and some other properties with eukaryotes, they are considered to be a distinct group (Woese *et al.*, 1983; Lake, 1988; Table **1.1**). Archaea are an interesting group from a biological point of view, because many are adapted to extreme environments that are high in salt (halophiles) or high in temperature (thermophiles and hyperthermophiles), extreme in pH (acidophiles and alkalophiles), or anaerobic (methanogens).

The composition of the archaeal rRNAs probably represents the adaptative changes brought about by the extreme niches the organism came to inhabit (Woese *et al.*, 1984). In extreme thermophiles, RNAs are rich in G-C base pairs (75-80%) and post-transcriptionally modified nucleotides are five times more prevalent than in the other archaeal organisms (Woese *et al.*, 1984) which probably enhances their thermal stability. The extreme halophilic rRNAs are also rich in G-C base pairs. The methanogens have relatively high levels of guanosines and uridines (Østergaard *et al.*, 1987), resulting in many G•U base pairs (21% in *M. thermoautotrophicum* 23S rRNA). The primary sequences of all archaeal 16S rRNAs are at least 70% similar and this suggests that they diverged from a common ancestor between 2.0 and 3.5 billion years ago (Woese, 1987).

1.5.2 Halophiles

The extreme halophiles are aerobic and require a minimum NaCl concentration of 1.5 M with an optimum growth usually in the range of 2.5 M to 4.5 M. Environments inhabited by halophilic archaea contain Na^+, K^+, Mg^{++} and Ca^{++} ions as the predominant

Table **1.1** Characteristics of the three domains (Woese *et al.*, 1990). Some distinguishing features of the eubacteria, archaea and eukaryota are listed. Abbreviations are: Chloramphenicol (Cm); Anisomycin (Ani); Kanamycin (Kan); Pseudouracil (Y); α-amanitin (Ama) and Rifampicin (Rif).

Characteristic	Eubacteria	Archaea	Eukaryota
Cellular Organization	anucleate	anucleate	nucleated with organelles
Genome size (bp)	5×10^5 to 5×10^6	5×10^5 to 10^7	1.5×10^7 to 5×10^{11}
Membrane Lipids	ester-linked straight chain	ether-linked branched chain	ester-linked straight chain
Cell Walls	peptidoglycan	various but not peptidoglycan	various or none
Ribosomes			
(a) rRNA	5S, 16S, 23S	5S, 16S, 23S	5S, 5.8S, 18S, 28S
(b) diptheria toxin	insensitive	sensitive	sensitive
(c) antibiotic sensitivity	Cm^S Ani^R Kan^S	Cm^R Ani^S Kan^R	Cm^R Ani^S Kan^R
Transfer RNA			
(a) TYC loop	TYCG	1-methyl YYCG	TYCG
(b) 1-methyl adenine	absent	present	present
(c) 5'-end of initiator tRNA	5'-monophosphate	5'-triphosphate	5'-monophosphate
(d) initiator amino acid	N-formyl methionine	methionine	methionine
RNA Polymerase			
(a) number of types	1	1	3
(b) subunits	5	6-13	12 or greater
(c) antibiotic sensitivity	Ama^R Rif^S	Ama^R Rif^R	Ama $(PolII)^S$ $(PolI+III)^R$ Rif^R
mRNA	Uncapped	Uncapped	7-methyl G cap and polyadenylation

cations; the relative concentrations of these ions fluctuate due to their different solubility properties which vary with changing conditions. These organisms are known to concentrate predominantly K^+ ions, to actively extrude Na^+ ions, and to balance the overall intracellular cation concentration with that of the external environment (Christian and Waltho, 1962; Ginzburg *et al.*, 1970; Lanyi and Silverman, 1972). Members of the genus *Halobacterium* produce a unique purple membrane containing the transmembrane protein bacteriorodopsin, a light driven proton pump that generates ATP (Stockenius *et al.*, 1979; Stockenius and Bogomolni, 1982).

Halophiles exhibit various morphologies; they can be pleomorphic (*Haloarcula marismortui)* (Oren *et al.*, 1988), box-shaped (*Halobacterium sp. GN*), rod shaped (*Halobacterium salinarium* and its relatives including *Hb. cutirubrum* and *Hb. halobium*), or disc-shaped (*Haloferax volcanii)* (Mullakhanbhai and Larsen, 1975). Recently, it was shown that two members of the genus *Ha. vallismortis* and *Ha. hispanica* display complex cellular morphogenesis under unusual growth conditions (Cline and Doolittle, 1992). Morphogenesis of these species appears to be closely related to the uncharacterized halophilic archaea isolate *Gp9* (Wais 1985). These morphological types at least bear superficial similarity to examples found within the methanogenic genus *Methanosarcina* (Boone and Mah, 1990). Genetic exchange has recently been demonstrated for the first time in archaea using auxotrophic mutants of *Hf. volcanii*, a genomically stable and non-purple-membrane producing species (Mevarech and Wecyberger, 1985). It is believed that exchange proceeds by way of cell fusion to produce a transient diploid cell.

1.6 Archaeal rRNA Operons

1.6.1 Organization

The structural organization of archaeal rRNA operons exhibits considerable variability. The archaeal species characterized contain a low copy number ranging from one to four operons per genome (Table **1.2**). In most of the sulfur-dependent thermoacidophiles, the 5S

gene is unlinked to the 16S-23S rRNA operon and transcribed separately (Kjems and Garrett, 1987; Reiter *et al.*, 1987). *Thermoplasma acidophilum* is exceptional in that all three genes are unlinked and transcribed separately (Ree *et al.*, 1989; Larsen *et al.*, 1986). The rRNA operons from methanogens and halophiles share a common 5'-16S-23S-5S-3' gene organization that is reminiscent of the rRNA gene organization found in most eubacteria (Jarsch and Bock, 1985; Hui and Dennis, 1985). The primary rRNA operon transcripts also contain a $tRNA^{Ala}$ sequence in the 16S-23S intergenic spacer and a $tRNA^{Cys}$ sequence distal to the 5S gene. These tRNA sequences are inefficiently processed from the primary transcript; as a consequence, their stoichiometry relative to ribosomes is considerably less than one (Chant and Dennis, 1986). Eubacterial rRNA operons also contain various tRNA genes in these locations (Brosius *et al.*, 1981).

1.6.2 Transcription Signals

The rRNA genes of halophiles and *Desulfurococcus mobilis* (a thermophilic organism) are generally expressed from multiple promoters located upstream of the 16S rRNA gene (Mankin *et al.*, 1984; Dennis, 1985; and Larsen *et al.*, 1986). This complex and unusual promoter alignment distinguishes halophiles from methanogens and thermophiles, where the latter two generally contain only a single promoter. In archaea, the 16S and 23S rRNAs are usually cotranscribed. In halophiles, the 5S rRNA, an intergenic tRNA (usually coding for alanine), and in some cases, a distal tRNA gene (usually coding for cysteine), are also cotranscribed with 16S and 23S rRNAs. A weak internal promoter within the 16S-23S spacer region of halophile operons has been reported. It is assumed that the internal promoter rectifies the stoichiometric imbalance between 16S and 23S rRNAs caused by premature transcriptional termination (Mankin *et al.*, 1987; and Garrett *et al.*, 1993). In *S. acidocaldarius*, a cryptic promoter-like sequence in the 16S-23S spacer was reported, but transcripts initiated from this site could not be identified (Durovic and Dennis, 1994).

Transcription initiation has been studied *in vitro* using purified RNA polymerase and suitable promoter–containing template DNA (Zillig *et al.*, 1985; Gropp *et al.*, 1986; Reiter *et*

Table **1.2** Number of rRNA Operons in archaea. The number of rRNA operons vary from one in sulfur-dependent thermoacidophiles (Böck *et al.*, 1986), one to four in methanogens (Neuman *et al.*, 1983; Lechner *et al.*, 1985) and one to probably four in halophiles (Sanz *et al.*, 1988).

Organism	Number of rRNA Operons
Halophiles	
Halobacterium salinarium NCMB 777	1
Halococcus morrhuae NCMB 757	2
Haloarcula marismortui	3
Haloarcula californiae ATCC 33799	4
Methanogens	
Methanobacterium thermoautotrophicum	2
Methanothermus fervidus	2
Methanococcus vanielli	4
Sulfur-dependent Thermoacidophiles	
Sulfolobus acidocaldarius	1
Thermoproteus tenax	1
Desulfurococcus mobilis	1

al., 1990; Klenk *et al.*, 1992). Characterization of the archaeal transcription signals was initially confined to nucleotide sequence comparisons. Such analyses reveal two conserved sequence elements, one located between positions -38 and -25 (distal promoter element) and the other located between positions -11 and -2 (proximal promoter element). The distal promoter element contains the TATA-like (also called "box A") motif, located approximately

26 nucleotides upstream of the transcription start sites of archaea. By site specific mutagenesis, this element was shown to be important for the efficiency of transcription and start site selection (Reiter *et al.*, 1990). In some halophiles, another conserved element [consensus T(C or T)CGA] occurs ten nucleotides upstream of the "box A" sequence (Chang *et al.*, 1981; Das Sarma *et al.*, 1984; Dennis, 1985; Hamilton and Reeve, 1985).

The similarities in promoter structures parallel those found between the archaeal and eukaryotic RNA polymerases. Particularly, the similarities found with polymerases II (Klenk *et al.*, 1992; Huet *et al.*, 1983; Berghofer *et al.*, 1988; Leffers *et al.*, 1989; Pühler *et al.*, 1989a and; Pühler *et al.*, 1989b) suggest that the expression signals and transcribing enzymes are conserved between the two primary lineages. Transcription termination sites are not as well characterized as the promoter sites. Transcription termination sites for the polypeptide encoding genes of methanogens and in the *bop* gene of *Hb. halobium* have been determined by nuclease S1 protection assays (reviewed in Brown *et al.*, 1989). In most cases, termination occurs at sites that are similar to the *rho*-independent termination sites in *E. coli*. In the RNA transcripts, sequences that contain inverted repeat symmetry followed by poly dT regions are found downstream of many archaeal polypeptide-encoding genes.

Transcription termination signals downstream of the 16S-23S rRNA and the 5S rRNA operons of *D. mobilis* are similar to those found in the 16S-23S rRNA operon of *Thermoproteus tenax* and the distal (of the two) rRNA operon of *M. thermoautotrophicum*. These terminations occur within, or very close, to the pyrimidine-rich regions which contain CTCCT or CTCCCT sequences (Kjems and Garrett, 1987; Klein *et al.*, 1988). Termination of the transcripts of the *sod* (superoxide dismutase) family genes found in *Halobacterium cutirubrum*. sp. GRB, *Hf. volcanii* and *Ha. marismortui*, and the ribosomal protein genes, L11e and the tricistronic L1e-L10e-L11e present in *Hb. cutirubrum* were located near or within the T rich polypyrimidine tracks (May *et al.*, 1987; Shimmin and Dennis, 1989; Joshi and Dennis, 1993). Termination of the rRNA operon of the *M. vanielli* has been shown to occur at the beginning of a TTTTTAATTTT sequence located immediately downstream of the 5S rRNA-encoding sequence (Wich *et al.*, 1986a; Wich *et al.*, 1986b). In most of the characterized termination sites, longer T rich polypyrimidine tracts appear to result in more

efficient termination. On the basis of above examples, it is reasonable to suggest that the initial event of termination in archaea involves T rich polypyrimidine sequences. However, the mechanisms for this event have not been characterized in any of the cases described above. Helical structure in the RNA transcripts do not appear to be an absolute requirement, as it is missing from most termination sites.

1.6.3 Secondary Structural Organization of Transcripts

Primary transcripts of all the archaeal rRNA operons exhibit several partially conserved inverted repeats which presumably form stem loop structures. Figure **1.5** shows putative secondary structures within the regions surrounding the primary transcripts of 16S and 23S rRNAs of some archaea. Primary transcripts of methanogens and extreme thermophiles form simple structures, which generally include helices A, B, E, and F (Kjems and Garrett, 1990). However, transcripts of extreme halophiles, exemplified by *Hb. cutirubrum*, form the most complex structures. Three of the stem structures or helices C, D and G shown in Figure **1.5** are exclusive to extreme halophiles. The stem structures C and G directly preceding tRNAs and 5S RNAs respectively, may participate in their processing (Kjems and Garrett, 1990). The helix D, in the 16S-23S spacer, overlaps with an active promoter sequence of the DNA.

All the putative helices shown in Figure **1.5** are specific to either extreme halophiles or archaea. These helices are defined by their location relative to the processing stems, the tRNA, and the 5S RNA sequences. The transcripts of *Thermoproteus tenax* and *Pyrodictium occultum* exhibit exceptionally short spacers lacking helices E and F. The *T. tenax* transcript forms a bifurcated processing stem that contains two large rRNAs where the 5'- and 3'- ends of the mature RNAs lie within the processing stem. This relatively simple structure may reflect the minimal function of the 16S-23S RNA spacer, permitting the rRNAs to associate with their respective ribosomal protein and undergo maturation *in vivo* without mutual interference.

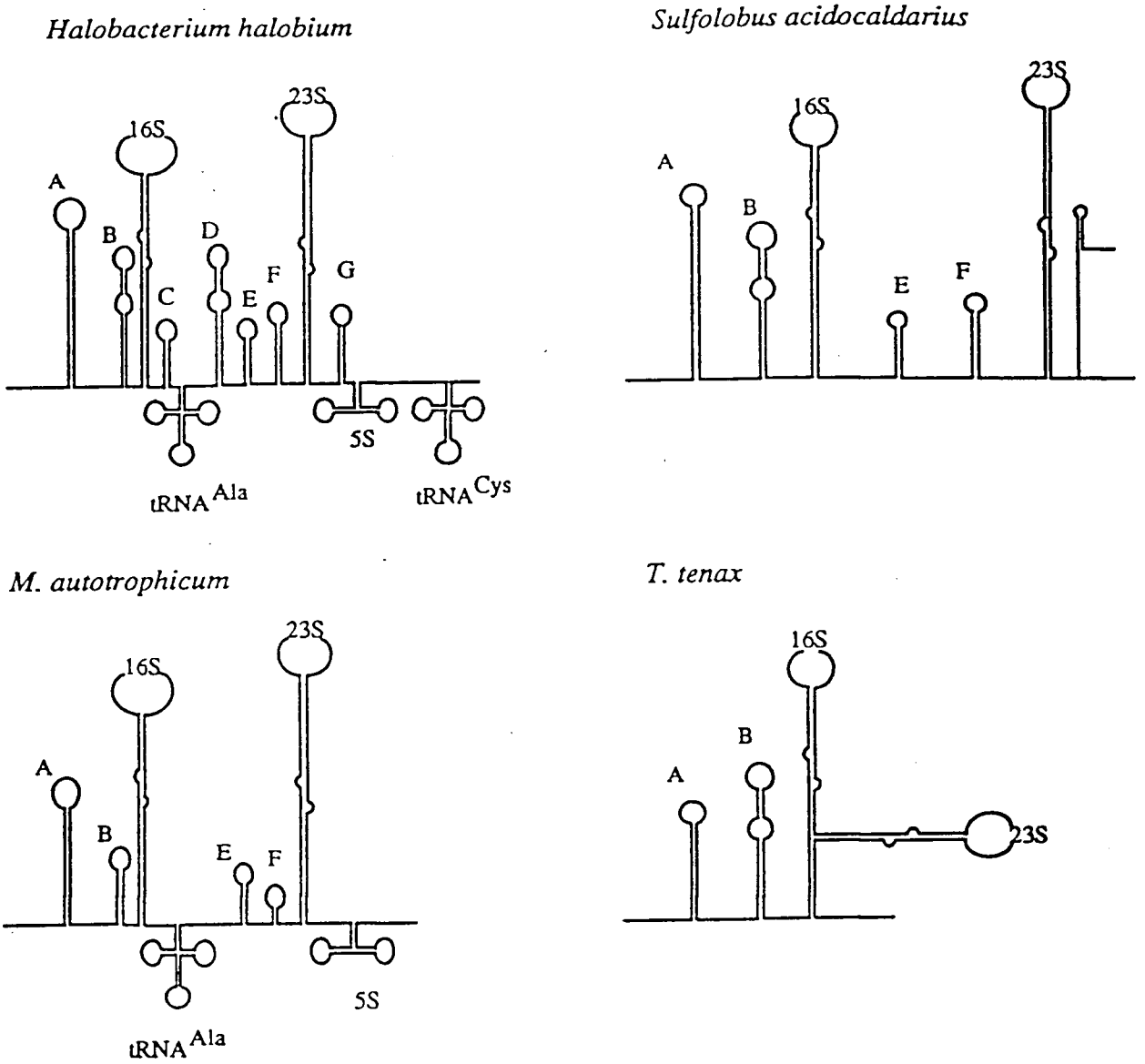

Figure **1.5** Gene Organization and Structure of Archaeal rRNA Transcripts. In *H. halobium* and *M. autotrophicum*, the organization of genes within the operons resemble those of eubacteria. The other two, sulfur-dependent thermoacidophile strains, represent the extremes of structural variability where the 5S genes are unlinked. In *T. tenax*, the 16S and 23S processing stems bifurcate from a common stem structure.

All of the long processing stems for the archaeal 16S and 23S rRNAs, with the exception of the 16S RNA stem of *S. acidocaldarius* and *T. acidophilum*, contain a well-defined structural motif which serves as the recognition site for the excision endonuclease. The aforementioned motif consists of two three-base bulges on opposite strands which are

precisely separated by four helical base pairs near the center of a much longer helical structure (Figure **1.6**). The nucleotides on the bulge are usually A or U and cleavage within the two bulges results in products with a 5'-hydroxyl and a 2' 3'-cyclic phosphate (Thompson *et al.*, 1989). Remarkably, this activity was first characterized as the excision endonuclease that removes intron sequences from the transcripts of archaeal tRNA and rRNA genes (Thompson and Daniels, 1988). The exon-intron junction in these cases forms the requisite "bulge-helix-bulge motif". The mature 16S and 23S rRNAs are formed by subsequent trimming at the 5'- and 3'-ends.

The rRNA operons in *E. coli* are similar to those of archaea in that the processing occurs within the long processing stems surrounding the 16S and 23S RNAs (Figure **1.6**) but the enzymology of precursor processing is somewhat different. In *E. coli* and related eubacteria, the enzyme RNase III is responsible for these initial excisions. The enzyme recognizes duplex RNA of generally undefined sequence, often with one or more bulge or unpaired nucleotides near the site of the cleavage. Cleavage occurs at staggered positions on each of the two strands and produces a 5'-phosphate and 3'-hydroxyl product (Court, 1993). Subsequent trimming at the 5'-and 3'-ends is required to produce the mature 16S and 23S rRNAs.

In the sulfur-dependent thermoacidophile, *S. acidocaldarius*, the processing and maturation of 23S rRNA appear to follow the typical archaeal pathway, utilizing a "bulge-helix-bulge" motif within the 23S processing helix (three–base bulges on opposite strands, precisely separated by four helical base pairs) as the substrate for an excision endonuclease. The pathway for processing and maturation of 16S rRNA is distinct and does not involve the "bulge-helix-bulge" motif in the processing stem (within the 5'-flanking region, the bulge motif contains only two nucleotides). Instead, the transcript is cleaved at several novel positions in the 5'-leader and in the 3'-intercistronic sequence. *In vitro* processing studies showed that the 16S helix is neither used nor required for leader processing. The same studies also showed that the complete maturation of the 5'-end of 16S rRNA can occur in the absence

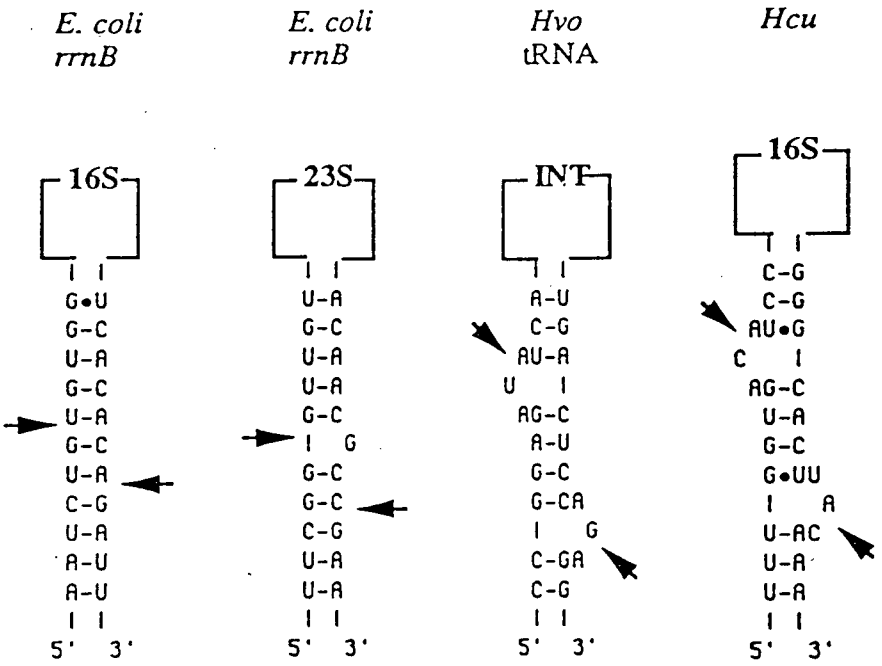

Figure **1.6** The secondary structures containing the recognition features for endonuclease activities that excise precursor 16S and 23S rRNAs from the primary transcripts. In eubacteria (e.g., *E. coli*), the processing enzyme RNase III recognizes duplex RNA of generally undefined sequence, often with one or more bulges or unpaired nucleotides near the site of cleavage. Cleavage sites are shown by arrows. The *Hvo* tRNA (tRNA intron from *Hf. volcanii*) and *Hcu* (the 16S processing stem from *Hb. cutirubrum*) represent the archaeal endonuclease recognition motif consists of two, three nucleotide bulges on opposite strands separated by four helical base pairs.(Thompson *et al.*, 1989). The two structures from *E. coli rrnB* represent the 16S and 23S processing stem structures.

of concomitant ribosome assembly and in the absence of all but the first 72 nucleotides of the 16S rRNA sequence (Durovic and Dennis, 1994).

1.6.4 Secondary Structures of Ribosomal RNAs

In terms of secondary structure, the various archaeal groups exhibit a mixture of unique, eubacterial and eukaryotic features built upon the same structural themes seen in all 5S rRNAs (Fox, 1985). The archaeal 5S rRNA models were categorized into four groups: I, II, III and IV. Group I contains members of the genera *Methanobacterium*,

Methanobrevibacter, and possibly *Methanococcus*. The extreme halophiles and *Methanosarcina* exhibit typical group II 5S rRNA (Figure **1.7**). Group III includes *Methanomicrobium* and *Methanospirillum*, and Group IV includes *Thermoplasma* and *Sulfolobus*. Table **1.3** provides a summary of the structural features seen in specific archaeal strains, as well as the eubacterial and eukaryotic cytoplasmic 5S rRNAs (Fox, 1985).

A detailed analysis of the secondary structure of the 16S-like rRNA from the three evolutionary lineages has shown that, in spite of the significant sequence variability, there exists a common secondary structure core, which is responsible for its main functions in the ribosomes (Mankin and Kopylov, 1981; Maly and Brimacombe, 1983; Woese, 1987). Using *E. coli* and *Hf. volcanii* as examples it was shown that, despite the differences in their primary sequences, the secondary structures of the 16S rRNAs are similar to each other (Woese *et al.*, 1983). In terms of primary structural homology, a comparison of the *Hf. volcanii* and *E. coli*, 16S rRNA genes gives a value of 58%, which is indeed low compared to the 75% seen in the very distant *E. coli-Zea mays* chloroplast (Fox, 1985). Figure **1.8** schematically summarizes the results of a structural comparison encompassing about 100 different small subunit rRNAs (SSU rRNAs) from a wide variety of organisms covering virtually the entire evolutionary spectrum (Raue' *et al.*, 1988).

Although the sequences of 23S rRNAs of the eubacteria (*E. coli*) and archaea (*Hb. halobium* or *Hb. cutirubrum*) differ significantly, the major part of their secondary structures are highly conserved (Mankin and Kagramanova, 1986; Raue' *et al.*, 1990). Relative to eubacterial rRNAs, small inserts are often found in archaeal rRNAs, at positions where the larger inserts occur in eukaryotic rRNAs (Leffers *et al.*, 1987). For example, in the 23S rRNA, two helices in domains I and II show archaeal-specific structural variations (Figure **1.9**). In the archaea, the size of the inserts ranges from 50-58 nucleotides in domain I and 26-37 nucleotides in domain II whereas in eukaryotes, the lengths of the inserts range from 183-836 and 49-123 nucleotides, respectively. Kingdom-specific nucleotides have been identified that are associated with antibiotic binding sites at functional centers in 23S-like RNAs: at the elongation factor-dependent GTPase center (thiostrepton domain II) they resemble bacteria

Table **1.3** Summary of characteristic structural features in various 5 S rRNAs. The 5S rRNAs are as follows: Eb, consensus eubacterial 5S rRNA (excluding organelles and mycoplasmas): 1, *Methanobrevibacter ruminantium*; 2, *Methanobrevibacter smithii*; 3, *M. thermoautotrophicum*; 4, *Methanobacterium bryanti*; 5, *M. vanielli*; 6, *Hc. morrhuae*; 7, *Hb. cutirubrum*; 8, *Hf. volcanii*; 9, *M. barkeri;* 10, *Methanogenium marisnigri;* 11, *Methanogenium cariaci*; 12, *Methanospirillum hungatei*; 13, *Methanomicrobium mobile*; Ec, consensus eukaryotic 5S rRNA; 14, *T. acidophilum*; 15, *S. acidocaldarius*. Structural property refers to the distance between positions 66 and 78 in the eubacterial consensus sequence.

	RNA Source																
		Group I					Group II				Group III				Gp IV		
Structural Property	Eb	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Ec
Helix I looped-out base	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-
Helix II extension	+	+	+	+	+	+	+	+	+	+	-	-	-	-	+	+	+
Helix III loop length	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	13	12
CGAAC- sequence	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	-
Helix IV looped-out base	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+
Extended helix V obviously feasible	-	-	-	+	+	+	-	+	+	-	-	+	+	+	+	+	+
Helices I/V or II/V may be coaxail	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+
Number of bases between main portion of helix II and beginning of helix IV	13	17	17	17	16	16	16	16	16	16	16	16	16	16	15	14	15
Total number of bases in helix IV loop and stem	19/20	20	16	19	18	16	21	21	21	21	21	21	21	21	21	21	21

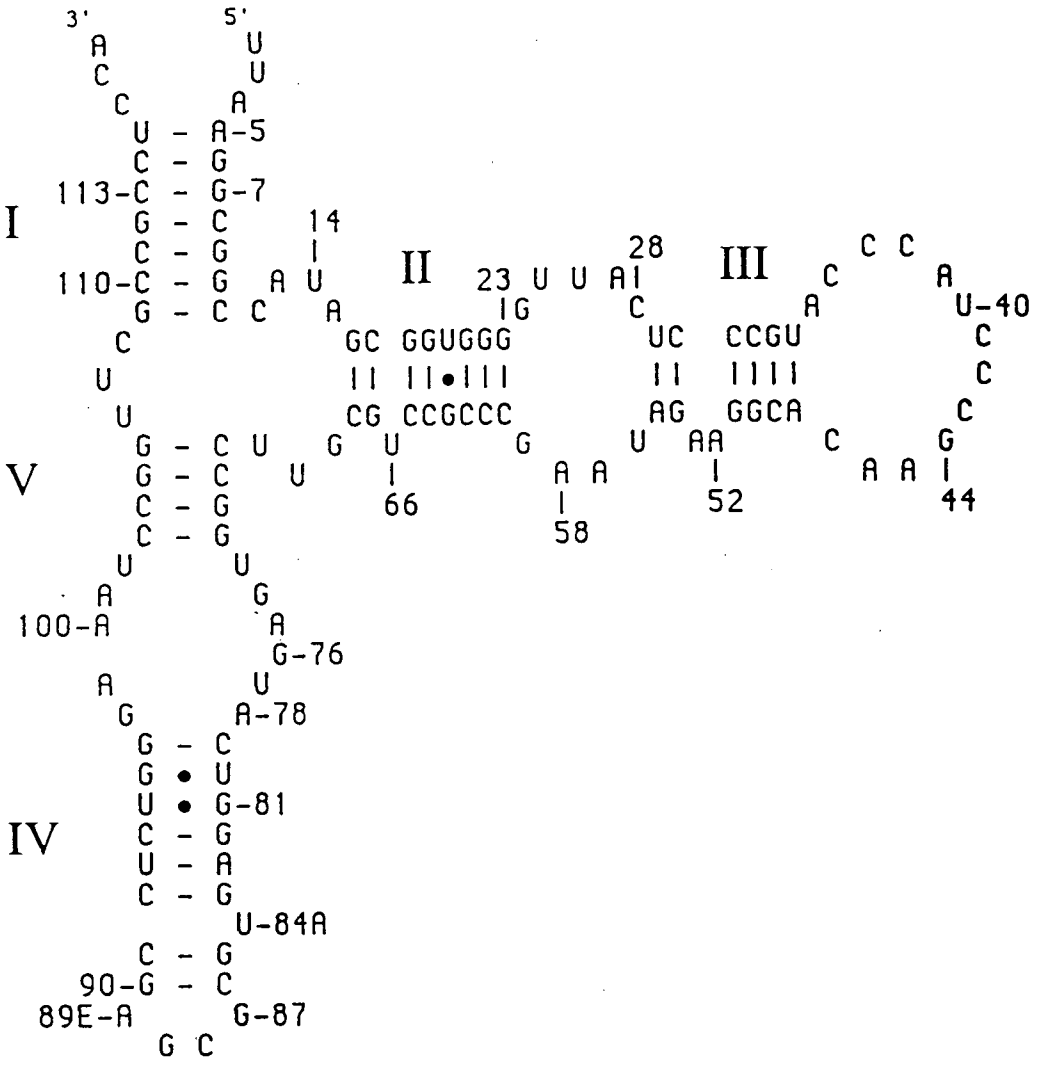

Figure **1.7** *Halobacterium cutirubrum* 5S rRNA (Fox, 1985), a typical group II 5S rRNA. Helix I has a larger number of unpaired bases at its terminii than is usual, but the molecule is totally eubacteria-like in its structural features through position 69. Helix IV has a four-base loop as is typical of eukaryotes. The looped-out base in helix IV is followed by two base pairs instead of the three that characterize the eukaryotes in this region. This feature is best regarded as uniquely archaeal and the relevant position is accordingly numbered 84A, which distinguishes it from the eukaryotic analog, 89E.

and in the peptidyl transferase center (erythromycin-domain V), and at the putative aminoacyl tRNA site (α-sarcin domain VI) they resemble eukaryotic large subunit rRNAs (Raue' *et al.*, 1990; Jarsch and Böck, 1985; Leffers *et al.*, 1987). Figure **1.9** shows a schematic

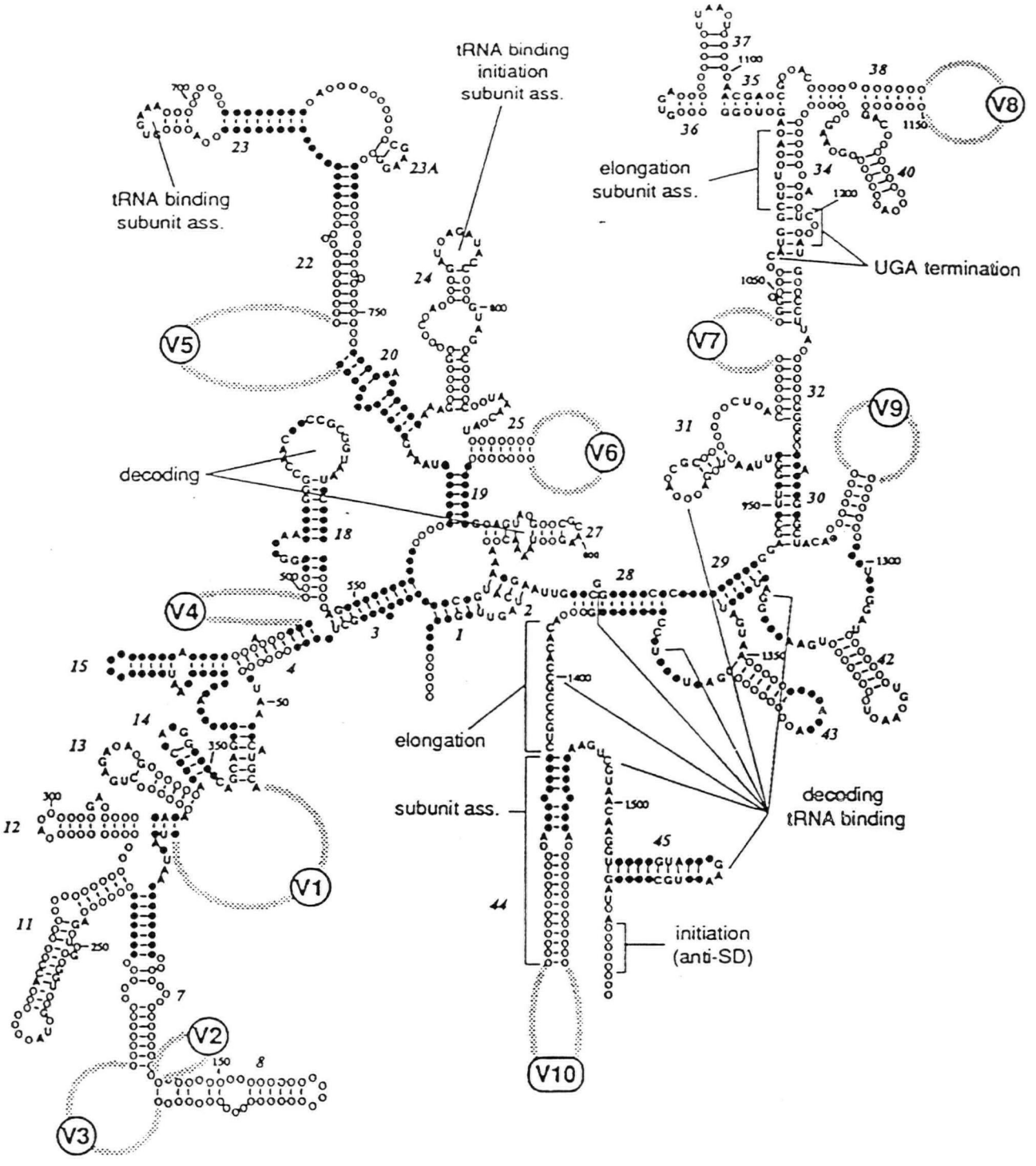

Figure **1.8** Functional regions in small subunit rRNA (adapted from Raue' *et al.*, 1990). The figure shows a schematic representation of the SSU rRNA structure based on the secondary structural model for *E. coli* 16S rRNA. Variable expansion segments are numbered consecutively from the 5'-end. Filled circles indicate regions conserved in all classes of ribosomes; open circles correspond to regions conserved in all classes except the mitochondrial ones. Highly conserved nucleotides (present in >90% of all known SSU rRNA sequences) are indicated. Helices are numbered according to Raue' *et al.* (1988). The functional regions are also labelled.

representation of the LSU rRNA structure based on the secondary structure model for *E. coli* 23S rRNA where the universally conserved regions, expansion segments and functional regions are indicated.

In two thermophilic archaea *S. marinus* and *D. mobilis*, introns have been discovered in the 23S rRNAs (Kjems and Garrett, 1987; Kjems and Garrett, 1990). These introns, like those of eukaryotes, are located within the functionally important domains IV and V of LSU rRNA. The similarity of the locations of the archaeal and eukaryotic LSU rRNA introns suggests that they are functionally related. In archaea, splicing of the rRNA introns occurs at a "bulge-helix-bulge" motif that forms at the exon-intron junctions (Kjems and Garrett, 1987). This structural feature, first characterized for the intron-containing tRNAs of extreme halophiles and thermophiles (Kjems and Garrett, 1990), also occurs in the processing stems of the archaeal 16S and 23S RNAs (see section **1.6.3**).

1.7 Ribosomal RNA Heterogeneity

Distinct population (or subpopulations) of ribosomes, i.e., ribosome heterogeneity in an organism, may result from a change in one of the ribosomal components during its life cycle. The presence or absence of a new ribosomal protein (r-protein) or rRNA, a unique covalent modification, or a stoichiometric difference in one cell type or developmental stage may all lead to ribosome heterogeneity in eukaryotic organisms (Ramagopal, 1992).

Several examples of low level rRNA sequence heterogeneity in a variety of species have been reported. In *E.coli,* which has seven rRNA operons in its genome, eight sites of sequence heterogeneity have been observed in the analysis of bulk 23S rRNA (Branlant *et al.*, 1981). Three rRNA operons from *Rhodobacter sphaeroides* have been sequenced; the three 16S genes are identical whereas the 23S genes exhibit a single nucleotide substitution in one gene and three single nucleotide deletions in the second gene. The 5S genes exhibit slightly higher heterogeneity; one 5S gene differs at four positions from the other two genes (Dryden and Kaplan, 1990). The mitochondrial genome of *Tetrahymena* contains two genes encoding

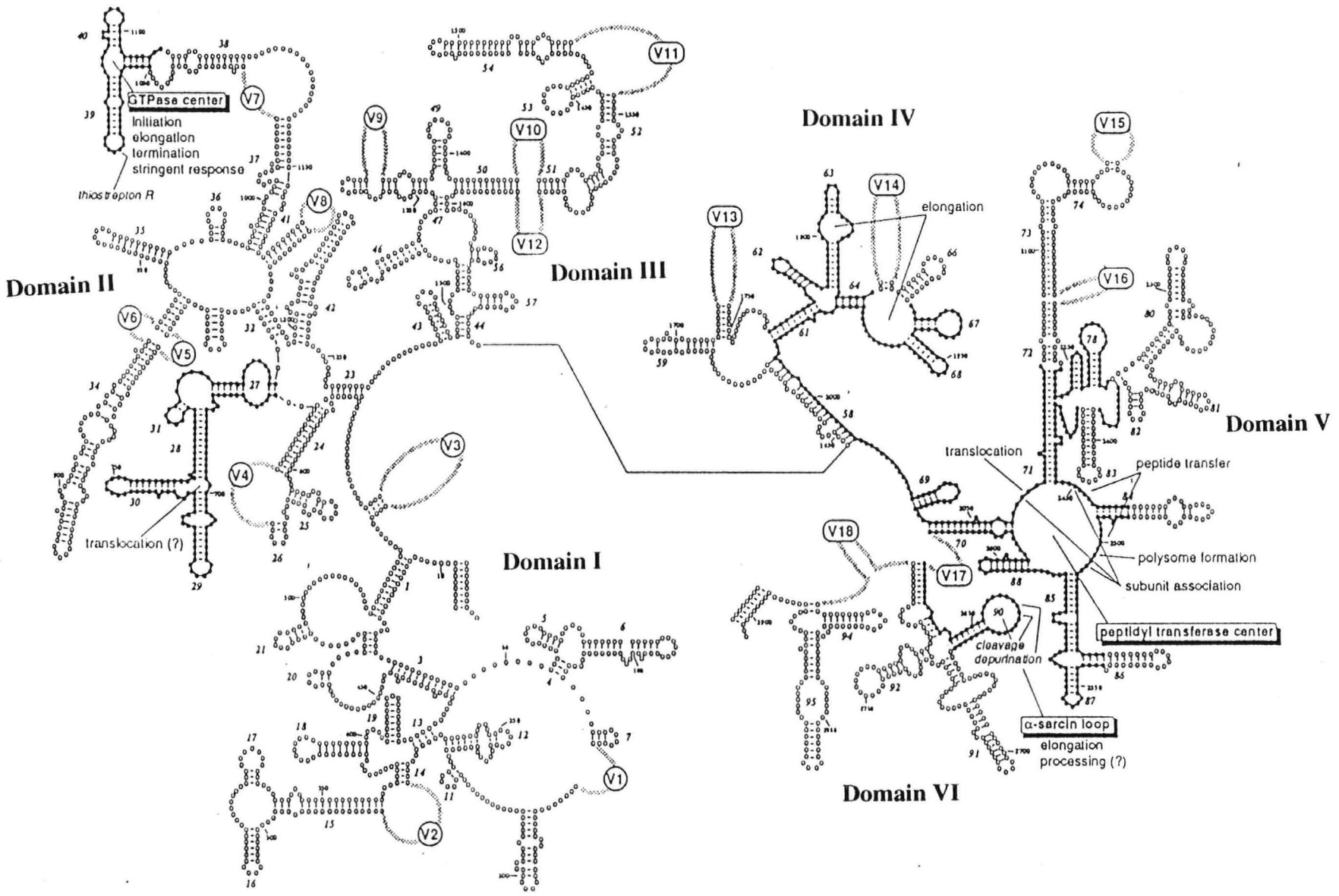

Figure **1.9** Functional regions in large subunit (LSU) rRNA (adapted from Raue' *et al.*, 1990). The figure shows a schematic representation of the LSU rRNA structure based on the secondary structure model for *E. coli* 23S rRNA. Variable expansion segments are numbered consecutively from the 5'-end. Filled circles indicate universally conserved regions. Functional regions are also labelled in the diagram.

large subunit rRNA; the two genes differ at five out of 2595 positions, and both genes are expressed (Heinonin *et al.*, 1990). The nucleotide substitutions discussed above are not present in functionally important positions of the rRNA structures. These heterogeneous gene products are believed to be an intermediate stage in the process of concerted evolution (Hancock and Dover, 1990).

A more interesting example has been described in the blood parasite, *Plasmodium berghei,* where two 18S rRNA encoding genes referred to as A- and C- genes (Gunderson *et al.*, 1987) are expressed in gametocytes and sporozoites respectively. The two A genes and two C genes differ at 72 positions out of 2059 (i.e., 96.5% identical). It has been suggested that this differential expression may play a role in the types of proteins synthesized during different developmental stages of the complex life cycle of *Plasmodium berghei.* It was shown by Walters *et al.*, (1989) that the switch from A to C gene expression involves the control of rRNA processing. In gametocytes, precursor transcripts from C-type genes are not processed and ribosomes contain predominantly A type RNA. In the zygote and the early ookinate, transcription and processing of the rRNA from C type genes is accelerated. As C-type ribosomes accumulate, a defined and limited pattern of breakdown of the dominant A-type ribosomes occurs, during which conserved and functionally important sequences involved in the elongation and termination of translation are targeted. By the late oocyst stage, the A-type ribosomes have essentially been replaced by C-type ribosomes. Sequence comparison of the C and A type genes showed that the distribution of the differences between the two sequences is not random. There are only four differences in the 3'-domain, whereas considerable variation is found in the middle and 5'-domains of 18S rRNA. The secondary structural analysis suggests that these differences are restricted to regions that are otherwise phylogenetically variable (Gunderson *et al.*, 1987; Woese *et al.*, 193). In eukaryotes, several examples of ribosome heterogeneity, caused by ribosomal proteins were also reported (Ramagopal, 1992).

It was shown by Sanz *et al.*, (1988) that the genome of *Ha. marismortui* contains three ribosomal RNA operons in its genome. However, Mevarech *et al.*, (1989) showed that *Ha.*

marismortui contains two unlinked rRNA operons in its genome and are different in several respects. These two operons, designated as *rrnA* and *rrnB*, were previously cloned as separate genomic restriction fragments. Preliminary characterization by RFLP (Restriction Fragment Length Polymorphism) indicated that the two copies of the 16S and 23S rRNA genes differ at a number of positions. Furthermore, analysis of the 5'-flanking regions indicated that the *rrnB* operon was unusual in that it contained only a single recognizable promoter and apparently lacked the consensus processing site within the 5'-portion of the putative 16S processing stem.

1.8 Research Objectives

To provide a more complete characterization of the ribosomal RNA operons, *rrnA* and *rrnB*, present in the genome of *Ha. marismortui*, several approaches were made. The genomic DNA derived from the progeny of a single cell was probed for the presence of both *rrnA* and *rrnB* operons. Then, the nucleotide sequences of the flanking, coding and intergenic spacer regions of the *rrnA* and *rrnB* operons were determined and analysed for the conservation of primary sequence and secondary structure. Using structural models, primary transcripts and processing intermediates derived from the two operons were characterized. The *rrnA* and *rrnB* gene sequences (rRNAs) were also compared to the corresponding gene sequences from a number of other halophilic archaeal species in order to understand the significance of their sequence divergence from each other and from other halophilic sequences. Using nuclease S1 protection assays, it was shown that the 16S rRNAs from the *rrnA* and *rrnB* operons were expressed and present in active 70S ribosomes. To determine the start and end sites of the transcripts and the processing intermediates, *in vivo* transcript analysis were performed using total RNAs from *Ha. marismortui*. Finally, the rRNA gene sequences from the *rrnA* and *rrnB* operons were aligned to the corresponding gene sequences from a number of other halophiles, a methanogen and an eubacterium (*E. coli*) and phylogenetic analysis was performed using PAUP (Phylogenetic Analysis Using Parsimony).

CHAPTER 2

Materials and Methods

2.1 Materials

Bacterial culture components—yeast extract, tryptone, casamino acids, and agar were purchased from Difco Laboratories, ampicillin from Sigma Chemical Co., IPTG from GIBCO BRL, Xgal from Bethesda Research Laboratories or Biosynth AG., and D-glucose from BDH. *E. coli* strains JM101 and JM109 are available from Promega. Strain DH5α is available from Bethesda Research Laboratories.

The vector pBR322 is available from New England Biolabs. Vector phages, M13mp18 and M13mp19 are available from Pharmacia. The vectors pGEM-3Zf(+/-), pGEM-5Zf(+/-), and pGEM-7Zf(+/-), and the helper phage R408 were purchased from Promega.

Most of the restriction enzymes and DNA modifying enzymes were purchased from Pharmacia or New England Biolabs (NEB). Enzymes purchased from other sources are: T4 polynucleotide kinase from PL Biochemicals, Pharmacia; modified T7 DNA polymerase (Sequenase) and shrimp alkaline phosphatase (SAP) from United States Biochemical Corp. (USB); AMV reverse transcriptase from Boehringer Mannheim Canada (BMC); lysozyme and ribonuclease A from Sigma; exonuclease III from Promega; T4 DNA ligase and S1 nuclease from Pharmacia; and Klenow fragment from BMC.

Deoxyribonucleoside triphosphates, dideoxyribonucleoside triphosphates, and (1) -phosphorothioate deoxynucleotide triphosphates were obtained from Pharmacia. Radioactive α (^{32}P) dNTPs, γ (^{32}P) NTPs and α (^{35}S)dNTPs were obtained from Dupont New England Nuclear (NEN) Research Products. The T7 and SP6 primers were purchased from NEB. Oligonucleotides used for sequencing or probes for genes or for primer extension analyses were obtained from Dr. Carl Woese's laboratory (University of Illinois, Urbana Champaign, USA) or synthesized by T. Atkinson (University of British Columbia) and Dr. Ivan

Sadowski's laboratory (University of British Columbia). The synthesized oligonucleotides obtained from UBC were supplied as lyophilized crude powders. Crude oligonucleotides were purified by C18 Sep-Pak reverse phase chromatography and quantified by measuring the A260 (Atkinson and Smith, 1984).

Acrylamide and N, N'- methylenebisacrylamide were purchased from BioRad Laboratories, genetic technology grade agarose was purchased from Schwartz/Mann Biotech, cesium chloride was purchased from Cabot chemicals and β-mercaptoethanol (β-ME) was purchased from Matheson Coleman & Bell. All other chemicals were purchased from either BDH, Fisher, or Sigma.

Hybond-N nylon membranes and filters, and Hybond M & G paper were purchased from Amersham. Films for autoradiography (X-Omat and XAR) were purchased from Kodak and films for visualization of ethidium bromide stained DNA were purchased from Polaroid.

2..2 Methods

2.2.1 Media and Culture Conditions

Haloarcula marismortui was grown at 42°C in a rich medium as described by Oren *et al.*, (1988). The medium contains the following composition (g/l): NaCl, 206; $MgSO_4 \cdot 7H_2O$, 36; KCl; 0.373; $CaCl_2 \cdot 2H_2O$, 0.5; $MnCl_2$, 0.013 mg/l, and yeast extract, 5.0; pH 7.0. The plates for plating the cells were prepared by adding 15 g/l agar into the medium and autoclaving.

2.2.2 Isolation of *Haloarcula marismortui* Genomic DNA

For isolation of *Ha marismortui* DNA, a 1 L culture was grown to an A_{600} of 1.0 - 1.5 and pelleted. The cells were washed in a solution containing 204 g of NaCl and 39.6 g of $MgSO_4 \cdot 7H_2O$ per liter. The washed bacteria were resuspended in 100 ml of 10 mM $MgCl_2$ -10 mM Tris hydrochloride (pH 7.5), aliquoted into five tubes and each tube was extracted twice with phenol. NaCl was added to a final concentration of 0.5 M and the solution was

cooled on ice. 40 ml of 95% ethanol was added to each tube and the DNA was collected by spooling on a sealed pasteur pipets. The DNA was washed twice with absolute ethanol, air dried, dissolved in TE (10 mM Tris-HCl (pH 7.5), 1 mM EDTA), and stored at -20°C.

2.2.3 5'-end Labelling of Oligonucleotides

Oligonucleotides (about 5 pmol) were 5'-end-labelled at 37°C for 40 minutes with 1 unit of T4 Polynucleotide Kinase and 50 μCi of 3000Ci $mmol^{-1}$ [$\gamma^{32}P$] ATP in 20 μl of kinase buffer (1 X = 0.1 M Tris-HCl, pH 8.0, 5 mM DTT, 10 mM $MgCl_2$). The reaction was stopped by adding 1 μl of 0.5 M EDTA (pH 8.0) and heating at 65°C for 5 minutes. Carrier tRNA (8 μg) and 80 μl of distilled water were added and the preparation then precipitated twice with 2.5 volumes of 95% ethanol in the presence of 0.3 M NaOAC and dried. Each sample was counted by Cerenkov and then dissolved in 10-20 μl TE buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA).

2.2.4 Southern Blot Analysis

DNAs were digested with various restriction endonucleases (as described in section **3.2**) and separated by size on agarose gels alongside ^{32}P-labelled size standards. The DNA was denatured *in situ* by soaking the gel for 30 minutes in denaturing solution (1.5 M NaCl, 0.5 M NaOH). The DNA was then transferred to a nylon membrane by blotting with transfer buffer (1.5 M NaCl, 0.25 M NaOH) for 12 hours. The membrane was washed in 2 X SSPE (SSPE: 180 mM NaCl, 100 mM NaH_2PO_4, 10 mM EDTA (pH 7.7)) and dried at room temperature for an hour. The DNA was crosslinked to the membrane by a 2 minute exposure to UV light. The membranes were prehybridized for one hour in hybridizing solution (5 X SSPE, 5 X Denhardt's solution (Denhardt's solution: 0.02% w/v BSA, 0.02% w/v Ficoll, 0.02% w/v polyvinylpyrollidone, 0.5% w/v SDS) before the addition of the radioactive DNA probe. The hybridization was carried out at 37°C overnight and the membrane was washed with low and medium stringency buffers respectively.

2.2.5 Plasmids and Phage Preparations

Small-scale plasmid DNA purifications were done by using the "Magic™ Minipreps DNA Purification System" from Promega or by the alkaline lysis method (Maniatis *et al.*, 1982). Large scale plasmid preparations were done by the methods described by Maniatis *et al.* (1982). The *E. coli* strains DH5α and JM101 were used for plasmid propagation and generation of single stranded phagemids respectively (Dente and Cortese, 1983). The helper phage R408, was used in conjunction with pGEM (Promega) and dideoxy sequencing reactions using KITS (i. e. sequenase) contain modified procedures as described by Sanger *et al.* (1977).

2.2.6 Gel Electrophoresis

2.2.6.1 Native Gel Electrophoresis

Restriction endonuclease digested DNA fragments were separated electrophoretically through agarose (genetic technology grade) slab gels (0.8% or 1%) or 5% nondenaturing polyacrylamide gels (acrylamide: N N'-methylene bisacrylamide, ratio of 39:2). The buffer used for electrophoresis was 0.5 X TBE buffer (45 mM Tris, 45 mM Boric acid, 1 mM EDTA, pH 8.2) or 1 X AGB buffer (for agarose only; 20 mM Sodium acetate, 40 mM Tris, 1 mM EDTA, pH adjusted to 8.0 with glacial acetic acid) and the electrophoresis was performed at between 100 and 150 volts. In order to visualize the bands, the gels were run in the presence of 0.25 μg/ml ethidium bromide (for agarose only) or were stained with it after electrophoresis. The bands were cut out using a clean scalpel and the DNAs were extracted by using the Gene Clean Kit from Pharmacia (for agarose gels) or electroeluted against 0.5 X TBE buffer (for acrylamide gels) and then ethanol precipitated.

2.2.6.2 Denaturing Polyacrylamide Gel Electrophoresis

The DNA products from sequencing reactions, nuclease S1 transcript mapping reactions, and from primer extensions were separated on 0.35 mm denaturing vertical

polyacrylamide gels. The gels (ratio of acrylamide to N N'-methylene bisacrylamide 39:2) were composed of the following: (i) acrylamide concentration-6% or 8%, (ii) 0.5 X TBE buffer, (iii) 8.3 M urea, and (iv) the reagents used for polymerization (260 µl $(NH_4)_2S_2O_8$ and 30µl TEMED) for a total volume of 50ml. Electrophoresis was carried out at 32 watts constant power. After electrophoresis, gels were dried on to Whatmann 3 MM paper and exposed to Kodak x-ray film.

2.2.7 Ligation

For cohesive-end and blunt-end ligations, 40 fmoles of plasmid vector DNA and 3 fold molar excess of insert DNA were ligated in 10 µl of reaction mix containing 1X ligase buffer (20 mM Tris-HCl, pH 7.6, 5 mM $MgCl_2$, 5 mM DTT, 50 µg/µl BSA) and 0.1 Weiss unit of T4 DNA ligase. The incubation was carried out at 16°C overnight.

2.2.8 Transformation

E. coli competent cells were prepared by the $CaCl_2$ method for transformation. The *E. coli* cells were grown in 100 ml YT medium to an O.D at A_{600} of about 0.5-0.8 (1 cm path length) and harvested at 4°C at 4000 rpm in the Sorvall centrifuge for 10 minutes. The cells were collected by centrifugation, resuspended in 50 mM $CaCl_2$ (0.5 x volume of the original culture,) and incubated on ice for 40 minutes. The cells were then centrifuged in the Sorvall centrifuge at 4°C at the speed of 4000 rpm for 10 minutes, resuspended in 2 ml of 50 mM $CaCl_2$, 15% glycerol, and incubated on ice for 1 hour. The cells were either used directly or pipetted into small aliquots (100 µl), frozen quickly in dry ice, and stored at -70°C for later use. Competent cells were gently mixed with 2-4 fmoles of DNA, incubated on ice for 30 minutes, heat-shocked at 42°C for 45 seconds, incubated on ice for a further 2 minutes, and then directly plated on YT-agar containing ampicillin, X-gal and IPTG.

2.2.9 Exonuclease III Deletions

Bidirectional deletions of insert DNA were constructed in the plasmids pGEM 7^+ and pGEM 3^+ using exonuclease III (Henikoff, 1987). This technique was used in the sequencing of the 23S rRNAs from the *rrnA* and *rrnB* operons where *XbaI* (gives a 5'-cohesive end) and *SphI* (gives a 3'-protecting end) enzymes were used. A molar excess of exonuclease III digests DNA in the 3'- to 5'- direction in a time- and temperature-dependent manner. Nuclease S1 was used to digest the remaining single stranded DNA, klenow fragment along with a mix of the four deoxyribonucleotides was used to fill in any recessed 3'-ends, and T4 DNA ligase was used to circularize the plasmid. The plasmids were transformed into DH5α strains. The colonies were picked randomly and grown into 5ml YT media. The plasmids were isolated and screened for the insert sizes and those that contained deletions were sequenced using sequencing KITS (sequenase) contain modified version of the dideoxy sequencing method (Sanger *et al.*, 1977).

2.2.10 DNA Sequencing

Most of the 16S rRNA sequencing reactions were carried out using the primers obtained from Carl Woese's laboratory. The 23S rRNA genes, 5S rRNA genes, intergenic spacer regions, and the 5S 3'-flanking regions were sequenced by using either sequence specific primers or exo-nuclease III deleted template DNAs. The deleted DNAs were sequenced by the dideoxy chain termination method employing either T7 or SP6 primers. Both single- and double-stranded DNA molecules were employed as templates (Sanger *et al.*, 1977; and Zhang *et al.*, 1988). For most of the double stranded sequencing, a 7-deaza-2'-deoxyguanosine 5'-triphosphate (C^7dGTP) sequencing kit from Pharmacia was used.

2.2.11 Isolation of Total RNA

Total cellular *Ha. marismortui* RNA was isolated using the boiling SDS-lysis method described by Chant and Dennis (1986). Briefly, cells were rapidly cooled on ice, collected by centrifugation, resuspended, and lysed in an SDS-containing buffer (100°C, for 15-30 sec);

RNA was extracted with phenol and precipitated with ethanol. DNA contaminants were removed by ultracentrifugation through a 5.7 M CsCl block gradient. The RNA was resuspended in TE buffer and stored at - 20°C.

2.2.12 Isolation of RNAs From Mature Ribosomes

A one liter culture was grown to an A_{460} of 0.5 and pelleted. The cells were resuspended in 1 ml ribosomal preparation buffer (3.4 M KCl, 100 mM $MgCl_2$, 6 mM 2-mercaptoethanol and 10 mM Tris-HCl, pH 7.6; Shevaik *et al.*, 1985) and disrupted by passage through a french pressure cell. The lysate was centrifuged at 2500 rpm for 5 minutes and again centrifuged at 9500 rpm for 15 minutes to remove debris and unbroken cells. The supernatant was layered on to a 5-30% sucrose density gradient in ribosomal preparation buffer and spun at 27,000 r.p.m using an SW27 rotor for 6 hours at 10°C. The fractions containing 30S, 50S, and 70S were collected, dialysed against ribosomal preparation buffer for 12 hours, extracted with phenol/$CHCl_3$ three times in the presence of 1% SDS, ethanol precipitated with 2.5 times the volume with 95% ethanol, and purified by CsCl centrifugation.

2.2.13 5'-End Labelling of DNA Fragments

Restriction DNA fragments containing 5'-overhang ends were dephosphorylated with SAP (Shrimp Alkaline Phosphatase) at 37°C in a solution containing 10 mM $MgCl_2$, 20 mM Tris-HCl, pH 8.0, for 1 hour. The reaction mixture was then heated for 30 minutes at 65°C to denature the SAP. Once the mixture was cooled down to room temperature, Tris-HCl (pH 8.0), DTT, and spermidine were added to a final concentration of 30 mM, 5 mM, and 0.1 mM respectively, and then 0.1 units of T4 poly nucleotide kinase (T4 PNK) and 50 µCi [$\gamma\,^{32}P$] ATP were also added. The reaction mixture was incubated at 37°C for 40 minutes. The labelled fragments were precipitated using salt and 2 volumes of 95% ethanol and the radioactivity was measured by Cerenkov counting. If necessary (especially for Maxam and Gilbert sequencing), the fragments were then digested with appropriate restriction enzymes and purified by isolating the product fragments on a 5% non denaturing gel. After the

digestion with restriction enzyme, if the unwanted product fragment was comparatively small, the product fragments were ethanol precipitated after phenol/ chloroform and used directly for Maxam and Gilbert sequencing.

2.2.14 3'-End Labelling of DNA Fragments

Restriction DNA fragments containing recessed 3'-ends were end labelled using the Klenow fragment of the *E. coli* DNA polymerase I and the appropriate [α^{32}P]dNTP (specific activity of 3000 Ci/mmol, 10 mCi/ml). The DNA fragments of approximately 100 - 500 ng were labelled at the 3'-end in a solution containing 1x Klenow buffer (10 mM NaCl, 10 mM Tris-HCl, pH 7.5, 7 mM $MgCl_2$), 15 µl of 3000 Ci/mmol [α^{32}P] dNTP and 5-7 units of Klenow at room temperature for 15 minutes.

2.2.15 Maxam and Gilbert Sequencing

Chemical sequencing of DNA fragments was carried out according to the method described by Maxam and Gilbert (1980). The DNA fragment was either labelled uniquely at the 3'-end by Klenow enzyme or labelled at both 5'-ends with T4 PNK and digested internally with a restriction endonuclease to yield product fragments with a uniquely labelled end. The labelled fragments were then purified electrophoretically through a polyacrylamide gel. The Cerenkov count was taken and the DNA was dissolved in dH_2O to give 50K cpm/reaction. Since chemical sequencing was only performed to establish a size ladder with nucleotide precision for transcript mapping on DNA probes whose sequences had already been determined, only two sequencing reactions were required. End-labelled DNA was spotted on to strips of Hybond M&G filter for the G reaction and double the amount was used for the A+G reaction. The filters were rinsed twice with dH_2O, once with 95% ethanol, and air dried for 3 minutes. The G reaction was performed by treating with 50 mM ammonium formate (pH 3.5) and 0.7% dimethyl sulfate for 40 seconds at room temperature. The A+G reaction was performed by treating with 66% formic acid for 10 minutes. The reactions were terminated by removing the strips from the reaction solutions, washing twice with dH_2O,

once with ethanol, and air drying. Cleavage and removal of the DNA from the filter was achieved by submersion of the filter in 100 µl of 10% piperidine and heating at 90°C for 30 minutes. After treating with piperidine, the strips were removed and the mixture was lyophilyzed. The pellet was lyophilized with 40 µl of dH_2O twice and the DNA was finally resuspended in 4 µl of dH_2O.

2.2.16 Transcript Mapping

2.2.16.1 Nuclease S1 protection Analysis of the Total RNA

Nuclease S1 protection of *Ha. marismortui* rRNA transcripts was performed as follows. Fragments of DNA, labelled at either the 5'-or the 3'-end, were ethanol precipitated in the presence of 1 – 2 µg of total RNA and resuspended in 4 µl of S1 hybridization buffer (40 mM PIPES (pH 6.8), 400 mM NaCl, 1 mM EDTA) and 16 µl of deionized formamide. The samples were denatured for 15 minutes at 80°C and hybridized to the RNA for three hours at temperatures optimized for G-C content (52°C-59°C). Unhybridized single stranded DNA and RNA was digested by the addition of 300 µl nuclease S1 digestion buffer containing 280 mM NaCl, 30 mM NaOAc (pH 4.4), 4.5 mM $ZnCl_2$, 6 µg nonspecific single stranded DNA, and nuclease S1 (200-500 U/ml). The reaction mixture was incubated for 30 minutes at 37°C. The protected products were isopropanol precipitated, dried under vacuum, resuspended in 4 µl of dH_2O, and then 4 µl of FDM was added. The samples were heat denatured and loaded on denaturing polyacrylamide gels alongside the DNA probe and the Maxam and Gilbert sequencing ladder.

2.2.16.2 Nuclease Protection Analysis of the Active 70S Ribosomal RNAs

Nuclease S1 protection experiments were carried out as described above. This experiment was performed to show that the 16S rRNAs derived from both operons (*rrnA* and *rrnB*) during exponential growth were present in the active 70S ribosomes (see section **2.2.12**). The restriction fragments used for this experiment were the homologous but nonidentical 272 nucleotide long *SacII-SmaI* fragments (nucleotide positions 463-734 in 16S

rRNA sequence) from the *Ha. marismortui* pHC8 (*rrnA*) and pHH10 (*rrnB*) clones and from the *Hb. cutirubrum* p4W clone. The fragments were 5'-end labelled at the *SmaI* site (nucleotide position 736) as described above. Approximately 20-30 ng of the respective fragments (DNA excess; between 10^5 and 10^6 dpm per assay) were hybridized to 200 ng of total RNA (in the case of *Hb. cutirubrum)* or 50ng of RNA isolated from 70S ribosomes (in the case of *Ha. marismortui*) in S1 hybridization buffer for three hours at temperatures between 50°C and 60°C. Hybrids were digested with 200-500 units/ml of nuclease S1 at 32-35°C for 30 minutes, and protected products were analysed for length by electrophoresis in 8% polyacrylamide gels. As a negative control, total tRNA from *Saccharomyces cerevisiae* was used in place of *Ha. marismortui* or *Hb. cutirubrum* RNA to protect the end labelled DNA fragments. In these experiments, the distribution and intensity of partial protection products were sensitive to (i) hybridization temperature, (ii) the S1 concentration, and (iii) to a lesser extent, the digestion temperature. The above conditions were optimized in order to obtain unambiguous results.

2.2.16.3 Primer Extensions

Primer extension reactions by AMV reverse transcriptase enzyme were carried out according to the method described by the manufacturer. A sequence-specific oligonucleotide primer (~2ng) labelled at the 5'-end was ethanol precipitated with ~ 5 µgs of total RNA, resuspended in 10 µl reverse transcriptase annealing buffer, annealed for 5 minutes at 65°C, cooled slowly to the reaction temperature (37-45°C), and incubated for an additional hour. Primer extensions were carried out by the addition of 10 µl of reaction buffer (10 mM MgCl2, 1 mM dNTPs, and 10 mM β-mercaptoethanol), 5 units of AMV reverse transcriptase, and 5 units of RNase inhibitor, followed by incubation at appropriate temperatures (either 37°C or 42°C) for 30 minutes. After that time, more AMV reverse transcriptase was added and incubated for an additional half an hour at 37°C, 42°C, or 52°C. The reaction was stopped by treating the reaction mixture with 1 µl of RNase A (10 mg/ml), incubating at 37°C for 15 minutes, and then by adding the stop mix (10 mM EDTA and 300 mM NaOAc). The

products were ethanol precipitated, dried under vacuum, resuspended in 4 µl dH_2O and 4 µl FDM, denatured, and run on denaturing polyacrylamide gels alongside a template sequencing ladder.

2.2.17 Sequence Alignment of the 16S 5'-Flanking Regions

The 5'-flanking sequence from the *rrnA*, *rrnB* and the rRNA operon from *Hb. cutirubrum* were first aligned using Clustal V method (Des Higgins, European Molecular Biology Laboratory, Germany). Although the stretch of nearly identical sequences from the three operons (Mevarech *et al.*, 1989; Dennis, 1991) were aligned (Figure **3.10**), the conserved sequence motifs within multiple promoter sequences were not recognized by the program. In the *rrnA* and *Hb. cutirubrum* operons, these nearly identical sequences extend beyond the primary processing sites at positions -86 and -110, respectively. The *rrnB* operon also contains the major portion of this conserved sequence between positions -94 and -145; however, it lacks the processing site located at the end of the sequence. In order to get the maximum alignment, the promoter sequences were then aligned manually. The *rrnA* promoter sequences P_1, P_2, P_3 and P_4 were aligned to the *Hb. cutirubrum* promoters P_3, P_4, P_5 and P_6, respectively. The *rrnB* promoter sequences P and Px were aligned to the *Hb. cutirubrum* promoters P_5 and P_2, respectively. When aligning the promoter sequences, first, the Box A and Box B sequences were aligned and then the gaps were introduced in order to get the maximum alignment.

2.2.18 Molecular Phylogeny Method

The phylogenetic method used in this study is maximum parsimony analysis or PAUP analysis with stepwise addition. The PAUP analysis package was created by David Swofford (1993). The principle of this method is minimum evolution, involving the identification of a tree that requires the smallest number of evolutionary changes to explain the differences observed among the operational taxonomic units under study. Such trees are assumed to be closest to the true tree and are therefore preferred (Fitch, 1971). To use this method, the sequences are first aligned using Clustal V method (packaged by Des Higgins, European Molecular Biology Laboratory, Heidelberg, Germany). Gaps are then introduced within the

sequences in order to optimize the alignment. The gaps represent insertion or deletion events in the sequences that are assumed to have occurred at these sites since their divergence from the common ancestor. Next, "informative sites" on the sequence alignment are determined. A site is considered to be phylogenetically informative only if it favours some trees over others. The minimum number of substitutions at each informative site is then calculated and a tree describing these substitutions is assigned to this site. Finally, the incidence of each of the trees over all of the informative sites are calculated; the one that occurs most frequently is considered to be the most parsimonious and best represent the evolutionary relationship of the sequences under consideration (Felsenstein, 1988).

Since this is a statistical method, there are inevitable errors; for example, underestimation of distances between sequences due to multiple mutations at a single site. In order to estimate the uncertainty in the overall tree elucidation, the data are subjected to resampling analysis. The most widely used resampling method in phylogenetic analyses is "bootstrapping" (Felsenstein, 1988). This consists of resampling with replacement; randomly selected samples in the set (corresponding to sites in aligned sequences) are analysed and returned to the set so that some sites of the sequences under consideration are sampled more than once and some are never sampled. If such resampling is performed frequently enough, it is possible to determine the confidence limits (P values) and hence the significance of any inferred evolutionary branch.

CHAPTER 3

The Gene Organization, Sequence Heterogeneity, Expression and RNA Processing of the *rrnA* and *rrnB* Operons From the Halophilic Archaeon *Haloarcula marismortui.*

3.1 Introduction

The initial report on ribosomal RNA heterogenity in the multicopy rRNA operons from the genome of a halophilic archaeon, *Ha. marismortui* was published by Mevarech *et al.* (1989). The observation of two nonidentical rRNA operons in the aforementioned report poses certain important questions: Why does a unicellular organism need two types of rRNA operons? Are the products of both rRNA genes from the two operons expressed and present in the active ribosomes? Are these operons expressed differentially at different stages of the growth cycle? How did the sequence heterogeneity originate and are there any selective forces which have allowed it to be maintained and propagated? Are there any unique mechanisms involved in the expression and processing of the rRNA genes from these operons? Are the two types of rRNA genes advantageous for the survival of *Ha. marismortui*? Are the presence of heterogeneous genes in this organism an adaptation to a changing environment?

To address some of these questions, a number of experiments were performed. The total number of rRNA operons present in the genome of this organism was determined by Southern hybridization analysis of genomic DNA isolated from progeny of a single cell. Next, complete sequence analysis of the *rrnA* and *rrnB* operons was performed. The two sequences were compared to ascertain the level of heterogeneity at the nucleotide level. Comparisons of the rRNA sequences from related halophiles were also performed to get more information about structurally and/or functionally important sites. The sequences of rRNA genes, spacers, and leaders were subsequently compared with related halophile sequences in order to identify conserved elements that are presumably important for function or expression. The primary

transcription products of the *rrnA* and *rrnB* operons, and the intermediates produced by a variety of steps (processing and maturation of rRNAs) involved in the assembly of functional ribosomal particles, were studied using nuclease S1 mapping and primer extension analyses. A nuclease S1 protection assay was developed using the rRNAs isolated from active 70S ribosomes with two different DNA probes which are specific for the 16S rRNAs from the *rrnA* and *rrnB* operons. This assay was used to demonstrate that both operons are active and their product 16S rRNAs are assembled into active 30S ribosome subunits.

3.2 Results and Discussion

3.2.1 Number of Operons

Mevarech *et al.*, (1989) reported that *Ha. marismortui* contains two non-adjacent ribosomal RNA operons, designated as *rrnA* and *rrnB*, within its genome. In their analyses, Southern hybridization to a *HindIII* digest of total genomic DNA identified two bands of 20-kbp and 10-kbp in length. The 10-kbp *HindIII-HindIII* fragment containing *rrnB* and a 8-kbp *HindIII-ClaI* fragment containing the *rrnA* (derived from the 20-kbp band) were cloned separately into the plasmid pBR322 (Mevarech *et al.*, 1989).

Since pBR 322 is a low copy number plasmid, the 10-kbp *Hind III–Hind III* fragment and the 8.0-kbp *Hind III- Cla I* fragments were isolated and subcloned into pGEM 7^+ vectors. and using these plasmid DNAs, restriction endonuclease digestions were performed in order to confirm that the two clones are from the *rrnA* and *rrnB* operons of *Ha. marismortui*. The results are summarized in Figure **3.1** and Appendix. The initial RFLP characterization of the *rrnA* (pHC8) and *rrnB* (pHH10) clones are indicated as A and B, respectively in Figure **3.1**. Digestion of the the *rrnA* and *rrnB* operons with the same restriction enzymes gave different products [e.g., lanes C+X (A) and C+X (B)] suggest the presence of polymorphic restriction sites in these operons. In Figure **3.1**, most of the bands below 2.8–kbp matches with the restriction maps shown below (see Appendix), however, some of the restriction sites (for example, *XhoI* and *PstI* sites; see Appendix) were not detected in the restriction maps obtained

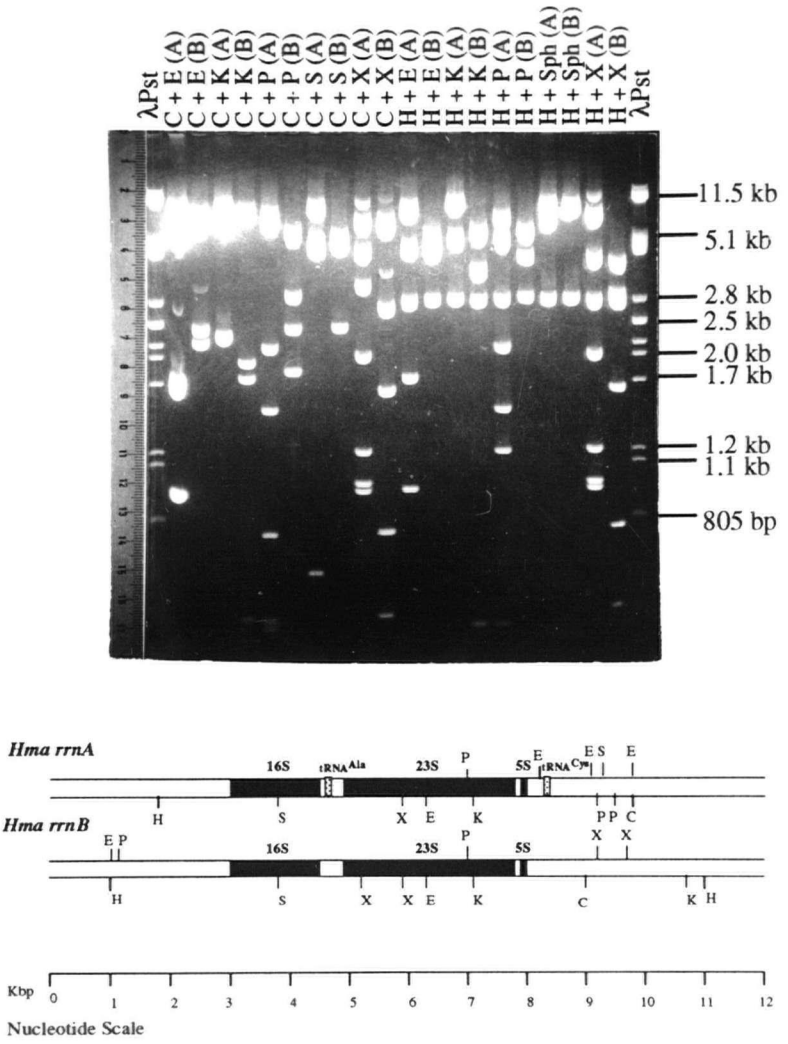

Figure **3.1** Restriction endonuclease digestion of the pHC8 (*rrnA*) and pHH10 (*rrnB*) operons from *Ha. marismortui*. The 8.0-kbp *HindIII-ClaI* fragment from the *rrnA* operon (A) and the 10-kbp *HindIII- HindIII* fragment from the *rrnB* operon (B), subcloned into plasmid pGEM 7^+, were used in this analysis. The following restriction enzymes were chosen so that they cut the plasmid at unique positions; *ClaI* (C), *HindIII* (H), *EcoRI* (E), *KpnI* (K), *PstI* (P), *SmaI* (S), *XhoI* (X) and *SphI* (Sph). The *rrnA* (A) and *rrnB* (B) operons were digested with similar enzymes and each pair of digestion products shows an RFLP [e.g., see H+E (A) and H+E (B)]. The fragments from λ DNA digested with *PstI* was used as a size marker.

by Mevarech *et al.*(1989). Also, some of the larger bands gave unexpected sizes. This may be due to overloading of the samples and incomplete digestion of the larger fragments. In addition,

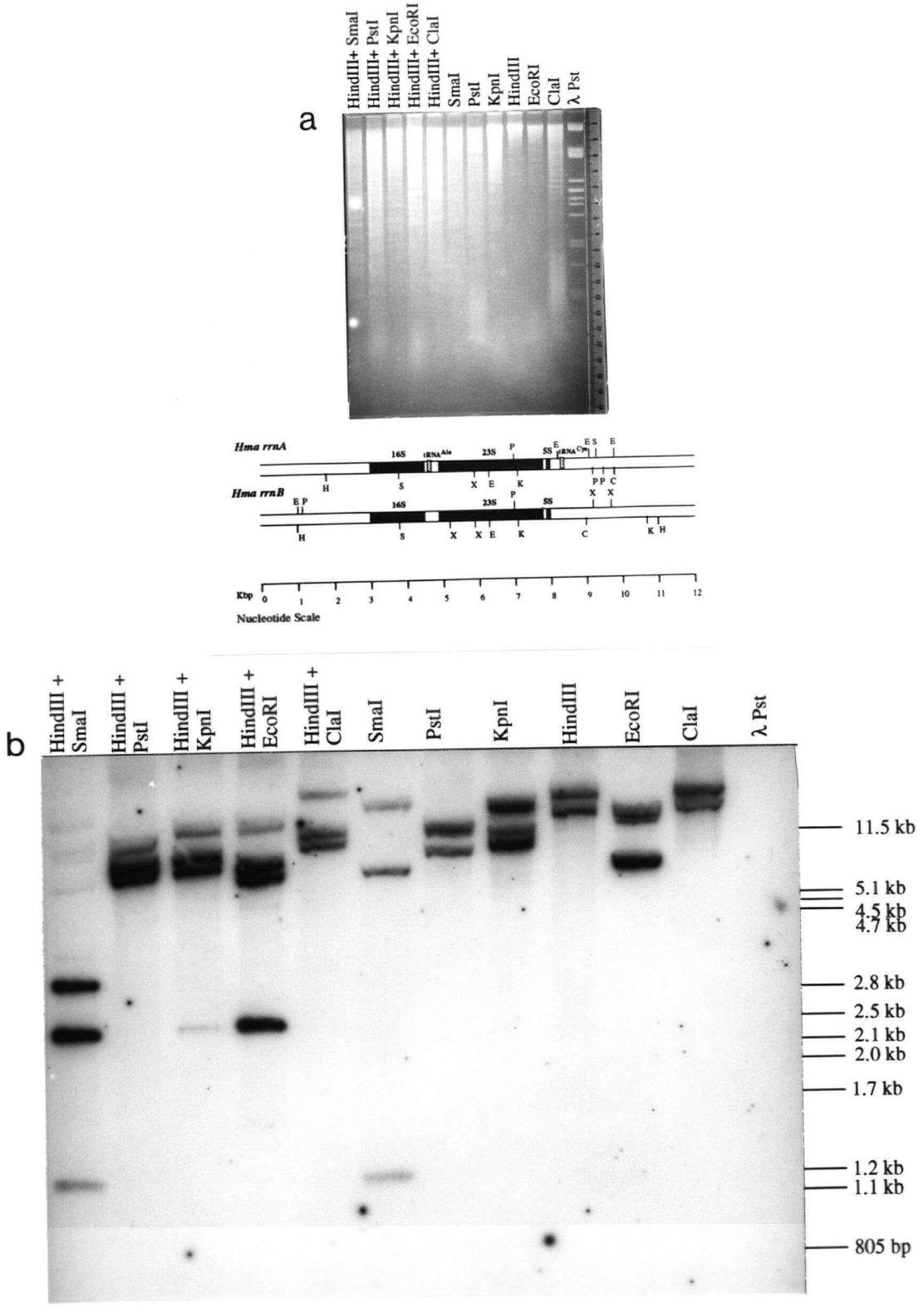

Figure **3.2** Figure caption on next page (page46)

Figure **3.2** Genomic southern hybridization with oligonucleotide, oPD 34.

(**a**) Genomic *Ha. marismortui* DNA isolated from progeny of a single cell was digested singly with the following restriction enzymes: *HindIII*, *SmaI*, *PstI*, *KpnI*, *EcoRI* and *ClaI*. For the double digests, *HindIII* was used along with one of *SmaI*, *KpnI*, *PstI*, *EcoRI* or *ClaI*. The digested DNAs were fractionated in a 1% agarose gel along with λ DNA digested with *PstI* as a size standard. An oligonucleotide oPD 34 was used as a probe in this analysis (see section **2.2.4**). The line diagrams indicated below are the restriction maps of the cloned *rrnA* and *rrnB* operons showing the digestion sites of *HindIII* (H), *EcoRI* (E), *SmaI* (S), *PstI* (P), *KpnI* (K), *XhoI* (X), and *ClaI* (C) (a similar restriction map was obtained by Mevarech *et al.*, 1989). (**b**) The autoradiogram showing the hybridization pattern of the DNA fragments from different digestion products of the genomic DNA. The lengths of fragments were estimated using the λ *PstI* fragments as standards.

sequencing analysis of the 5'-flanking region of the two operons were also performed and compared with the sequences obtained by Mevarech *et al.*, 1989. These results confirmed that the two clones are HC8 (*rrnA*) and HH10 (*rrnB*).

To confirm the presence of *rrnA* and *rrnB* operons in the genome of *Ha. marismortui*, a Southern hybridization analysis was performed (Figure **3.2**). Chromosomal DNA was isolated from progeny of a single cell and digested singly with the following restriction enzymes: *HindIII*, *SmaI*, *KpnI*, *PstI*, *EcoRI*, and *ClaI*. For the double digests, *HindIII* was used along with one of *SmaI*, *KpnI*, *PstI*, *EcoRI* or *ClaI*. An oligonucleotide, oPD 34, which anneals to conserved nucleotides at position 57-38 of the 16S rRNA genes, (sequence obtained from Mevarech *et al.*, 1989), was used as a probe in this analysis. The digested fragments were transferred on to a Hybond N membrane, hybridized overnight to the oligonucleotide probe at 37°C, and then washed with a low stringency buffer at room temperature.

The hybridization pattern shown in Figure **3.2** indicates that in all digestions, there are always two strong bands which correspond to the *rrnA* or *rrnB* operons (this is based on the restriction maps of the two operons). In some lanes, there is a third band with varying intensities and in some lanes, there are either two or four bands. Digestion with *Hind III–Sma I*, *Hind III–Cla I*, *Hind III– Pst I*, *Sma I* and *Kpn I* gave rise to three bands, suggesting that

there are three rRNA operons present in *Ha. marismortui* genomic DNA. In lane 1, the probe hybridized to three *HindIII-SmaI* digestion products that correspond to a 2.8-kbp fragment from the *rrnB* operon, a 2.1-kbp fragment from the *rrnA* operon, and a 1.1-kbp fragment presumably from a third operon, *rrnC*. The 1.1-kbp band also appeared in lane 6, where the DNA was digested with *SmaI* alone. This clearly indicates that the 1.1-kbp fragment is bounded by *SmaI* sites at each end. In lane 2, the probe hybridized to three *HindIII-Pst*I digestion products corresponding to a 6.0-kbp fragment from the *rrnA* operon, a 6.6-kbp fragment from the *rrnB* operon, and a 7.7-kbp fragment presumably from the *rrnC* operon. In lane 3, the probe hybridized to three *HindIII-KpnI* digestion products corresponding to a 5.8-kbp fragment from the *rrnA* operon, a 6.2-kbp fragment from the *rrnB* operon and an 8-kbp fragment presumably from the *rrnC* operon. In lane 4, the probe hybridized to four *HindIII-EcoRI* digestion products, corresponding to a 5.7-kbp fragment from the *rrnB* operon, a 5.3-kbp fragment from the *rrnA* operon, and a 2.1-kbp fragment presumably from the *rrnC* operon and a band >11.5-kbp, may due to partial digestion. The 2.1-kbp fragment is a product of *HindIII* and *EcoRI* digestion because the band was not apparent in the single digestions (lanes 8 and 9). In lane 5, the probe hybridized to three *HindIII-ClaI* fragments, corresponding to a 7.75-kbp fragment from the *rrnB* operon, an 8.0-kbp fragment from the *rrnA* operon, and a third fragment which is >15-kbp in length. Digestion with *ClaI* alone indicates that this >15-kbp fragment may be a *ClaI-ClaI* digestion product (lane 11). Single digestions with *SmaI* and *KpnI* shown in lanes 6 and 8 respectively (Figure **3.2b**), yielded three bands. In lanes 7 and 9, where the DNA was digested with *PstI* or *HindIII*, only two bands are apparent, it is presumed that a third band is not visible because two of the three fragments were not resolved.

These results suggesting that the presence of a third band in some digestions may be due to several reasons. First, a third operon, *rrnC* (see section **3.2.6**), is present in the genome of *Ha. marismortui*. This observation is in agreement with the previous report by Sanz *et al* (1988); they performed a Southern hybridization on restriction fragments separated by pulse field gel electrophoresis and observed three fragments that hybridized to ribosomal probes specific for 16S and 23S rRNAs (Table **3.1**). Second, a recombination product of the *rrnA*

and *rrnB* operons, present in small proportion of the population, may represent the third band. The low intensity bands observed in some lanes (e.g. in *Hind III–Sma I* and *Sma I* digestions) may be an indication for this assumption. Third, there may be some non-specific binding between the probe and a complementary sequence in some part of the genome, not containing any rRNA operon. The lanes showing four bands may be due to partial digestion of the DNA or non-specific binding of the probe. The lanes showing only two bands may indicate that there are only two operons present in the genome or that two of the hybridizing fragments comigrate. To confirm these assumptions, other probes have to be used and the third operon has to be isolated from the genomic DNA of *Ha. marismortui*, and characterized.

Mevarech *et al.*, (1989) have reported that there are only two operons present in *Ha. marismortui* genome. The experimental procedure employed by us and Mevarech *et al.* (1989), was similar (the genomic DNA was digested with *Hind III–Cla I*). They misinterpreted the third band as a partial digestion product. The basis for their interpretation is attributed to the reduced intensity and coincident mobility of a third fragment produced by HindIII digestion. The third operon in this thesis has been designated *rrnC* and is assumed to contain the 23S and 5S gene sequences determined by Brombach *et al.* (1989). Confirmation of this assumption awaits the cloning and sequence characterization of the genomic *rrnC* locus.

3.2.2 Operon Structures

3.2.2.1 Primary Structure

To study the degree of heterogeneity between the *rrnA* and *rrnB* operons, a complete sequence analysis was performed. Several strategies were used in generating the 6171-bp sequence from the *rrnA* operon (Figure **3.3A**) and the 5947-bp sequence from the *rrnB* operon (Figure **3.3B**). These strategies included the use of rRNA sequence specific primers, and the use of M13, T7 and SP6 primers to sequence small fragments that had been

Table **3.1** Summary of pulse field gel electrophoresis experiments using chromosomal DNA from different halophilic archaea, digested with different restriction enzymes and hybridized with 23S and 16S rRNA probes (adapted from the data of Sanz *et al.*, (1988).

Microorganisms	Enzymes Used				No.of operons
	Not I	*Dra I*	*Sfi I*	*BamHI*	
Ha. Californiae ATCC 33799		+	+	+	4
Hf gibbonsii ATCC 33959	+	+	+		4
Hb halobium NCMB 777	+	+	+		3
Ha. marismortui	+	+	+		3
Hc. morrhuae NCMB 757	+	+	+	+	2
Hb. salinarium CCM 2148	+	+	+	+	1

sub-cloned into pGEM and M13 vectors. Exonuclease III generated deletions were also made and sequenced using T7 and SP6 primers. About 95% of the sequences were performed by sequencing double stranded or single stranded DNAs from both directions and 5% on one direction only (within the 5'-flanking region of the 16S rRNA genes, 23S-5S spacer and the 5S distal region). However, when there were uncertainties, the opposite single strand was sequenced to confirm that region. Special attention was paid to any site exhibiting polymorphism between operons.

The figures **3.3 A** and **3.3 B** show the complete nucleotide sequences of the *rrnA* and *rrnB* operons respectively. The gene orders and the unique polymorphic restriction sites are also depicted. As in other halophilic archaea, the 16S, 23S and 5S rRNA genes are linked in

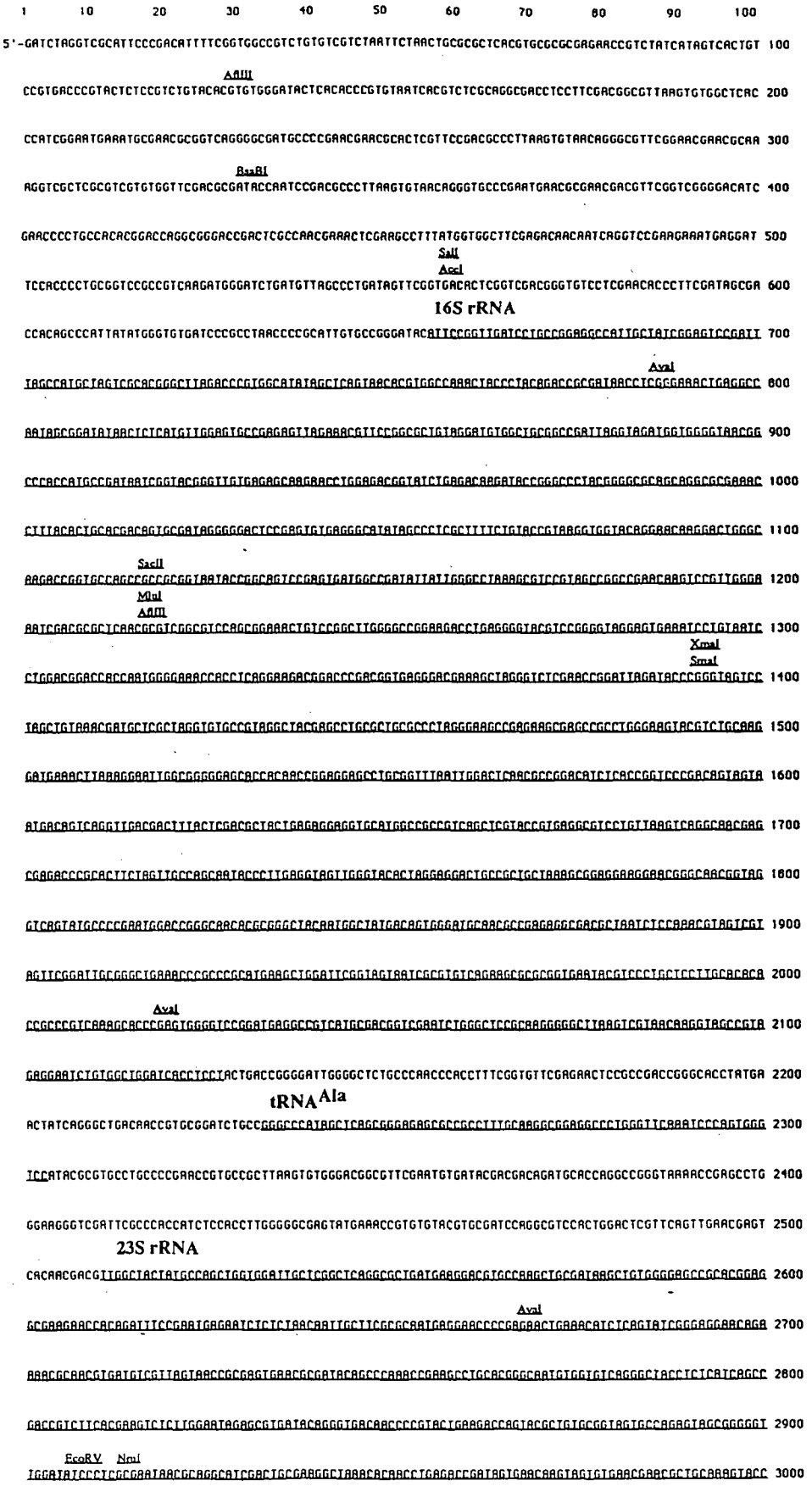

Figure **3.3A** Figure caption on page 54

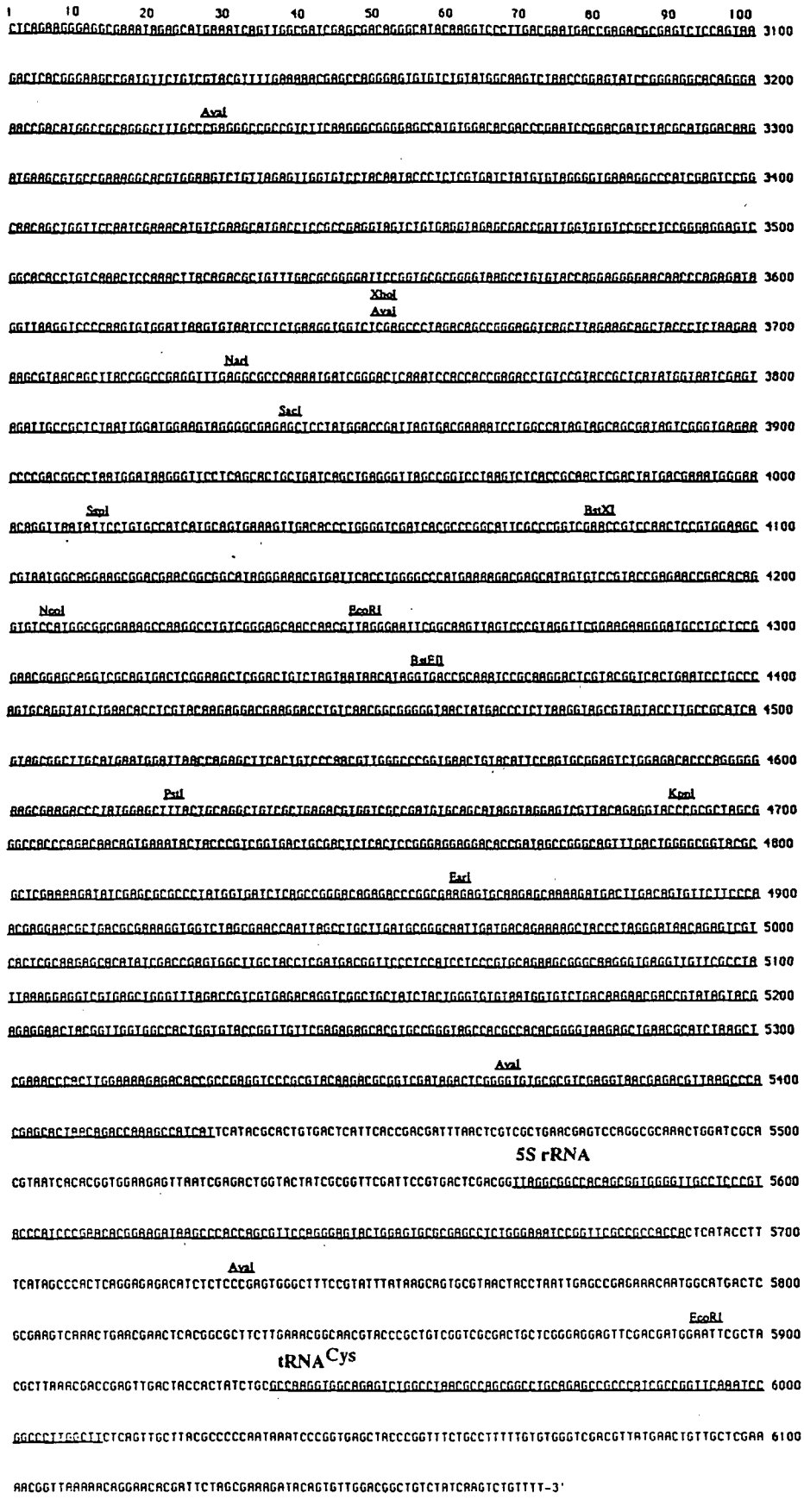

Figure 3.3A continued

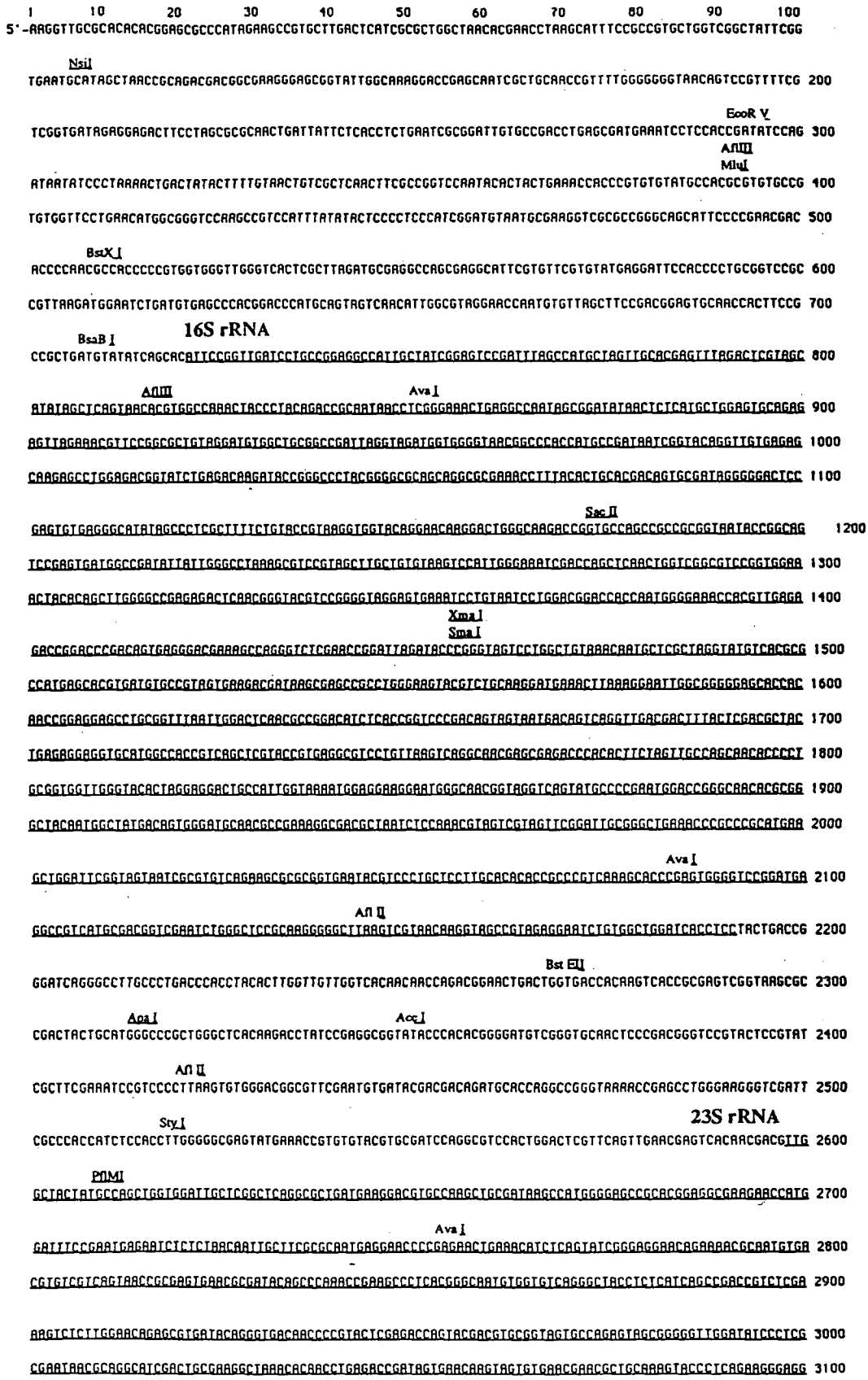

Figure 3.3B Figure caption on page 54

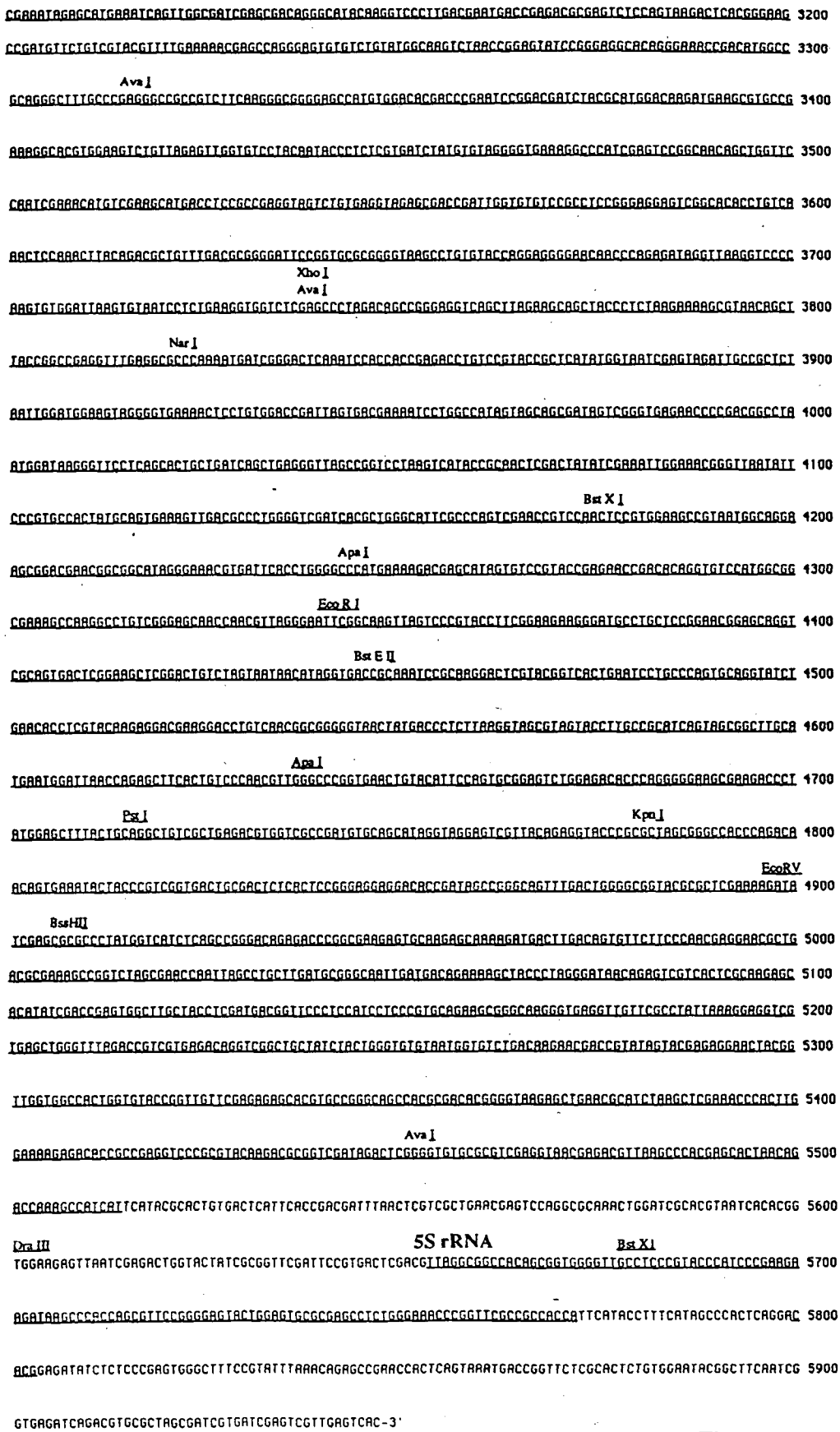

Figure 3.3B continued

Figure **3.3A** The complete sequence of the *rrnA* operon from *Ha. marismortui* is given from 5'- to 3'- direction. The 6171 nucleotide long sequence (+ strand) comprises the following: 16S leader, 16S rRNA, spacer containing $tRNA^{Ala}$ sequence, 23S rRNA, 23S-5S spacer, 5S rRNA, 5S distal region containing a $tRNA^{Cys}$. The rRNA and tRNA gene sequences are underlined and some important restriction enzyme sites are indicated above the respective sequences.

Figure **3.3B** The complete sequence of the *rrnB* operon from *Ha. marismortui* is given from 5'- to 3'- direction. The 5947 long nucleotide sequence (+ strand) comprises the following: 16S 5'flanking region, 16S rRNA, 16S-23S spacer, 23S rRNA, 23S-5S spacer, 5S rRNA and 5S distal region. The rRNA gene sequences are underlined and some important restriction enzyme sites are indicated above the respective sequences

both operons. Sequence analysis indicates that the *rrnA* operon contains a $tRNA^{Ala}$ gene within the 16S-23S intergenic spacer and a $tRNA^{Cys}$ gene located distal to the 5S gene. The *rrnB* operon possesses neither the $tRNA^{Ala}$ within the intergenic spacer nor the $tRNA^{Cys}$ in the 388-bp region distal to the 5S gene.

A Southern hybridization analysis was performed to test for the presence of the $tRNA^{Cys}$ gene in the *rrnB* operon (Figure **3.4a**). An oligonucleotide, oPD 45, specific for the $tRNA^{Cys}$ sequence in the *rrnA* operon was used as a probe in this analysis. As a control, a 1.0-kbp *EcoRI-EcoRI* fragment from the *rrnA* operon containing the $tRNA^{Cys}$ gene was used. The probe was hybridized to the DNAs at 42°C, washed first with a low stringency buffer at room temperature and then with medium stringency buffer at 42°C. The autoradiogram showed that the probe hybridized only to the 1.0-kpb fragment from the *rrnA* operon (PD 1099) and not to the 1.8-kbp or 2.0-kbp fragments from the *rrnB* operon (PD 1022). This experiment clearly demonstrates that the *rrnB* operon does not contain a $tRNA^{Cys}$ gene sequence in the 5S distal region (Figure **3.4b**).

Sequencing and Southern hybridization analysis indicate that the gene orders of the *rrnA* and *rrnB* operons from the *Ha. marismortui* are 5'-16S-$tRNA^{Ala}$-23S-5S-$tRNA^{Cys}$-3' and 5'-16S-23S-5S-3', respectively (Figure **3.5**). The sequence from the *rrnC* operon (Brombach *et*

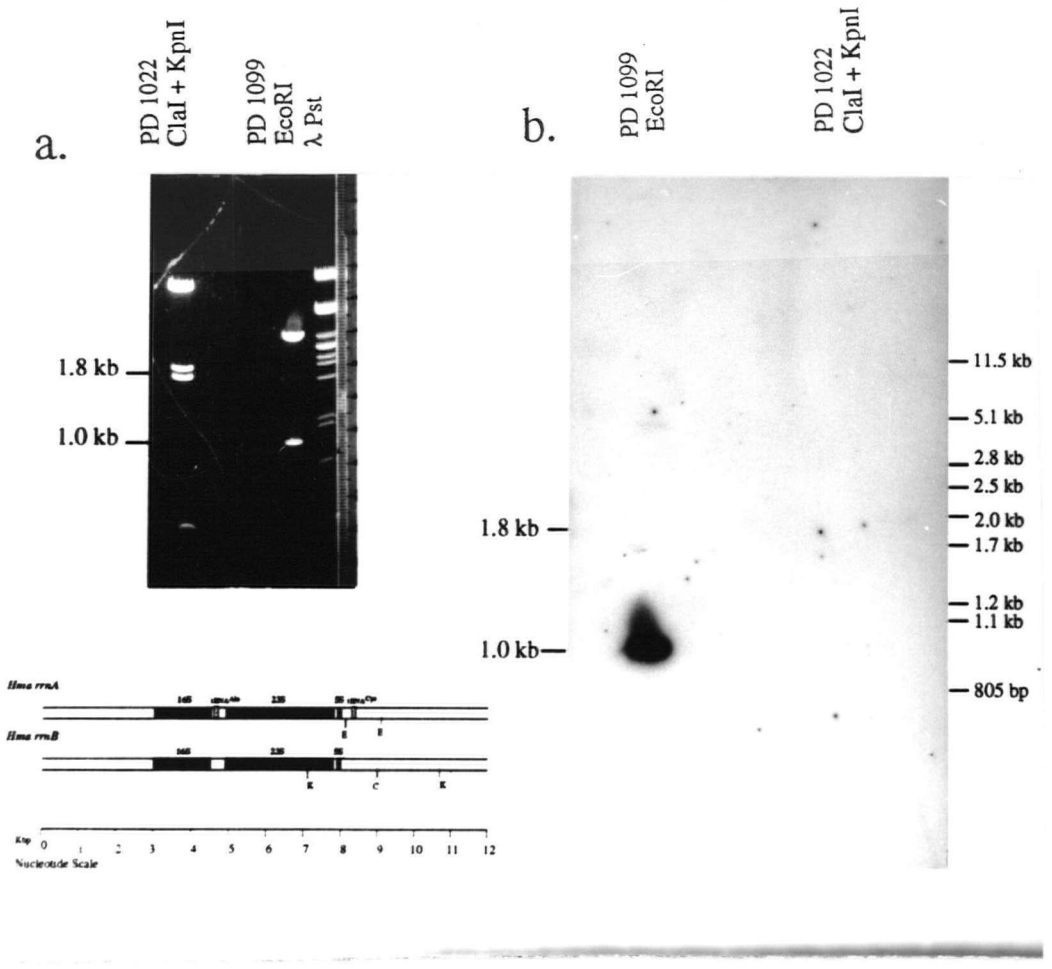

Figure **3.4** Identification of a $tRNA^{Cys}$ gene using Southern hybridization analysis. Plasmids containing the 5S distal regions of the *rrnA* (PD 1099) and *rrnB* (PD 1022) were digested with restriction enzymes *EcoRI* and *ClaI-KpnI* respectively and probed with the γ-^{32}P ATP-labelled oligonucleotide oPD 45, which is specific for the $tRNA^{Cys}$ gene (see section **2.2.4**).

(**a**) Gel shows the restriction enzyme digested fragments from plasmids PD 1099 and PD 1022. The fragments generated by digestion of λ DNA with *PstI* were used as a size standard for this analysis. The positions of the restriction enzyme sites indicated (*Cla I*, *Kpn I* and *EcoRI*) are identical to the restriction maps in Figures, **3.1** and **3.2**)

(**b**). Autoradiogram showing the 1.0 Kb fragment hybridized to the probe.

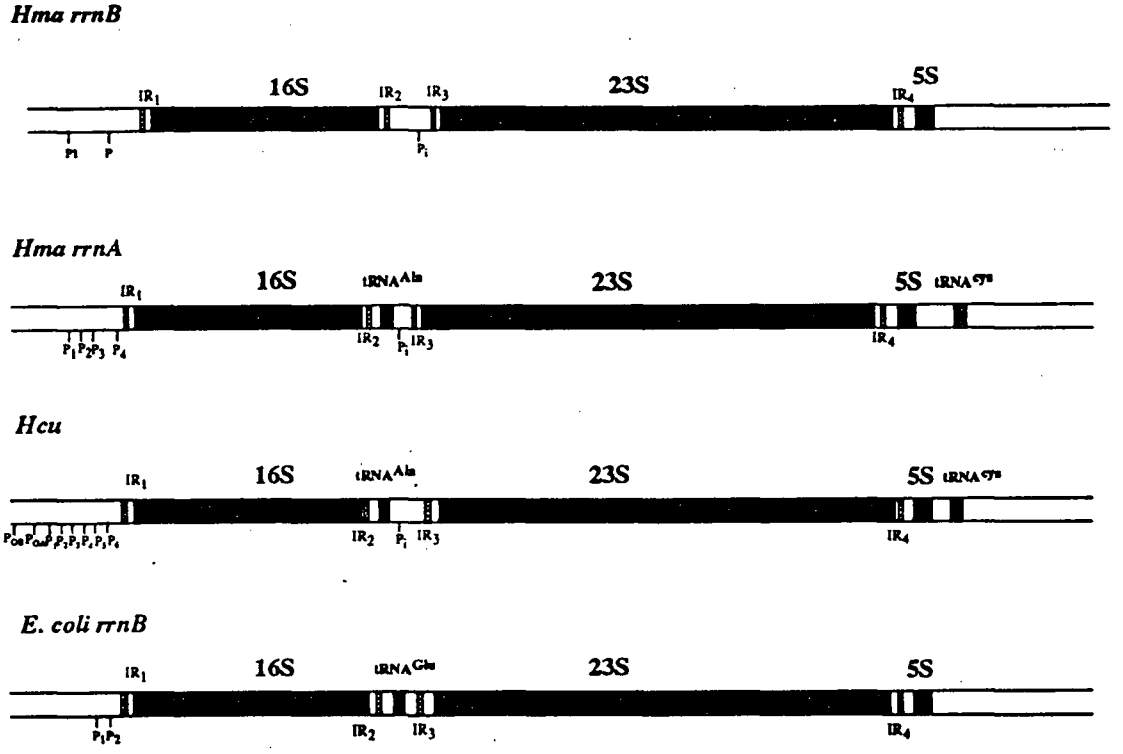

Figure **3.5** The gene organization of the *rrnA* and *rrnB* operons from *Ha. marismortui*. The gene organization of the *rrnA* and *rrnB* operons was compared with that of a related halophile, *Hb. cutirubrum* (Hui and Dennis, 1985) and a eubacterium, *E. coli* (Brosius *et al.*, 1981). The rRNA genes (16S, 23S and 5S) and the tRNA genes ($tRNA^{Ala}$ and $tRNA^{Cys}$) are shown as solid boxes and the 5' flanking regions, intergenic spacers and the distal regions are blank. The small boxes within the 5' flanking regions and spacers indicated as IR are the inverted repeat sequences surrounding the 16S and 23S rRNA genes. These sequences are usually involved in the processing of the respective rRNAs. The promoters are indicated as "P"s at the 5' flanking regions of all the operons. Internal promoters are indicated as Pi, within the 16S-23S rRNA spacer regions.

al., 1989) contains a portion of the *rrnB*-like 16S-23S intergenic spacer (254 nucleotides upstream of the 23S rRNA start site) and a 463 nucleotide long *rrnA*-like 23S distal region (including the 5S rRNA gene). The 5S distal sequence is insufficient to determine whether the *rrnC* operon contains a $tRNA^{Cys}$ gene or not. The gene order of the *rrnB* operon of *Ha. marismortui* is unusual because it deviates from the general pattern, 5'-16S-$tRNA^{Ala}$-23S-5S-$tRNA^{Cys}$-3', found in halophilic archaea, such as the *rrnA* operon of *Ha. marismortui* and the

single rRNA operons from *Halobacterium cutirubrum* (Hui and Dennis, 1985, Figure **3.5**), *Halococcus morrhuae* (Larsen *et al.*, 1986) and one of the two rRNA operons from *Haloferax volcanii* (Gupta *et al.*, 1983). The second operon from the *Hf. volcanii* contains a $tRNA^{Ala}$ gene within the 16S-23S intergenic spacer region but lacks the second, $tRNA^{Cys}$ gene, located downstream of the 5S gene (Daniels *et al.*, 1986).

To display the spectrum of the phylogenetic differences between halophiles and to identify functionally important elements in their molecular structures, the entire sequences from the *rrnA*, *rrnB* and the single operon from *Hb. cutirubrum* were aligned pairwise and the nucleotide differences within ten nucleotide intervals were recorded. *Hb. cutirubrum* was selected for the comparative analyses because it is a related organism and the sequence of the entire operon is available. Gaps (or deletions) were introduced to maintain alignment and were not considered as substitutions. Comparison of two DNA sequences usually cannot tell us whether a deletion had occurred in one sequence or an insertion had occurred in the other. Therefore, the outcomes of both types of events are collectively referred to as gaps. The distribution of substitutions along the length of the sequences of the three pairwise comparisons are illustrated by the histograms A, B, and C in Figure **3.6**. The two sequences which were compared are indicated on the upper right hand corner of Figure **3.6**.

As a first approximation, regions containing more than five substitutions per increment were considered to be poorly conserved. In the *rrnA-rrnB* case, most of the substitutions were observed within the 5'-flanking, 3'- flanking and spacer regions, although a substantial number was observed within the rRNA gene sequences. Differences observed within the 16S rRNA genes (positions 180 to 1652 in Figure **3.6**) sequences are less uniformly distributed than the surrounding regions. With a single exception at position 1216, they are confined to three intervals, bounded by nucleotides at positions 236-501, 688-1003 and 1166-1338 within the 16S rRNA genes.

The 23S rRNA genes (position 2140 to 5050 in Figure **3.6**) show fewer substitutions than the 16S rRNA genes and the substitutions are confined in four intervals, bounded by

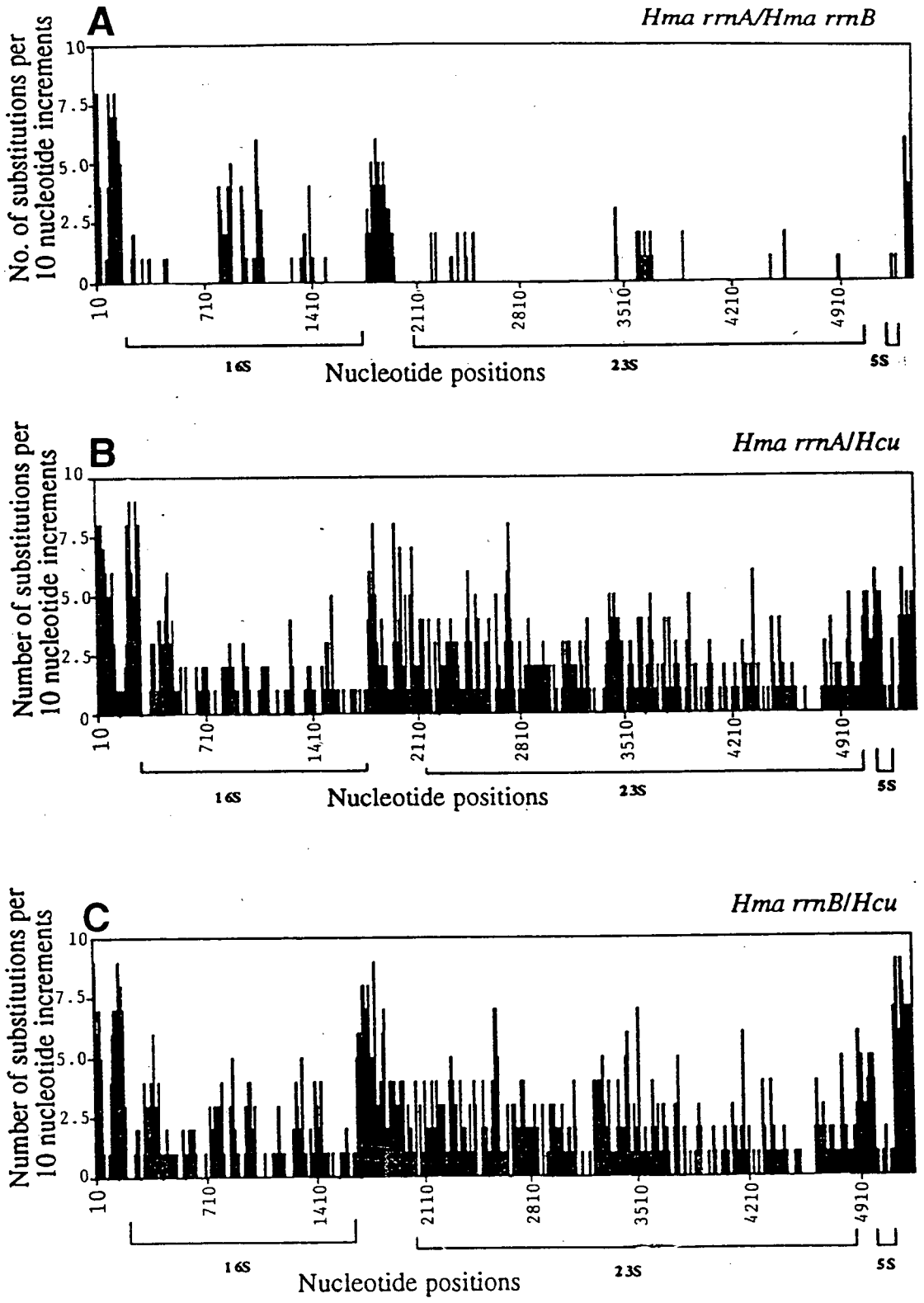

Figure 3.6 Figure caption on next page (page 59).

Figure **3.6** Distribution of sequence differences of the pairwise comparisons along the length of the three rRNA operons. Pairwise sequence alignments were made between the fellowing sets of operons; *rrnA* and *rrnB* of *Ha. marismortui*, *rrnA* of *Ha. marismortui* and *Hb. cutirubrum*, *rrnB* of *Ha. marismortui* and *Hb. cutirubrum*. Gaps were introduced in the alignment whenever needed. In each comparison, the nucleotide differences between the two sequences within ten nucleotide intervals were recorded and plotted against nucleotide positions. The two rRNA operon sequences in each pair are indicated on the right. The location of the 16S, 23S and 5S rRNA sequences are indicated at the bottom. Arrows indicate the peptidyl transferase center region. Abbreviations are as follows: *Hma rrnA* and *Hma rrnB* indicate *rrnA* and *rrnB* operons of the *Ha. marismortui* respectively; and *Hcu* indicates the single operon of *Hb. cutirubrum*. The bounds of the 16S, 23S and 5S rRNA sequences are indicated below the X axis.

nucleotides at positions 70-365, 1320-1765, 2320-2412 and 2759-2769 of the 23S rRNA genes. The 5S rRNA genes show only two nucleotide substitutions. In the 3'-flanking region, the *rrnA* and *rrnB* operon sequences were identical for approximately 60 nucleotides downstream of the 5S rRNA genes; after that point they are totally divergent. The biological and phylogenetic significance of the substitutions within the 5'-flanking regions, the rRNA genes, the spacers and the distal regions will be analysed in the remaining part of chapter 3 and chapter 4.

Unlike the *rrnA-rrnB* case, the substitutions within the *rrnA-Hb. cutirubrum* and *rrnB-Hb. cutirubrum* comparisons are distributed throughout the entire operon. However, in the latter cases, the substitution rates are relatively higher within the regions surrounding the rRNA genes. Within the 16S-23S spacer regions, the *rrnA* and the operon from *Hb. cutirubrum* show fewer substitutions than what was observed between *rrnB* and the operon from *Hb. cutirubrum*, which is attributed to the presence of nearly identical $tRNA^{Ala}$ gene sequences in the first pair of operons. Considering the histograms A, B and C shown in Figure **3.6**, the region between 4600 and 4800 in the first two and the one between 4460 and 4530 in histogram C show very few substitutions. This highly conserved region corresponds to the peptidyl transferase center within domain V of the 23S rRNA.

3.2.2.2 Secondary Structure

Putative secondary structures of the primary transcripts from the *rrnA*, *rrnB* and *Hb. cutirubrum* were deduced (Figures **3.7A**, **3.7B** and **3.7C**, respectively) from their primary sequences using the structural model of Kjems and Garrett (1987). Comparison of these structures reveals that the 16S and 23S rRNAs of the *rrnA* and *Hb. cutirubrum* operons and the 23S rRNA of the *rrnB* operon are surrounded by long processing stems each containing the expected "bulge-helix-bulge" motif. The inverted repeat structure surrounding the 16S rRNA of the *rrnB* operon is different from other processing stems in that it does not contain the motif. The spacer ($tRNA^{Ala}$) and the distal ($tRNA^{Cys}$) tRNAs are present in the *rrnA* and *Hb. cutirubrum* operons and are absent in the *rrnB* operon.

A number of other putative helices, designated A to I, are also present within the primary transcripts (see Figures **3.7 A, 3.7 B,** and **3.7 C**). Helix A is irregular and can be formed in all archaeal transcripts except for those of *D. mobilis* and *M. vanielii*. Although there is no formal phylogenetic support or conserved base pairings among all archaea, the lower part of the helix was aligned for three methanogens and this yielded positive compensatory base changes for 4 base pairs (Kjems and Garrett, 1990). In the transcripts of the three halophilic operons (the *rrnA* operon of *Ha. marismortui* (this work) and the single operons of *Hb. cutirubrum* and *Hc. morrhuae* (Kjems and Garrett, 1990; Garrett *et al.*, 1993) the helix A structures are conserved in that they contain the recognition elements for the most proximal promoter (Figures **3.7A** and **3.7C**). The AT-rich Box A promoter motifs lie in the terminal loops and the corresponding transcript start sites are located in the unstructured sequence between the helices. These structures can potentially form only in RNA transcripts initiated at a promoter located upstream of the repeat. Any transcripts initiated from this most proximal promoter would lack the conserved helix A at the 5'- end. In each case, extending upstream from helix A, somewhat similar inverted repeats overlap other upstream promoter sequences (Figure **3.8**). These repeats vary in their primary sequences and predicted secondary structure in ways such as the length of the repeats and the number of nucleotides between the two halves of the repeat. The *rrnB* operon

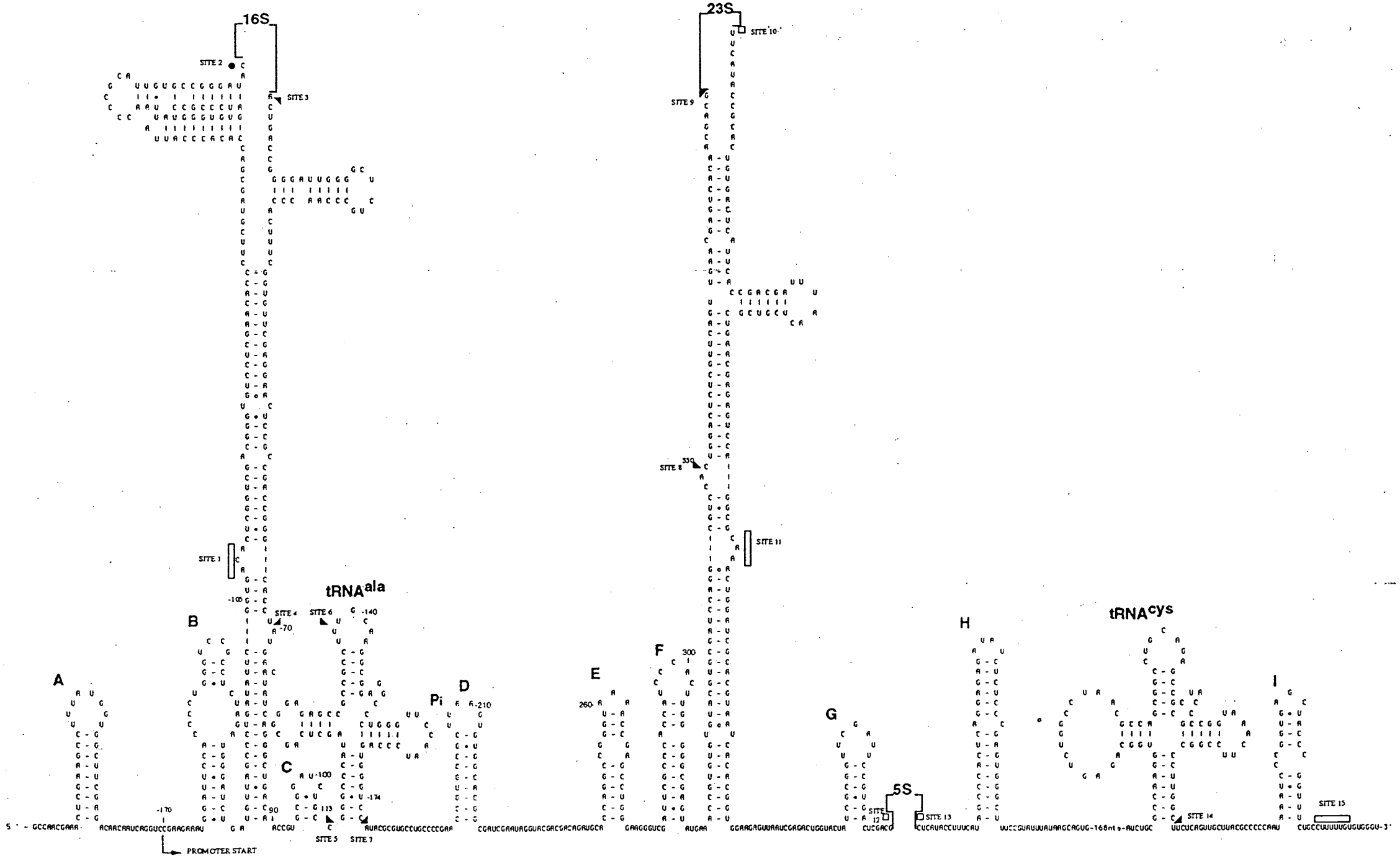

Figure 3.7A Primary transcript of the *rrnA* operon of *Ha. marismortui*.

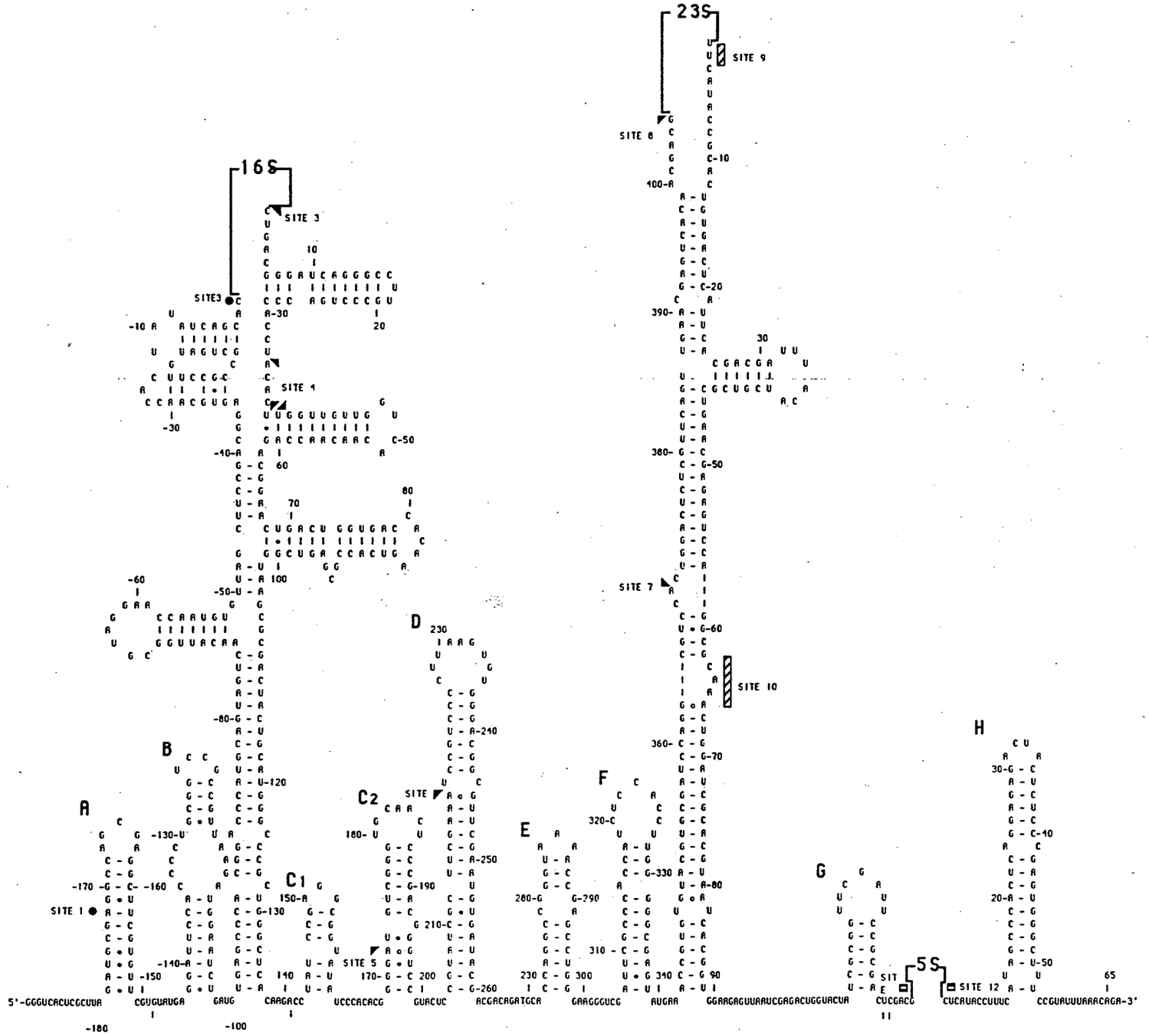

Figure 3.7B Primary transcript of the *rrnB* operon of *Ha. marismortui*.

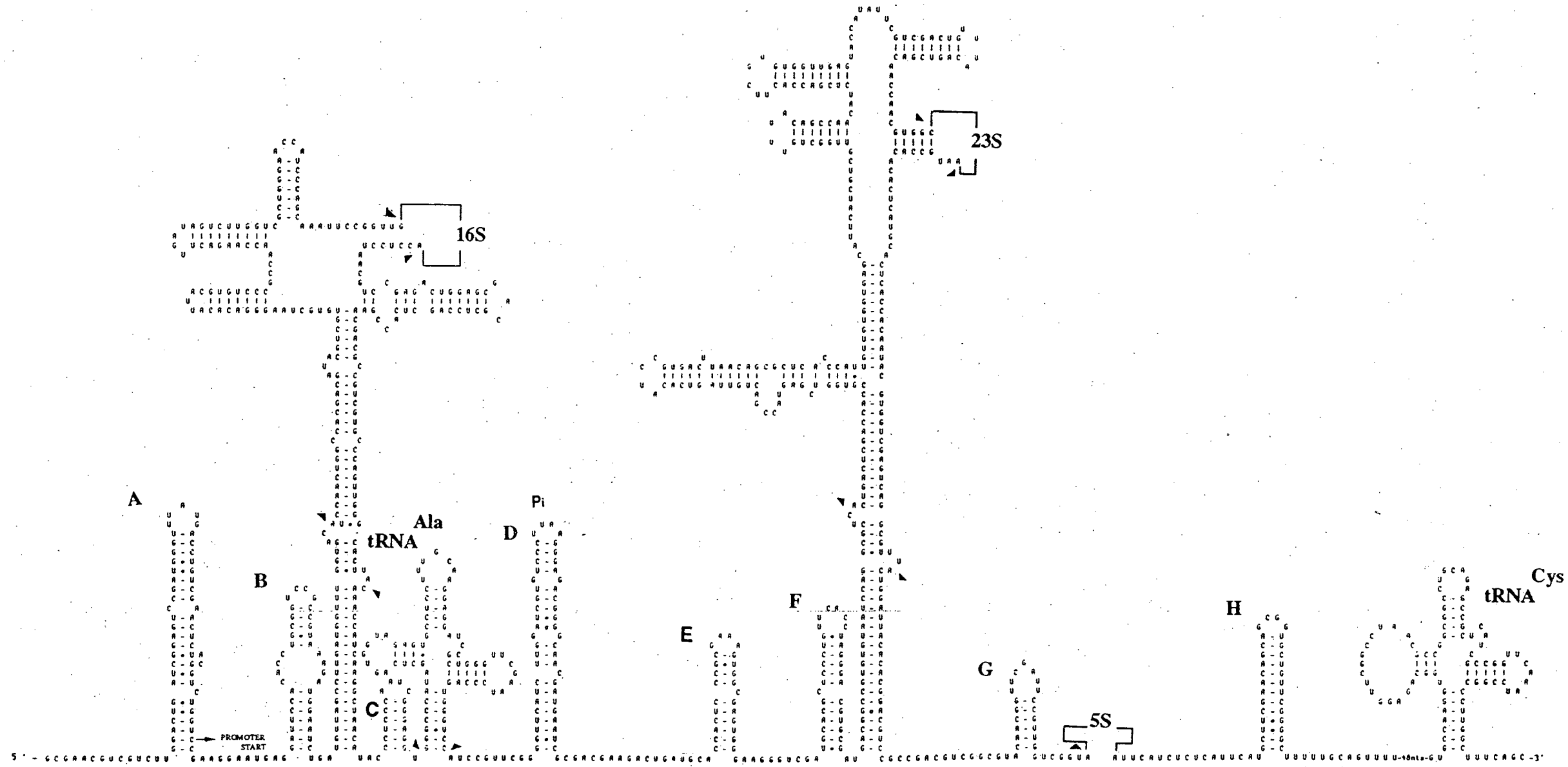

Figure 3.7C Primary transcript of *Hb. cutirubrum* rRNA operon.

Figure **3.7** Putative secondary structures from the *rrnA*, *rrnB* and *Hb. cutirubrum* operon transcripts.

Figure **3.7A** The putative secondary structure of the primary transcript for the *rrnA* operon from *Ha. marismortui*. The two tRNAs, $tRNA^{Ala}$ and $tRNA^{Cys}$, and a number of putative helices, A to I, are present within the primary transcript and are named according to the nomenclature of Kjems and Garrett (1990). The processing and maturase sites of the 5'-flanking regions that were mapped by nuclease S1 protection assays and confirmed by using primer extension analysis (are indicated by dots (•). The arrows indicate the processing and maturase sites within the 16S-23S spacer and $tRNA^{Cys}$ processing site (only if they were mapped at their nucleotide positions by sequencing). Other sites mapped by S1 nuclease protection assays to approximate nucleotide positions are shown by boxes. The processing and maturase sites are indicated from sites 1 to 15. The promoter P4 start site is indicated at position -170 in the 5'-flanking region. The box A sequence (UUAA) of the internal promoter, Pi, is shown in the loop of the helix D.

Figure **3.7B** The putative secondary structure of the primary transcript of the *rrnB* operon from *Ha. marismortui*. A number of putative helices A to H are shown. The processing and maturase sites mapped at the nucleotide positions are indicated by arrows within the 16S-23S spacer region and by dots (•) in the 5'-flanking regions. The approximate sites identified by nuclease S1 protection assays are indicated by boxes. Sites indicated from 1 to 12 are involved in the processing and maturation of the 16S, 23S and 5S rRNAs. The box A sequence of the putative promoter Pi is shown in the loop of the helix D. Helices C1 and C2 replace helix C and the $tRNA^{Ala}$ from the *rrnA* operon.

Figure **3.7C** The putative secondary structure of the primary transcript of the *Hb. cutirubrum* (sequence obtained from Hui and Dennis, 1985). The processing and maturation sites mapped by nuclease S1 protection assays (Chant and Dennis, 1986) are shown by arrows. The box A sequence of the internal promoter sequence (TTAA) is located within the loop of helix D.

transcript appears to be unique, in that it contains a 16S processing site within helix A and the known *rrnB* promoter sequences are all located upstream of the helix A region.

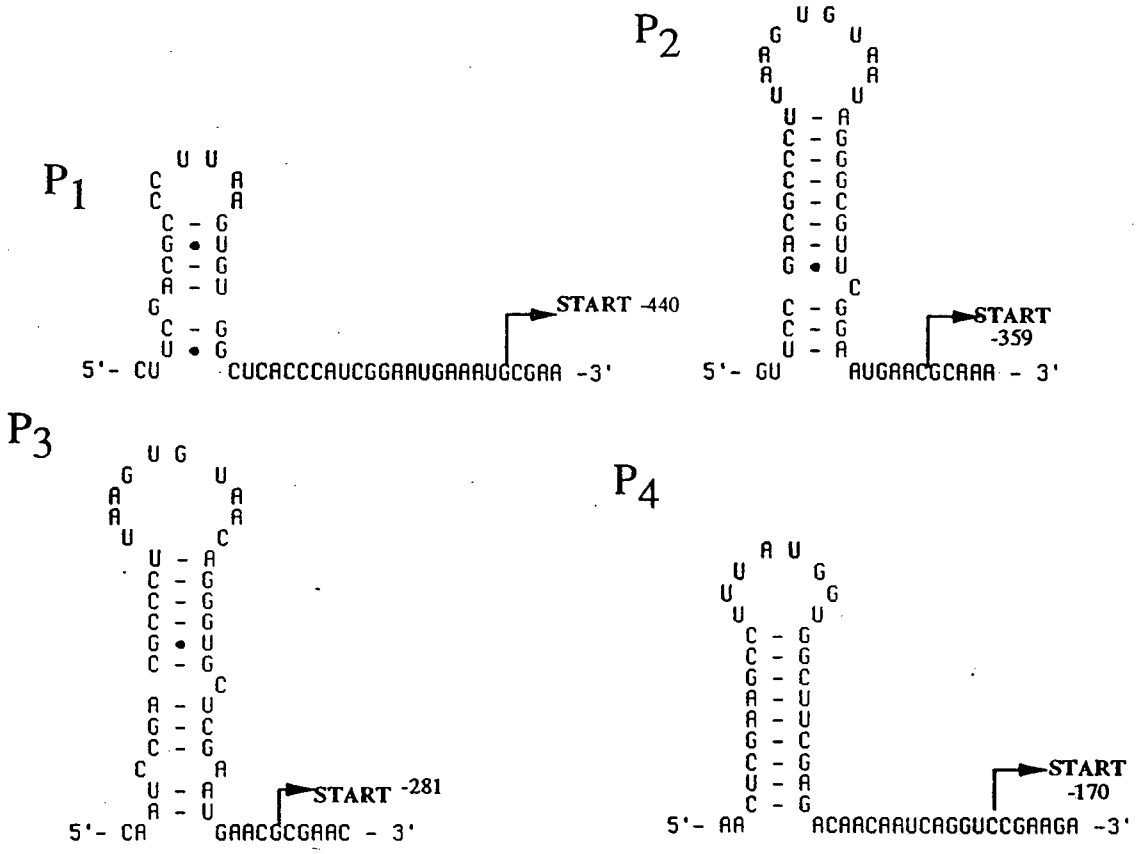

Figure 3.8 A series of inverted repeat structures of the 16S 5'-flanking region of the *rrnA* operon containing the promoter motifs. The AT-rich Box A motifs invariably lie in the terminal loop regions whereas the putative transcript start site is located beyond the regions of inverted repeat symmetry. These nucleotide positions of the sequences are indicated. The P4 structure corresponds to helix A in Figure 3.7A.

In the case of the *rrnA* operon transcript, the P4 promoter sequence (see Figures **3.10** and **3.11**) is located within helix A; the other three promoter motifs are within regions of inverted repeat symmetry (Figure **3.8**). In the four helices depicting these repeats, the TTAT promoter elements are located within the terminal loop.

Helix B is a universal structure in archaea and is clearly evident in the *rrnA* and *rrnB* transcripts. This structure is defined by its position and by the specific sequences at the base of the helix (Figure **3.9**). The appropriate sequences from some archaeal organisms are aligned in Figure **3.9**, where it can be seen that several base pairs in each helix are supported by

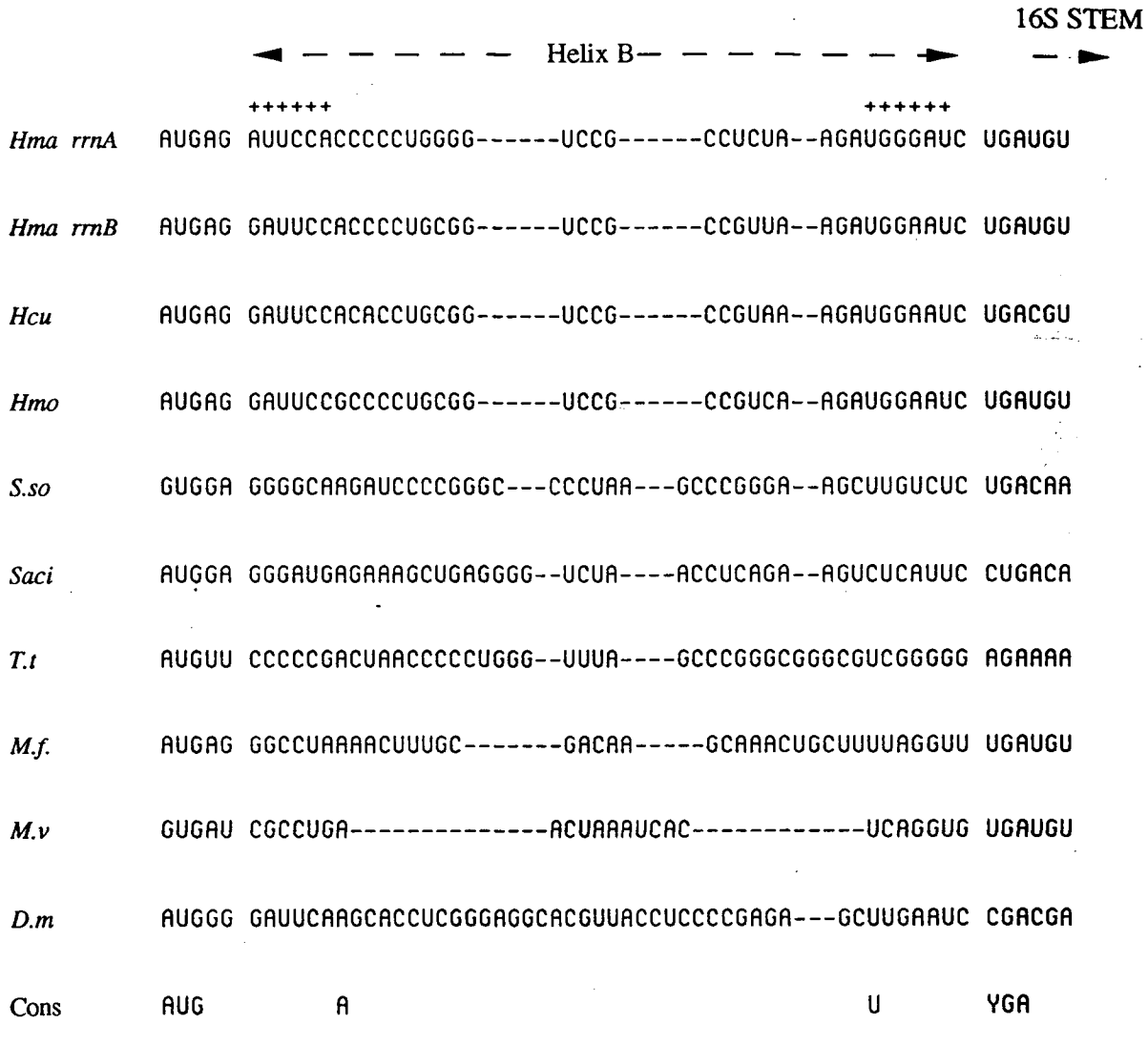

Figure **3.9** Alignment of the partial sequences that generate a helix B in the 16S leader regions of some archaeal organisms. Complementary sequences which generate the putative helices are indicated by arrows over the sequences. A + indicates that the putative complementary nucleotides in the helix are supported by compensatory base changes. The approximate start point of the 16S RNA processing stem is indicated by an arrow at the right. Consensus sequences common to at least 80% of the organisms are given on the last line. Abbreviations used are: *Hma rrnA* and *Hma rrnB* refer to the *Ha. marismortui rrnA* and *rrnB* operons; *Hcu*, *Halobacterium cutirubrum*; *Hmo*, *Haloccus morrhuae*; *Sso*, *Sulfolobus solfataricus*; *Saci*, *Sulfolobus acidocaldarius*; *Tf*, *Thermofilum pendens;* and *Mva*, *Methanococcus vanielli.*

compensatory base changes. An internal loop, observed usually in helix B structures of halophilic and thermophilic archaea, is present in the *rrnA* and *rrnB* operon transcripts (Figure **3.7A, 3.7B, 3.7C**). The sequences surrounding the helix B structures of *rrnA* and *rrnB* , 5'-AUGGA-helix-UGA-3', are identical in all the halophilic sequences known. These structures may constitute maturation signals, transcription termination-antitermination signals or have some other functions (Mason *et al.*, 1992; Nodwell and Greenblatt, 1993).

In halophiles, helix C is always present 5'-to the spacer tRNA (Figures **3.7 A** and **3.7 C**), suggesting that it may be related either to 5'-tRNA processing by RNase P or to 3'-end processing of 16S rRNA. The *rrnB* operon which lacks the $tRNA^{Ala}$ contains two helices designated C1 and C2 at this position.

Helix D is only present in halophilic archaeal organisms, which contain an intergenic, putative promoter sequence. Within this structure, a TTAA motif (Box A) is located in the apical loop region. The number of base pairs involved in the formation of the helix structure varies from species to species. In the *rrnA* and *rrnB* operons from *Ha. marismortui*, the sequences between the TTAA motifs of the Pi and the 5'-ends of 23S rRNAs are identical. However, the helix D structures are different in that the sequences upstream to the TTAA motifs are different. It was shown that these intergenic promoters found in the *rrnA* and *rrnB* operons are active (discussed under section **3.2.5.3**). In the thermophilic archaeon, *S. acidocaldarius*, an internal promoter-like sequence was observed within the 16S-23S spacer, however, activity was not detected (Durovic and Dennis, 1994).

Helix E, formed from a pair of perfect inverted repeat structures which are bordered by sequences AUGCA and GAAG, is conserved in all halophiles. Helix F, defined as the helix which directly precedes the 23S precursor processing stem, is formed from a pair of long, imperfect, inverted repeats. The sequence conservation and the secondary structural conservation of E and F, along with their primary sequence conservation within the flanking sequences have been observed in many species (Leffers *et al.*, 1987). Helix G is always observed in the 5'-flanking region of 5S rRNA transcript and is assumed to be involved in the

processing of 5S rRNA (Kjems and Garrett, 1987). Helix H, found downstream of the 5S rRNA transcript of the *rrnA*, *rrnB* and *Hb. cutirubrum* operons, is followed by poly dT stretches. These structures are similar to the *rho*-independent termination sites observed in *E. coli.* In the *rrnA* operon transcript, another helix structure, helix I, is located downstream to the $tRNA^{Cys}$, followed by poly dT stretch. Nuclease S1 protection assays indicate that termination occurs at this site (section **3.2.7.4**; Chant and Dennis, 1986). Therefore, it is suggested that the transcript reads through the poly dT stretch associated with helix H and the $tRNA^{Cys}$ and terminates at the poly dT stretch associated with helix I. In *Hb. cutirubrum*, the sequence beyond the $tRNA^{Cys}$ is not sufficient to determine whether the transcript can form a structure like helix I (Hui and Dennis, 1985).

3.2.3 The 16S Leader Region

3.2.3.1 Sequence Alignment

The 5'-flanking regions of the *rrnA* (664 nucleotides) and *rrnB* (720 nucleotides) operons from *Ha. marismortui* were determined. The two sequences were aligned and compared with the *Hb. cutirubrum* sequence (714 nucleotides) as described in section **2.2.17** (Figure **3.10**). The alignment shows that the *rrnB* operon differs from the *rrnA* and *Hb. cutirubrum* operons in several respects. The *rrnA* and *Hb. cutirubrum* operons contain multiple promoters which are active (this work; Mevarech *et al.*, 1989; Chant and Dennis, 1986; Mankin and Kagramanova, 1986); however, only one major promoter activity (P) was observed in the case of *rrnB* operon so far (activity of the Px promoter was not investigated). The 75 to 80 nucleotides preceding and including the processing sites are about 90% identical in the case of the *rrnA* and *Hb. cutirubrum* operons. The *rrnB* operon contains the preceding sequence but lacks the processing site sequence (GTGACA sequence located within the "bulge-helix-bulge" motif of the secondary structure; Mevarech *et al.*, 1989; Dennis, 1991). In the *rrnB* operon, the sequence associated with a putative precursor processing site CGAGGCC is located at position -172 within the helix A of the universal secondary structure. A related sequence CGAAGCC is

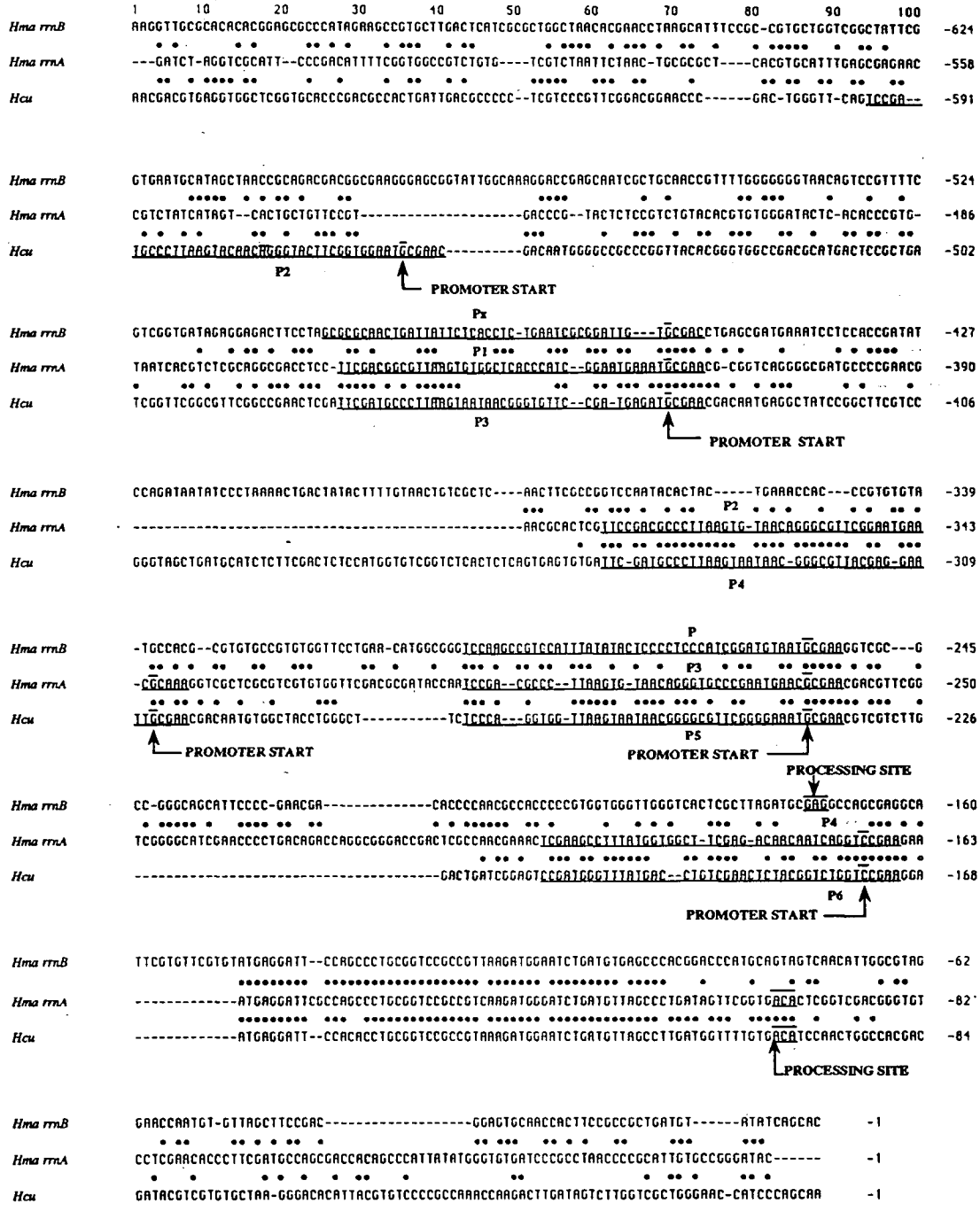

Figure **3.10** Alignment of the 5' flanking sequences from the *Ha. marismortui* operons, *rrnA* and *rrnB*, and from the *Hb. cutirubrum*. The methods used in the alignment are described in section **2.2.17**. The nucleotide identities shared between two sequences (*rrnA-rrnB*, top row and *rrnA-Hcu*, bottom row) are indicated by dots (•). The AT-rich putative promoter motifs (Ps), promoter start sites, and the processing sites are under and/or over-lined (see Figures **3.7A**, **3.7B** and **3.12**). The nucleotide positions at the end of each row is indicated on the right and the gaps are not included as nucleotides.

present in the *rrnA* operon transcript at the corresponding position within the helix A but no cleavage has been detected at this site.

To analyze the substitution rates within the 5'-flanking regions of the *rrnA*, *rrnB* and *Hb. cutirubrum* operons, pairwise comparisons were made using the available sequences from each operon. When entire sequences were compared between the *rrnA-rrnB* 16S 5'-flanking regions (597 nucleotides), they show 49.2% identity; between position -1 and the P_1 promoter sequence of the *rrnA* operon (position -484), they show 53.2% identity. In the *rrnA-Hb. cutirubrum* comparison, the entire sequences from both operons show 55.8% identity; between positions -1 and the P_1 promoter sequence of the *rrnA* operon (position -484), they show 59.6% identity. In the *rrnB-Hb. cutirubrum* comparison, the 599 nucleotides show 51.4% identity; between positions -1 and the Px sequence of the *rrnB* operon (position -500), they show 53.6% identity. These results indicate that the 5'-flanking region of the *rrnA* operon is more similar to the *Hb. cutirubrum* operon than to the *rrnB* operon; between positions -1 and -484 of the *rrnA* operon, *rrnA* and *Hb. cutirubrum* are almost 60% identical. Sequence comparisons also indicate that the *rrnB* sequence is more closely related to the *Hb. cutirubrum* sequence than the *rrnA* sequence.

3.2.3.2 Promoters

The *rrnA* operon of *Ha. marismortui* contains four tandem promoter sequences, P_1, P_2, P_3, and P_4, located upstream of the 16S rRNA gene (see section **3.2.3.4**, and Mevarech *et al.*, 1989). In the *rrnB* operon a single promoter, P and a promoter-like sequence Px were identified (see section **3.2.3.4**, and Mevarech *et al.*, 1989). All six promoter sequences from the *rrnA* and *rrnB* operons were aligned with promoters from the single rRNA operons of *Hb. cutirubrum* and *Hc. morrhuae.* (Figure **3.11**). Three regions with high nucleotide sequence similarity are apparent. The first is centered around the transcription initiation site; the majority of the transcripts initiates at a G residue and when the start nucleotide is not G, the promoter is generally low in efficiency (Dennis, 1985; Chant and Dennis, 1986; Mankin and Kagramanova, 1986). In *Ha. marismortui,* the weak P_4 promoter from the *rrnA* operon and the putative Px

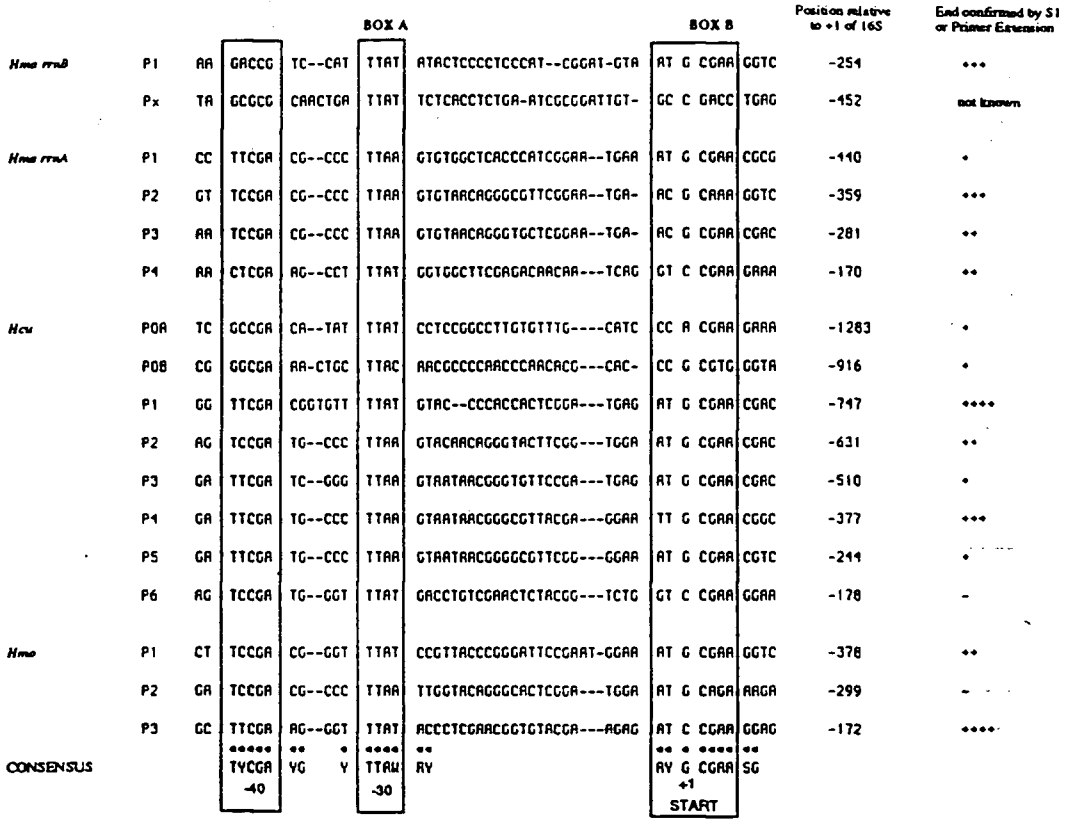

Figure **3.11** The alignment of promoters and putative promoter-like sequences from halophilic rRNA operons. Seventeen rRNA promoter sequences from halobacterial species are summarized. *Ha. marismortui* (*Hma*) contains two operons, designated *rrnA* and *rrnB* containing four and two tandem promoters respectively. *Hb. cutirubrum* (*Hcu*) has a single operon with eight tandem promoters. *Hc. morrhuae* (*Hmo*) has a single operon with three apparent tandem promoters. The start sites within the highly conserved Box B and the other conserved motifs, designated Box A (at position -30) and the halophilic-specific TYCGAYG motif (at position -40), are aligned. Consensus nucleotides within the rRNA promoters are also presented: Y, pyrimidine; R, purine; W, A or T; S, G or C. The relative strengths of the promoters based on the intensity of the bands from the S1 nuclease analysis (see section **3.2.3.4**) are indicated by +, ++, +++ and ++++.

promoter from *rrnB* starts with a C residue; all the others start with G. The second conserved element is a well defined AT-rich box (Box A) 25-38 nucleotides upstream from the first

transcribed nucleotide. This feature may be the equivalent of the eukaryotic TATA box, located about 25 nucleotides upstream of the transcription start site in RNA polymerase II promoters (Reiter *et al.*, 1988). The distance between the two consensus promoter elements (Boxes A and B) is approximately the same in archaea and eukaryotes (Reiter *et al.*, 1990). The third element is a TYCGAYG sequence at the -40 position from the start site which appears to be unique to halophiles (Mankin and Kagramanova, 1986).

Pairwise alignments of the promoter sequences (80 nucleotides in length) from the *rrnA* and *rrnB* operons were performed and the percentage similarities from each comparison are summarized in Table **3.2**. The percentage similarities given in Table **3.2** show that the promoters P_2 and P_3 from the *rrnA* operon are more closely related than the rest of the pairs. Among the four *rrnA* promoters, the weak promoter P_4 is the least similar. The P and Px sequences from the *rrnB* operon are more similar to each other than they are to the four promoters from the *rrnA* operon.

A pairwise comparison of the *rrnA* and *rrnB* promoter sequences with those of the rRNA promoters of *Hb.cutirubrum* and *Hc. morrhuae* revealed that they can be grouped into three subsets (data not shown). The P_1, P_2 and P_3 sequences from the *rrnA* operon are related to the P_2, P_3, P_4 and P_5 sequences from *Hb. cutirubrum*. The P_4 of *rrnA*, P_{OA} and P_6 of *Hb. cutirubrum*, and promoters P_1, P_2 and P_3 of *Hc. morrhuae* are related and can be grouped as a subset. The P and Px sequences from the *rrnB* are related to the P_{OB} and P_1 promoters of *Hb. cutirubrum.*

It is envisaged that the presence of multiple promoters in the halophiles described above provides the flexibility to maintain transcription under different salt conditions, ranging from 1.5M – 5M NaCl (Christian and Waltho, 1962; Ginzburg *et al.*, 1990; Lanyi and Silverman, 1972). Perhaps the presence of template specific multiple transcription factors, i.e., the DNA-binding proteins, which recognize short nucleotide sequences on different promoters, enable efficient transcription. An alternate possibility is that at different salt concentrations, the DNA-

Table **3.2** The percentage similarities between the promoter sequences from the *rrnA* and *rrnB* operons of *Ha. marismortui.* Pairwise alignments of the promoter sequences consisting of 80 nucleotides were compared and the percentage similarities between each pair were recorded. Promoters from the *rrnA* operon are indicated as *rrnA* P1, *rrnA* P2, *rrnA* P3 and *rrnA* P4, and from *rrnB* as *rrnB* P and *rrnB* Px.

	rrnA P_2	*rrnA* P_3	*rrnA* P_4	*rrnB* P	*rrnB* P_x
rrnA P_1	66	66	45	54	58
rrnA P_2		73	56	38	45
rrnA P_3			57	50	51
rrnA P_4				52	55
rrnB P					64

dependent RNA polymerases present in *Ha. marismortui, Hb. cutirubrum* and *Hc. morrhuae* respond differently in their DNA-binding ability.

Louis and Fitt showed that the salt response of the DNA-dependent RNA polymerase of *Hb. cutirubrum* was dependent on the type of DNA template supplied (linear, circular and supercoiled DNAs; Louis and Fitt, 1971a; Louis and Fitt, 1971b; Louis and Fitt, 1972a; Louis and Fitt 1972b). Louis and Fitt indicated that the DNA-dependent RNA polymerase of *Hb. cutirubrum* is a dimer containing single alpha and beta subunits (Louis and Fitt, 1971b; Louis and Fitt, 1972a; Louis and Fitt 1972b). However, Zillig *et al* (1985) later showed that *Hb. cutirubrum* RNA polymerase contains eight components, A, B', B", C, D, H, I and J (Zillig *et al.*, 1985). Louis and Fitt had shown that the beta subunit of DNA-dependent RNA polymerase is responsible for chain initiation and that in a primed reaction, catalyzed by the alpha subunit, the template-dependent salt effects were no longer observed (Louis and Fitt 1972b). These findings suggest that the template specificity of the RNA polymerase of *Hb. cutirubrum* might

reside in a site which recognizes short DNA sequences. On the basis of the above discussion, an explanation for the presence of different promoter sequences would be that the beta subunit carries two or more initiation sites with different salt dependent properties. Furthermore, since there is only one active site in the beta subunit, the effect of salt concentrations could change the conformation of the protein, thereby resulting in the recognition of different nucleotide sequences.

With the exception of the *rrnB* operon of *Ha. marismortui*, all other halophilic promoter elements contain an inverted repeat sequence surrounding the Box A element (see section **3.2.2.2**). The inverted repeat sequences containing all four promoter motifs from the *rrnA* operon are shown in Figure **3.8**. In all four cases, the TTAT promoter element is present in the region between the repeats. There is little apparent conservation in either the sequence or lengths of the repeats or the distance separating the repeat sequences. The functional role of these structures was not investigated.

Inverted repeat sequences within promoters were also observed in eukaryotic and eubacterial organisms (Vogeli *et al.*, 1981; Orosz and Adhya, 1984; Majumdar and Adhya, 1984; Dunn *et al.*, 1984). Larsen and Weintraub (1982) have suggested that, in the chick α_2 collagen gene promoter, the inverted repeat sequences might provide a recognition site for trans-regulatory proteins; and that, once bound, these proteins would inhibit nucleosome formation in that region of DNA. Inverted repeat sequences were also observed in *E. coli* promoters of the *gal* (Orosz and Adhya, 1984; Majumdar and Adhya, 1984) and *ara* (Dunn *et al.*, 1984) genes. The significance of the presence of these inverted repeat sequences was not studied in detail.

3.2.3.3 Primary Transcript Analysis of the 16S Leader Region

Protection from S1 nuclease digestion of end labelled DNA fragments by precursor and product rRNA transcripts was used to identify the positions of transcription initiation sites, processing sites, and mature 16S maturation sites within the 5'-flanking region of the two rRNA operons from *Ha. marismortui*. To investigate the 5'-end sites of the transcripts within the 5'-flanking region, restriction fragments were obtained from plasmids PD 927 (*rrnB* operon) and

PD 928 (*rrnA* operon) digested with *AflIII* and used as probes. The enzyme *AflIII* cleaves at position 95 within the 16S gene of both operons, at position -544 in the 5'-flanking region of *rrnA*, and at position -331 in the 5'-flanking region of *rrnB*. The 638 nucleotide long *AflIII-AflIII* fragment from *rrnA* (Figure **3.12A**) and similar 425 nucleotide long fragment from *rrnB* (Figure **3.12 B**) were 5'- labelled on the minus strand at position 95 within the 16S gene, hybridized to *Ha. marismortui* total RNA, and digested with S1 nuclease. Protected DNA fragments were analyzed on an 8% polyacrylamide gel.

With the two probes, the shortest protected fragment had a size of about 95 nucleotide corresponding to protection by the mature 5'- ends of the 16S rRNAs—the *AflIII* site is conserved within the 16S rRNA genes of both operons. Multiple bands observed in this region may be due to imprecise maturation of the 5'- end of 16S rRNA or imprecise trimming of the protected DNA by S1 nuclease.

With the *rrnA* probe, additional major protected fragments of approximately 191, 265, 376, 454, and 535 nucleotides were observed. The end of the 191 nucleotide fragment corresponds to the precursor processing site at position -96 and the ends of the other four fragments correspond to the transcription initiation sites that mapped by S1 nuclease analysis at or near the G residues at positions -440, -359, -281, and -170.

With the negative strand DNA 5'-labelled at the *AflIII* site on the 425 bp fragment from the *rrnB* operon, a full length product and partially protected fragments of 95, 267, 349, and 425 nucleotides were observed. These partial products are believed to represent protection by (i) the mature 5'- end of the 16S rRNA; (ii) the precursor processing site (at position -172); (iii) the promoter start site (at position -254); and (iv) the full length product corresponding to the transcript initiated upstream of the *AflIII* site within the 5'-flanking region, respectively (Figure **3.12B**). The minor band X, may correspond to another processing site located upstream of the promoter P (since there is no putative promoter sequences are present within this region, this band may not be a promoter start site). The band Y corresponds to a contaminant DNA fragment that is present within the conserved region of the 16S genes which ran closer to the 425 nucleotide probe during isolation. Further purification of the probe was not needed since

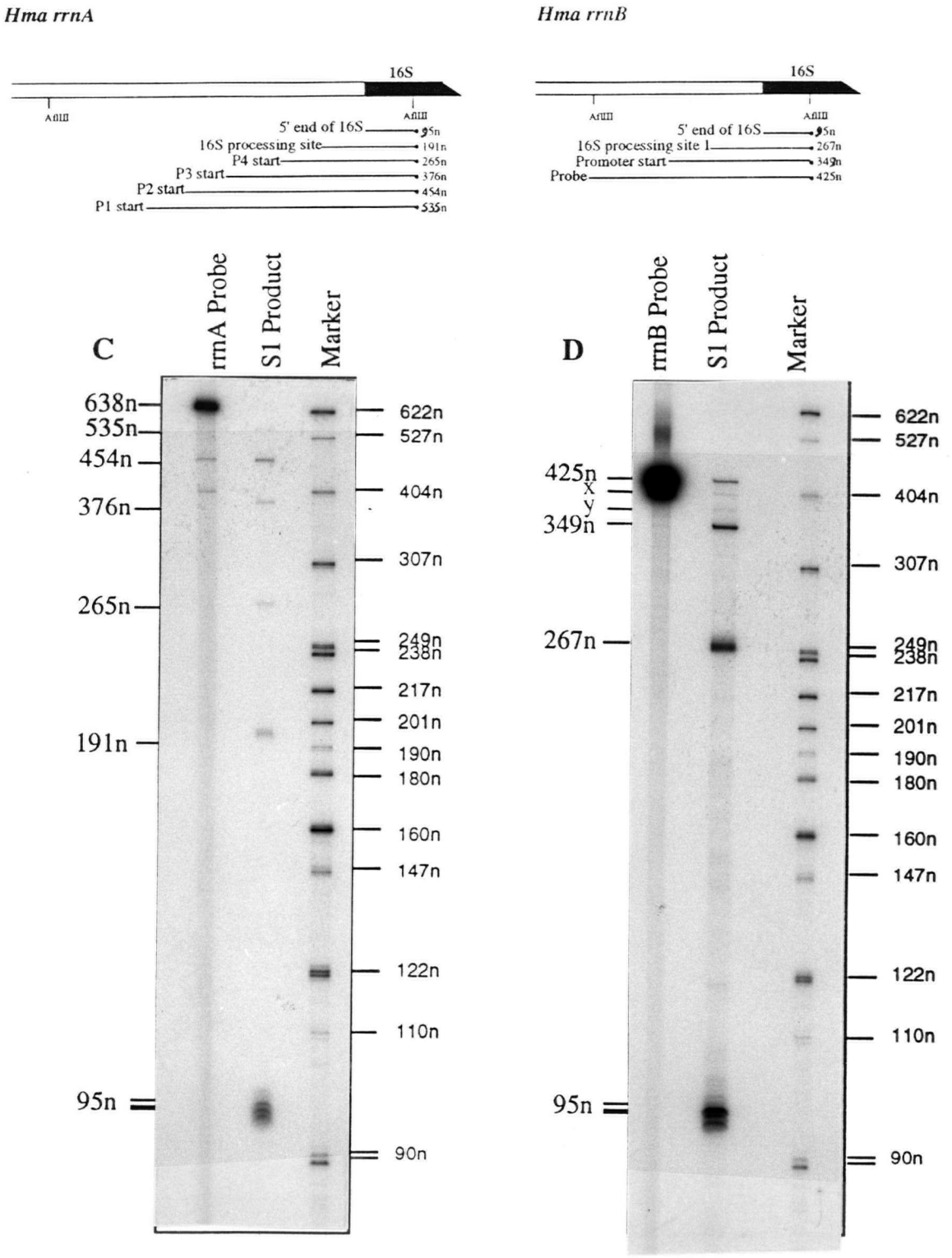

Figure **3.12** Figure caption on next page (page 77).

Figure **3.12** Nuclease S1 protection assays using probes from the 5'-flanking regions of the *rrnA* and *rrnB* operons of *Ha. marismortui.*

Figure **3.12A** A line diagram showing the 16S 5'-flanking regions of the *rrnA* operon (from plasmid PD 928). The two *AflIII* sites within the 16S rRNA and 5'-flanking region were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The positions where the probes were labelled with γ-^{32}P ATP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.12B** A line diagram showing the 5' flanking region of the 16S rRNA gene of the *rrnB* operon (from plasmid PD 927). The two *AflIII* sites within the 16S rRNA and the 5'-flanking region were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The positions where the probes were labelled with γ-^{32}P ATP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.12C** The autoradiogram showing the nuclease S1 protection assay products from the 5'-flanking region of the *rrnA* operon. Above each lane, the *rrnA* probe, S1 product (S1 nuclease digestion product of the annealed probe with total RNA), and marker (pBR 322 plasmid digested with *MspI* and labelled at the 3'- end with α-^{32}P dCTP) are indicated. The sizes, of the protected fragments from the S1 treated DNA-RNA hybrids and the markers, are indicated on both sides of the autoradiogram.

Figure **3.12D** The autoradiogram showing the nuclease S1 protection assay products from the DNA-RNA hybrid of the *rrnB* operon. Above each lane, the *rrnB* probe, S1 product (S1 nuclease digestion product of the annealed probe with total RNA), and marker (pBR 322 plasmid digested with *MspI* and labelled at the 3'-end with α-^{32}P dCTP) are indicated. The sizes, of the protected fragments from the S1 treated DNA-RNA hybrids and the markers, are indicated on both sides of the autoradiogram.

this fragment is identical in the *rrnA* and *rrnB* operons, and would not affect the S1 analysis within the 5'-flanking region.

Since S1 nuclease analysis are only semi–quantitative, the band intensities do not indicate the exact amount of RNAs present in the cells. These analysis are mainly done in order to detect the transcription initiation, termination or processing sites of the transcripts. In the case of rRNA transcripts, usually we expect only about 1 - 2% of rRNA in the form of very long precursors and most of the rRNAs in mature form (King and Schlessinger, 1983). However, this is not what we have observed from the S1 nuclease analysis shown in Figure **3.12**; the intensities of the bands representing the mature 16S rRNAs (Figures **3.12 C** and **3.12 D**) are only about 3 or 4 times stronger than the bands representing the processed 16S rRNAs. This reflects non-quantitative aspects in the S1 assays. First, the efficiency of hybridization between rRNA and DNA probe may depend on the lengths of the rRNAs; longer RNAs may hybridize more efficiently than the shorter RNAs and may displace shorter RNAs from the RNA probe. Second, the conditions used in the S1 analysis (especially temperature) may not have permitted the mature rRNA to hybridize efficiently. Third, mature rRNAs may be folded into a secondary structural conformation and may not readily available to hybridize with the DNA probe.

Primer extension analysis were performed in order to confirm the end sites of the prodects identified from S1 nuclease assays (the approximate positions of the end sites were already determined using DNA restriction fragment ladders, as shown in Figures **3.12 A** and **3.12B**). The 267 and 349 nucleotide fragments, observed from the S1 mapping analysis, were confirmed and precisely positioned by primer extension analysis using an *rrnB*-specific oligonucleotide, oPD 48. This oligonucleotide is complementary to a sequence from -96 to -118 within the 5'-flanking region of the *rrnB* operon. The smallest extension product terminates at an A residue at position -172 and corresponds to the major processing site in the 5'-flanking region (Figure **3.13**). The second product terminates at a G residue at position -253 which corresponds to the transcription initiation site from the promoter P (data not shown). Precise mapping of the 95 nucleotide fragments observed from the S1 nuclease analysis (Figures **3.12 C** and **3.12 D**) were also confirmed by primer extension analysis using oPD 34

(complementary to the region 57 - 38 within the 16S genes of the *rrnA* and *rrnB* operons). The major extension products terminated at an A residue at position 1 of the 16S rRNA genes (data not shown).

3.2.3.4 The 16S rRNA Processing Pathways for the *rrnA* and *rrnB* Operons

In the *rrnA* operon, the 16S rRNA processing occurs within the predicted "bulge-helix-bulge" motif (see sections **1.1.7.2** and **3.2.2.2**) of the processing stem. This is similar to the recognition sites used by other archaeal organisms to excise precursor 16S and 23S rRNAs from the primary transcripts and to remove introns from the transcripts of intron containing rRNA and tRNA genes. Sequence comparison of the 5'-flanking sequences (Figure **3.10**) and secondary structural analysis of the primary transcripts (Figure **3.7B**) indicated that the *rrnB* lacks the conserved sequence associated with the cleavage site and the "bulge-helix-bulge" motif within the processing stem. Subsequently, the S1 nuclease protection assay (Figure **3.12**) and primer extension analysis (Figure **3.13**) confirmed that the sequence and the secondary structure associated with the cleavage site in the *rrnB* operon are different. This suggests that in the *rrnB* operon, a different endonuclease is involved in the processing of 16S rRNA from the primary transcript and that the recognition feature for this enzyme is different from that of the endonuclease involved in the processing of the 16S rRNA from the *rrnA* operon. Since the processing site is located within helix A in the universal secondary structure for the *rrnB* transcript (Figure **3.7B**), it is possible that endonuclease recognition site is associated with some aspect of helix A structure. The reason for two distinct 16S processing pathways in *Ha. marismortui* is not understood. It may be that the processing reactions involved are optimally active under different environmental conditions (e.g. different NaCl concentrations, ranging from 1.5 - 5M).

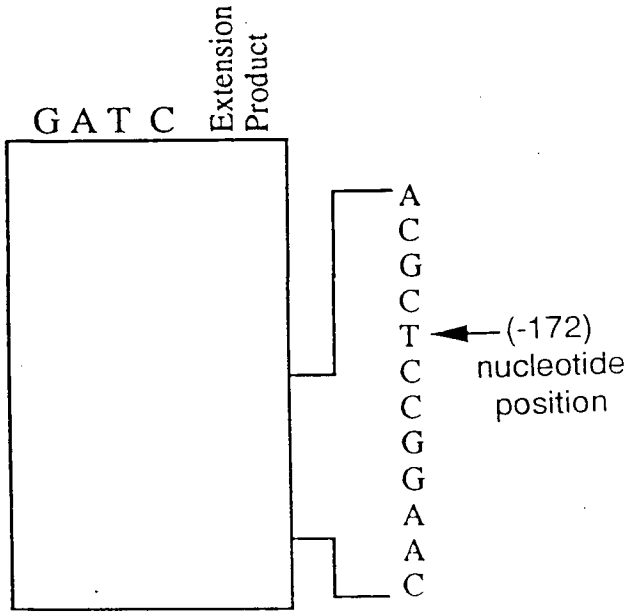

Figure **3.13**. Primer extension analysis within the 5'-flanking region of the *rrnB* operon using oPD 48. The oligonucleotide oPD 48 is complementary to the sequence from -98 to -118 within the 5'-flanking region of the 16S rRNA from the *rrnB* operon. A sequencing ladder using the same primer is shown adjacent to the primer extension product. The arrow is pointing at the nucleotide corresponding to termination of extension product. The poor quality of reaction products from primer extensions is caused by poor hybridization of certain primers to template DNA, and inefficient extension and premature termination caused by high GC content and higher order structure in the template DNA.

3.2.4 16S rRNA

3.2.4.1 Primary Structural Analysis

The complete nucleotide sequences of the *rrnA* and *rrnB* 16S rRNA genes have been determined and are both 1472 nucleotides in length. Comparison of the 16S rRNA gene sequences from the *rrnA* and *rrnB* operons showed that they differ at 74 nucleotide positions (Figure **3.14** and Table **3.3**). To gain insight into the functional and evolutionary significance of these nucleotide differences, the two *Ha. marismortui* 16S rRNA sequences from *rrnA* and *rrnB* were compared to the single 16S rRNA sequences of closely related genera: *Hb. cutirubrum*, *Hf. volcanii* and *Hc. morrhuae* (Gupta *et al*, 1983; Leffers and Garrett, 1984; Hui

and Dennis, 1985) (Figure **3.14**). The 16S sequence of *Hc. morrhuae* is 1475 nucleotides whereas the others are 1472 nucleotides in length. When aligning the sequences the secondary structural model for 16S rRNAs was also considered (Woese and Pace, 1993). To maximize the nucleotide identities in the end to end alignment of the five sequences, introduction of nine gap positions were required.

Results from the pairwise alignments are summarized in Table **3.3**. The 16S rRNA sequences from the *rrnA* and *rrnB* operons of *Ha. marismortui* are 95% identical, whereas the remaining comparisons show substantially less identity, ranging from 86.8% to 90.2%. Intra- and interspecies comparisons show that the substitutions caused by transitions outnumber transversions by about 2:1; in rRNA genes, the significance of this ratio has not been investigated. About 60% of these nucleotide substitutions are compensatory, affecting both components of a nucleotide base pair in a region of an RNA helical structure. All of the substitutions occur between *rrnA* and *rrnB* in what are recognized to be phylogenetically variable positions. These differences are not randomly distributed; with a single exception, all are confined to three domains, 58-321, 508-823, and 986-1158, within the universal secondary structure model for small subunit rRNA (Figures **3.15**, **3.16** and **3.17**, Woese *et al.*, 1983).

In contrast to the *rrnA-rrnB* case, the nucleotide differences in all interspecies pairwise alignments showed that the differences were not concentrated in specific domains but were more generally distributed throughout the 16S sequence (Figure **3.15**). For the three interspecies pairwise comparison between *Hf. volcanii* and *Hb. cutirubrum*, *Hf. volcanii* and *Hc. morrhuae*, and *Hb. cutirubrum* and *Hc. morrhuae*, the 508-823 domain exhibits essentially the same frequency of nucleotide substitutions as the entire 16S molecule (percentage similarities of 88.3%-88% versus. 88.6%-85.6% ; Table **3.3**, **A** and **B**).

3.2.4.2 Secondary Structural Analysis

Phylogenetic comparison of rRNAs from all three lineages provides support for the idea that despite often considerable differences in their sequences, the secondary and probably the

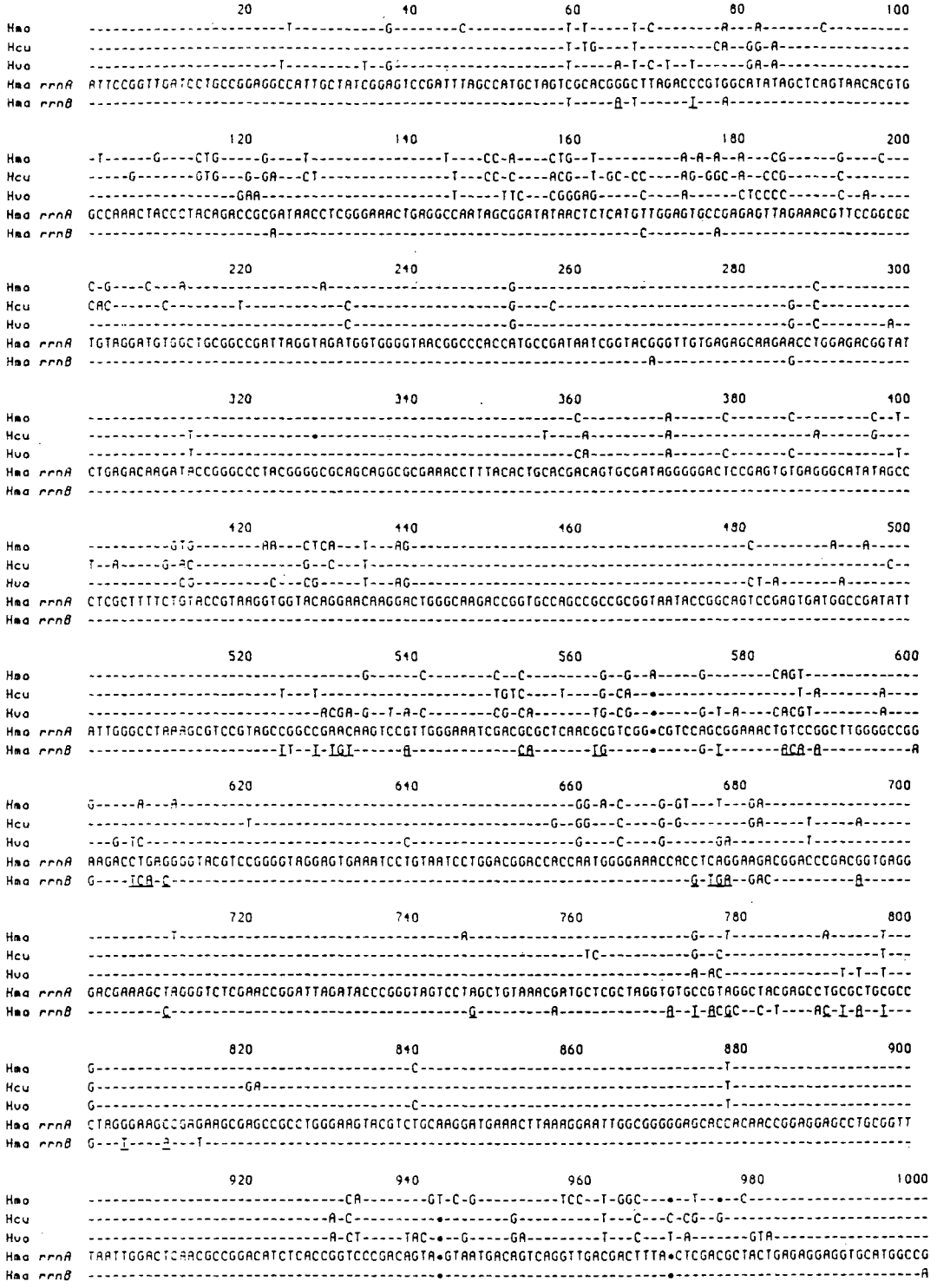

Figure 3.14 Figure caption on next page (page 83).

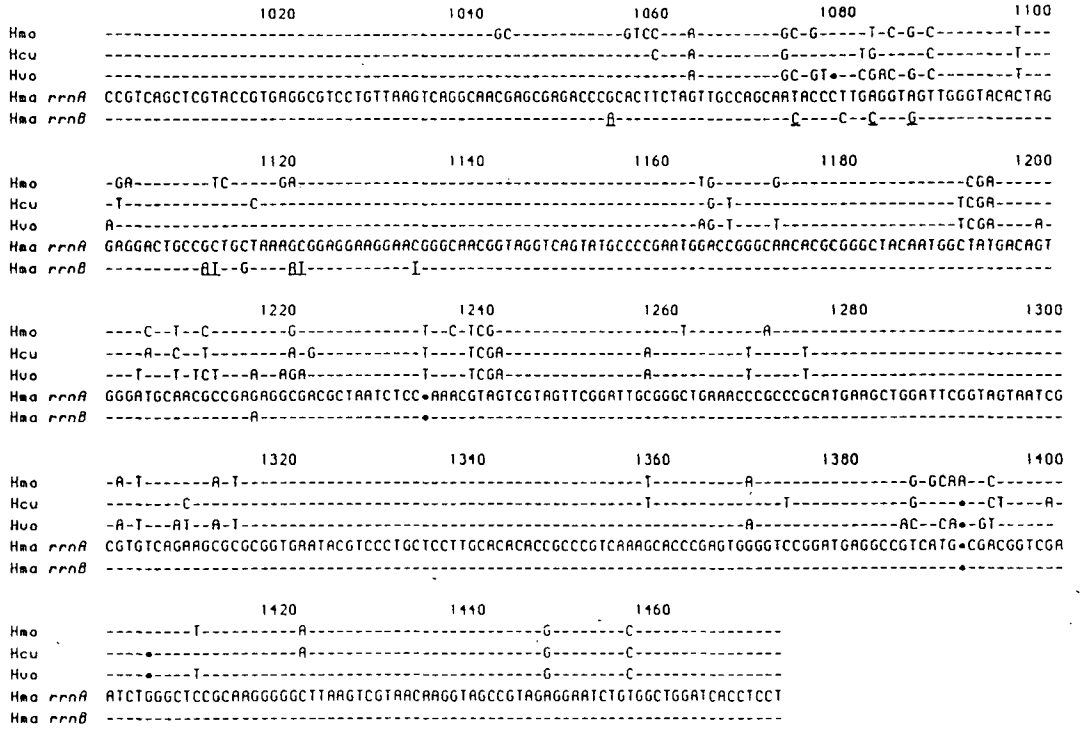

Figure **3.14** Nucleotide sequences and alignment of the 16S rRNA encoding genes. The complete nucleotide sequence of the *Ha. marismortui* 16S encoding gene from the *rrnA* operon is depicted (*Hma rrnA*). Below is the *rrnB* 16S rRNA encoding sequence (*Hma rrnB*); only nucleotides that differ from the *rrnA* sequence are indicated. Substitutions in the *rrnB* sequence that are compensatory, affecting both components of a base pair in a region of RNA secondary structure, are underlined. For comparison, the entire 16S rRNA sequences from *Hb. cutirubrum* (*Hcu*), *Hf. volcanii*(*Hf*), and *Hc. morrhuae* (*Hmo*) are also included; again, only nucleotides that differ from the corresponding nucleotides in the *rrnA* sequences are indicated. Dashes (--) indicate nucleotides identical to the *rrnA* sequence; dots (•) indicate single nucleotide gaps in the sequence(s) required to maintain alignment .

Table 3.3 Comparison of the nucleotide sequences of the 16S rRNA and the 508-823 domain from halophilic archaeal species are summarized. The abbreviation used are as follows: L is the length of the nucleotide sequences compared; Ts is the number of transitional substitutions; Tv is the number of transversional substitutions; Total is the total number of substitutions. Abbreviations of species names are as in the legend to Figure **3.14**.

Comparisons	Nucleotide Differences				Percentage similarities
	L	Ts	Tv	Total	
A. Complete 16S sequence					
rrnA/ rrnB	1472	54	20	74	95.0%
rrnA/ Hvo	1470	118	54	172	88.3%
rrnB/ Hvo	1470	124	60	184	87.5%
rrnA/ Hcu	1471	94	50	144	90.2%
rrnB/ Hcu	1471	120	62	182	87.6%
rrnA/ Hmo	1471	106	56	162	89.0%
rrnB/ Hmo	1471	128	66	194	86.8%
Hvo/ Hcu	1471	121	56	177	88.0%
Hvo/ Hmo	1470	107	66	173	88.2%
Hcu/ Hmo	1470	115	51	166	88.7%
B. 508-823 Domain					
rrnA/ rrnB	316	34	17	51	83.9%
rrnA/ Hvo	316	29	13	42	86.7%
rrnB/ Hvo	316	34	16	50	84.2%
rrnA/ Hcu	316	24	8	32	89.9%
rrnB/ Hcu	316	40	19	59	81.3%
rrnA/ Hmo	316	16	15	31	90.2%
rrnB/ Hmo	316	31	26	57	82.0%
Hvo/ Hcu	316	40	5	45	85.8%
Hvo/ Hmo	316	29	11	40	87.3%
Hcu/ Hmo	316	29	7	36	88.6%

tertiary structures of both small subunit (SS) and large subunit (LS) rRNA have been conserved to a remarkable degree across the entire evolutionary spectrum (Raue' *et al.*, 1988). In *E. coli*, several structural models which predict higher order RNA-RNA and RNA-protein interactions within the 30S subunit of the ribosome have been proposed (Woese *et al.*, 1983; Stern *et al.*,

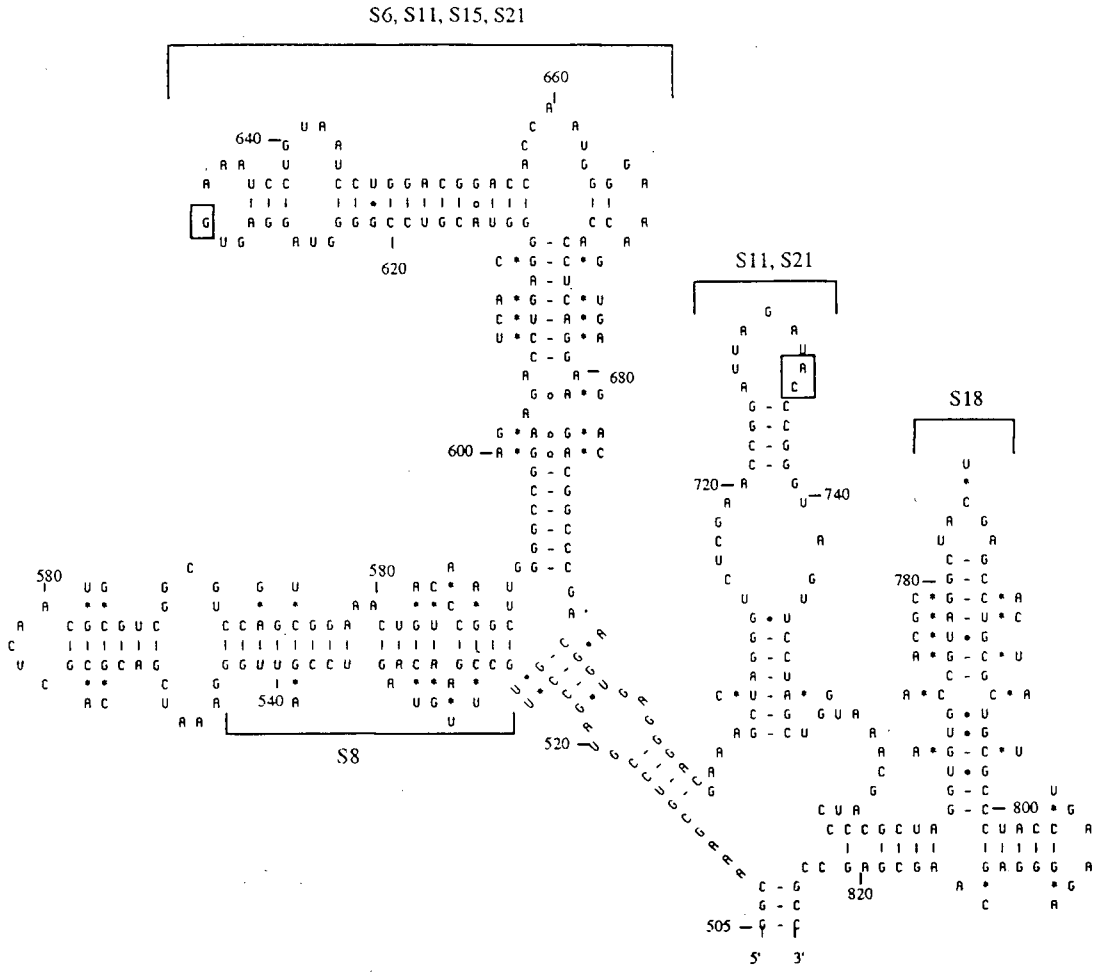

Figure **3.15** Predicted secondary structure for the 508-823 domain of the *rrnA* and *rrnB* 16S rRNAs from *Ha. marismortui.* The secondary structure for the region bounded by nucleotide positions 508-823 for the *rrnA* sequence is illustrated. Mutational differences between *rrnA* and *rrnB* (*), normal Watson-Crick (-), G-U (•), and A-G (o) base pairs are indicated. The boxed nucleotides correspond to the positions in *E.coli* 16S rRNA protected from chemical modifications by tRNA binding to the P site (Moazed and Noller, 1989). Regions that in *E. coli* 16S rRNA interact with the indicated small subunit ribosomal proteins S6, S8, S11, S15, S18, and S21 are also illustrated (Brimacombe *et al.*, 1990).

1988; Noller *et al.*, 1990; Brimacombe *et al.*, 1990; Oakes *et al.*, 1990). Studies on rRNA-protein interactions using archaeal rRNAs and *E. coli* ribosomal proteins indicated functional conservation of at least some ribosomal protein binding sites from one primary kingdom to another (Zimmermann *et al.*, 1980; Thurlow and Zimmermann, 1982; Leffers *et al.*, 1988). Because of the evolutionary conservation, it seems likely that *Ha. marismortui* 30S subunits

containing either the *rrnA* or *rrnB* 16S rRNA sequence will retain the important structural and functional features necessary for protein synthesis (Woese and Pace, 1993).

The sequences and predicted structure for the more variable 508-823 domain (the central domain) are illustrated in Figure **3.15**. Most of the nucleotide differences between *rrnA* and *rrnB* sequences are located within certain helical regions of RNA secondary structure and are compensatory, affecting both components of a nucleotide base pair. The two helical regions bounded by nucleotides 526-588 and 768-800 exhibit many differences between *rrnA* and *rrnB*; however, most of the differences are compensatory. The base of the 526-588 helix corresponds to the region in *E. coli* 16S rRNA where protein S8 binds and the apical portion of the 768-800 helix corresponds to the region where protein S18 binds. Another helical region bounded by nucleotides 599-611 and 673-684 also exhibits many nucleotide substitutions. Ribosomal proteins S6, S11, S15, and S21 bind to sites within one or both of these helical regions between positions 594-690. The two interrupted helical regions bounded by nucleotides 612-672 and 708-749 are invariant except for a compensatory pair of substitutions occurring at positions 710 and 747. The apical loops of these helices correspond to the loops in *E. coli* 16S rRNA which contain nucleotides (boxed in Figure **3.15**) protected from chemical modification by tRNA binding to the P site of the ribosome (Moazed and Noller, 1989). In the 30S subunit, these two loops are in close proximity and form a ring-like structure that defines the platform on the top of the 30S subunit (Oakes *et al.*, 1990).

In *E. coli*, the sequence UUAGAU, in the apical loop of the second helix adjacent to the two tRNA P site nucleotides, has been proposed to interact with a complementary sequence in 23S rRNA during 70S particle formation (Tapprich *et al.*, 1990). The homologous sequence, UUAGAU, is conserved in both the *rrnA* and *rrnB* 16S rRNAs (positions 727-732, Figure **3.15**). The complementary hexanucleotide is also conserved in the *rrnA*, *rrnB* and *rrnC* 23S rRNAs between position 2783 and 2788.

The 986-1158 region also contains 12 substitutions between the *rrnA* and *rrnB* operons; ten of these are compensatory. In *E. coli*, one of the key helices in this region, bounded by

nucleotides 1118-1159, can be crosslinked to proteins S3, S9, and S10. The corresponding helices in the *rrnA* and *rrnB* operons of *Ha. marismortui*, bounded by nucleotides 1052-1106, contain four substitutions within this helix and seven substitutions in the surrounding helices (Figure **3.16**). In *Ha. marismortui*, although the substitutions are not directly involved in the corresponding binding regions of S9 and S3 proteins in *E. coli*, two substitutions are located within the region involved in binding of S10.

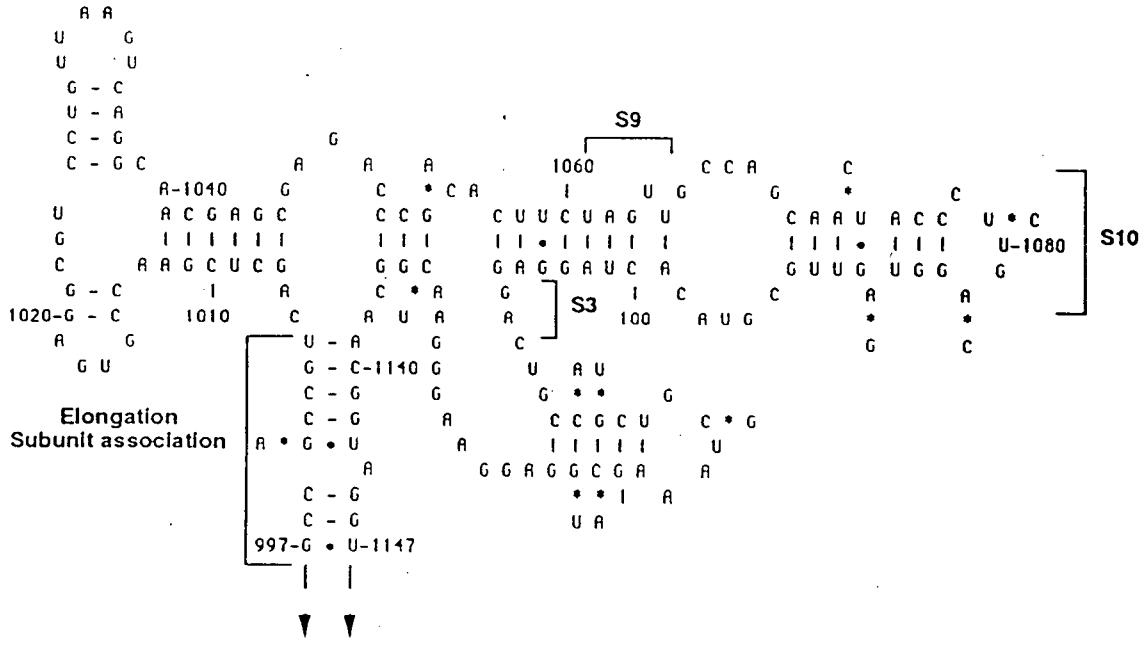

Figure **3.16** Predicted secondary structure for the 986-1158 domain of *rrnA* and *rrnB* 16S rRNAs from *Ha. marismortui*. A likely secondary structure for the region bounded by nucleotide positions 986 and 1158 for the *rrnA* sequence is illustrated. Nucleotide substitutions between *rrnA* and *rrnB* are indicated by asteriks (*). Watson-Crick (-), G-U (•), and A-G (o) base pairs are also indicated. The helices or regions in *Ha. marismortui* that correspond to the regions that interact with the small subunit proteins S3, S9 and S10 in *E. coli* 16S rRNA are also indicated.

In *E. coli*, the differences in sensitivity to chemical modification showed that the helix bounded by nucleotides 1048-1065 and 1191-1198 is crucial for subunit association in *E. coli*

ribosomes. Either removal of this helix (Zwieb *et al.*, 1986) or perturbation of its secondary structure (Meier *et al.*, 1986) causes a severe reduction in the efficiency of formation of the 70S ribosome. In the two operons from *Ha. marismortui*, the corresponding helices are bounded by nucleotides 995-1006 and 1137-1152. In the *rrnA* and *rrnB* sequences a single nucleotide substitution of G1000 -> A is located within this region; however, it does not disrupt the helical structure and probably does not affect the formation of 70S ribosomes.

The sequence and universal structure for the 58-321 domain from the *rrnA* operon is illustrated in Figure **3.17**. Mutational differences between the *rrnA* and *rrnB* operons are indicated by asteriks (*). Four of the substitutions observed within this domain are located within the helix bounded by nucleotides 60-78 and two nucleotide substitutions were located within the helix bounded by nucleotides 157-189. Another helix bounded by nucleotides 200-245 region in *Ha. marismortui* is identical in both the *rrnA* and *rrnB* operons. In *E. coli*, the small subunit protein S17 binding site is located in the corresponding helix (Brimacombe *et al.*, 1990).

3.2.4.3 Expression of *rrnA* and *rrnB* 16S rRNA

In prokaryotic organisms, three steps are required to produce functionally mature rRNAs from the *rrn* operon primary transcript: first, excision of monocistronic precursor RNAs from polycistronic primary transcript; second, accurate recognition and trimming to remove extraneous nucleotides to form mature 5'- and 3'-termini; and third, modification of the base or the ribose moiety at specific sites within the RNA chain. In *E.coli* and presumably also in halobacteria, only 30S subunits that are correctly assembled and active in translation are capable of associating with 50S subunits to form functional 70S particles (Tapprich *et al.*, 1990).

The heterogeneity between the *rrnA* and *rrnB* 16S rRNAs from the *Ha. marismortui* raises two important questions–are both operons actively transcribed and do both 16S sequences get assembled into active ribosomes? In this analysis, protection against nuclease S1 digestion of the end-labeled DNA fragments was used to demonstrate that both *rrnA* and *rrnB* 16S rRNAs

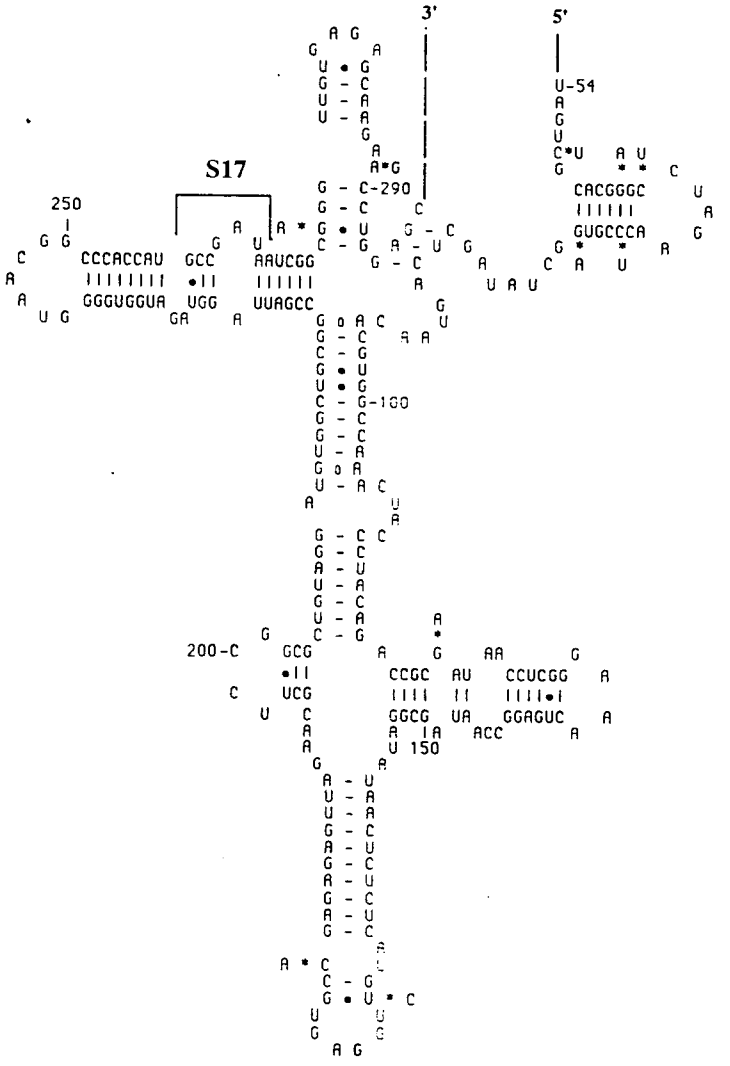

Figure **3.17** Predicted secondary structure for the 58-321 domain of the *rrnA* and *rrnB* 16S rRNAs from *Ha. marismortui*. A likely secondary structure for the region bounded by nucleotides 58-321 for the *rrnA* sequence is illustrated. Nucleotide differences between *rrnA* and *rrnB* 16S rRNAs are indicated by asterisks (*). The corresponding regions in *E. coli* 16S rRNA where the interaction of small subunit protein S17 is also shown.

are present in intact 70S ribosomes (Mylvaganam and Dennis, 1990; see section **2.2.12**). The RNA was obtained from 70S ribosomal particles isolated by sucrose density gradient sedimentation. The DNA probes were two homologous but nonidentical 272 bp *SacII-SmaI* fragments, isolated from the *rrnA* and *rrnB* operons, that overlap a portion of the 508-823 variable domain (nucleotides 462-734 in Figure **3.18**). As a control, the homologous fragment from the clone p4W (PD 655) containing the single copy 16S rRNA gene of *Hb. cutirubrum* was also used (Hui and Dennis, 1985).

In principle, a full length protection product is observed only if the DNA probe fragment is hybridized to an RNA with perfect sequence complementarity. During S1 digestion, partial protection products arise from cleavage of the labelled DNA probe within regions of RNA-DNA duplex that contain one or more mismatched base pairs. The thermostability of RNA-DNA duplexes with perfect sequence complementarity will be greater than the stability of duplexes containing a substantial number of mismatches. The hybridization temperatures and the concentration of RNA were optimized before performing the final analysis. At a temperature of 59°C, the probe hybridizes only to the RNA with perfect sequence complementarity and at 53°C, the probe hybridizes to both perfect and non-perfect RNAs.

Nuclease S1 protection analysis using DNA fragments from both *rrnA* and *rrnB* operons exhibited intense 272 nucleotide long, full length products resulting predominantly from protection by *Ha. marismortui* RNA with perfect sequence complementarity. These full length protection products were visible after hybridization at 53°C and 59°C (Figure **3.18**, lanes 1, 2, 7, and 8). Several clusters of partial length protection products were also visible following hybridization at 53°C (compare lane 1 with 2, and 7 with 8). These partial length products of 54-63, 149-153, 174-184 and 205-213 nucleotides in length are produced by S1 cleavage of the probe in the vicinity of clustered nucleotide differences that exist between the *rrnA* and *rrnB* DNA sequences at positions 675-684, 583-587, 555-566 and 524-532, respectively. The partial products observed from the *rrnA* and *rrnB* probes are expected to be the same because, the mispairings occur at the same positions. The lower bands of 54 - 63 nucleotides in length

appeared at identical positions in both cases, however, the upper bands seem to be digested at different rates in lanes 2 and 8. This may be due to some artifactual effects in the S1 experiments. For example, the S1 nuclease digestion reactions were performed at 35°C for 30 minutes; if the reaction continues for more than 30 minutes, the S1 nuclease continues to digest

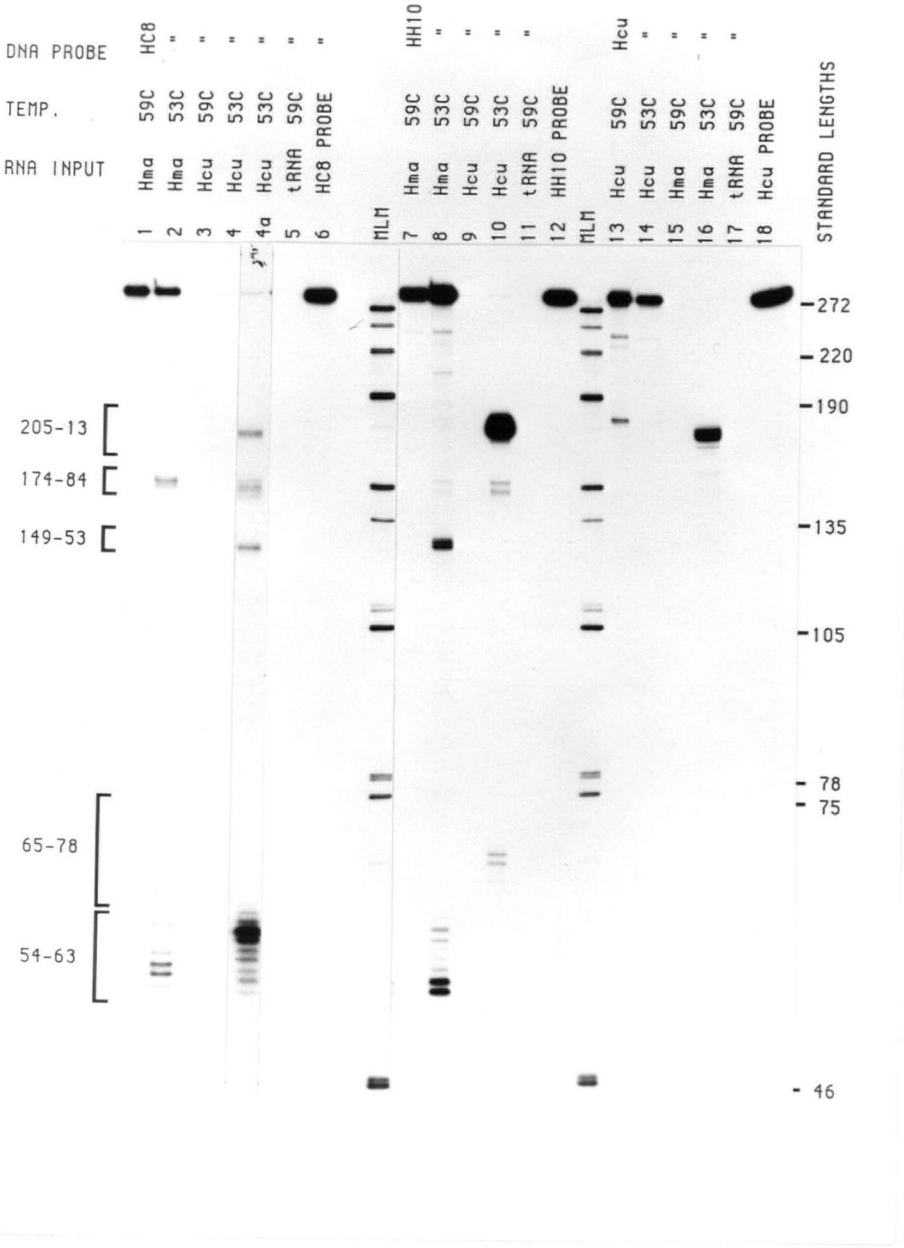

Figure **3.18** Figure caption on next page.

Figure **3.18** Ribosomal RNA protection of end-labelled DNA fragments. The 272-nucleotide long *SacII-SmaI* (nucleotide position 463-734) fragments were isolated from the *Ha. marismortui* pPD 928 (*rrnA* or *HC8*) and pPD 927 (*rrnB or HH10*) clones and from the clone p4W (PD 655) containing the single copy *Hb. cutirubrum* operon. The fragments were labelled at their 5'-ends in the minus strand (position 736) and hybridized to approximately 50 ng of *Ha. marismortui* RNA isolated from 70S ribosomes or to approximatly 200 ng of total *Hb. cutirubrum* RNA (see section **2.1.2.16.2**). For each lane, 1-18, the source of RNA, the hybridization temperature, and the source of DNA are indicated. The S1 digestions are not limit (i.e. exhaustive), but is sufficient to trim the ends of RNA-DNA hybrids and not cut at every internal mismatch sites. Exhaustive digestion produces additional artefacts. Abbreviations are as in the legend to Figure **3.14**.

the longer fragments into smaller fragments (data not shown). In lane 8, the bands of 149 - 153 nucleotides are much stronger than the bands of 174 - 184 nucleotides in length: nuclease S1 may also exhibit a sequence or contextual preference when cleaving the DNA strand in a RNA-DNA hybrid at the site of a mismatch.

As a control, heterologous RNA from *Hb. cutirubrum* was hybridized to the *Ha. marismortui rrnA* and *rrnB* DNA probes at both 53°C and 59°C (lanes 3, 4, 4a, 9 and 10). At higher temperature, little or no partial protection of the DNA probes was observed. At lower temperature, however, partial protection products of 54-63, 65-78,149-154, 174-184 and 205-213 nucleotides in length were clearly visible. Again, these products are produced by S1 cleavage of the DNA probes in the vicinity of clustered nucleotide differences that exist between the *Hb. cutirubrum* and both the *rrnA* and *rrnB* DNA sequences at nucleotide positions 675-684, 657-670, 583-587, 555-566 and 524-532.

As a second control, the DNA probe from *Hb. cutirubrum* was hybridized to either homologous *Hb. cutirubrum* or heterologous *Ha. marismortui* RNA at 53°C and 59°C. With the homologous RNA, a full length product is visible at both temperatures whereas with the heterologous RNA, little or no full length protection product was observed (lanes 13, 14, 15 and, 16). However, at 53°C, heterologous RNA hybridization resulted in partial length

products of 205-213 nucleotides in length. These correspond to S1 cleavage of the *Hb. cutirubrum* DNA probe in the vicinity of clustered nucleotide differences that exist between the *Hb. cutirubrum* DNA probe and both the *rrnA* and *rrnB* RNA sequences between nucleotide positions 524-532. Finally, for all three DNA probes, yeast tRNA failed to produce either full length or partial length protection products regardless of hybridization temperature (lanes 5, 11 and 17 and unpublished results).

These results clearly demonstrate that 16S rRNA sequences derived from both *rrnA* and *rrnB* 16S rRNA genes of *Ha. marismortui* are present in 70S ribosome particles. Particles 70S in size are presumed to have initiated protein synthesis and to be involved in active elongation (Tapprich *et al.*, 1990). At high stringency, both DNA probes exhibit substantial full-length protection. At lower stringency, both DNA probes exhibit many of the partial length protection products expected from S1 cleavage within regions of RNA-DNA sequence noncomplementarity. Although the S1 nuclease digestion assays are only semi-quantitative, the relative intensities of the full length autoradiographic bands suggest that the *rrnA* and *rrnB* type 16S RNAs are (to a first approximation) equally represented in 70S ribosomes. It is also possible that the 16S rRNA transcript of the *rrnC* operon, if present in the 70S ribosomes, may also have hybridized to either *rrnA* or *rrnB* probes and resulted in partial or complete protection products in the above S1 nuclease assays. The presence of the 16S rRNA from the *rrnC* operon in 70S ribosomes can be determined by sequencing the 16S rRNA gene from the *rrnC* operon and using a *rrnC* specific probe in S1 nuclease protection assays (similar to the assays discussed above).

Attempts were also made to verify the presence of two heterologous RNAs from the 70S particles by using the RNA as template in a primer extension-reverse transcription sequencing reaction. It proved problematic to generate an accurate and reliable sequence primarily because of non specific stops caused by RNA secondary structure and the high G+C content of the template RNA. Oren *et al*, (1988) used reverse transcription to determine a portion of the 16S rRNA sequence of *Ha. marismortui* without realizing that there is substantial sequence

heterogeneity; consequently, their sequence contains a large number of uncertain or unidentified nucleotides. This example clearly illustrates a potential error that can be encountered when using reverse transcriptase sequencing of small subunit rRNA in phylogenetic and systematic studies.

In summary, when the nucleotide sequences of the 16S rRNAs from the *rrnA* and *rrnB* operons of *Ha. marismortui* were compared, several features became apparent. The two sequences show differences at 74 of 1472 nucleotide positions, which are located with a single exception to three domain regions, 56-321, 508-823 and 986-1158, within the universal secondary structure model for small subunit rRNA. Most of the differences are located within certain helical regions of RNA secondary structure and are compensatory, affecting both components of a nucleotide base pair. None of the 74 nucleotide differences between the *rrnA* and *rrnB* 16S sequences occurs at positions that have been identified as functionally important for interaction with tRNA, mRNA, or factors during the protein synthesis cycle. The substitutions do not alter the predicted secondary structures significantly or affect nucleotide sites proposed to be important for tertiary interactions. None of the substitutions occurs at nucleotide positions that tend to be highly conserved during evolution. Using nuclease S1 protection analysis it was shown that both the *rrnA* and *rrnB* 16S rRNA sequences are expressed and present in the active 70S ribosomes.

The distribution of nucleotide substitutions between the 16S rRNA sequences of the *rrnA* and *rrnB* shows that they are similar to the differences observed between the two distinct small subunit rRNA genes from *Plasmodium berghei* (see section **1.8**). In *P. berghei*, the two genes are differentially regulated; one is expressed in sporozoites found in the insect host (C type), whereas the other is expressed in the asexual stage found in the blood stream of mammals (A type). The A- and C-types differ by substitution at 72 of 2059 (3.5%) aligned nucleotide positions . As with the two *Ha. marismortui* sequences, the distribution of differences between the two small subunit sequences of *P. berghei* is not random. Virtually all the substitutions and gap differences fall within two domains of the universal secondary structure model for small

subunit RNA (Gunderson *et al.*, 1987; Woese *et al.*, 1983). Interestingly, these two domains are the homologs to the 56-321 and 508-823 domains in *Ha. marismortui* 16S rRNAs. The significance of the clustering of substitutions in both cases is not understood. *Ha. marismortui* does not go through different developmental stages; nonetheless, it is possible that the 16S sequences from the *rrnA* and *rrnB* operons may exhibit differential activity in different conditions of growth.

Sequence comparisons indicate that the degree of identity between the 16S rRNA sequences of the *rrnA* and *rrnB* operons (95%) is substantially less than the 99% value expected in paralogous small subunit rRNA genes in the genomes of most organisms (Lewin, 1990). This indicates that one or more of the intracellular and intraspecies processes (selection, recombination gene conversion, etc.) has been altered in *Ha. marismortui* in such a way as to allow the accumulation of nucleotide differences within the 16S genes of the *rrnA* and *rrnB* operons.

Based on nucleotide sequence similarity alone, it would appear that the divergence of the *rrnA* and *rrnB* 16S rRNA sequences from each other is more recent than their divergence from the other three sequences (from *Hb. cutirubrum*, *Hc. morrhuae* and *Hf. volcanii*). Pairwise comparisons show less similarity (86.8%-90.2%; Table **3.3**)), and suggest that the orthologous small subunit genes from the halophiles are diverged from a common ancestral sequence within a relatively short period of evolutionary time (about 600 million years ago). However, within the 508-823 region, the divergence of the *rrnA* from the *Hb. cutirubrum*, *Hc. morrhuae* and *Hf. volcanii* are more recent (about 500-600 million years) than the divergence of the *rrnB* from the *rrnA* and the other three halophilic sequences (about 800-900 million years). This calculation is based on the assumption that in archaea, the mean rate of substitution is the same as the estimated value for the 16S rRNAs in eubacteria and 18S rRNA in eukaryotes (about 1% substitutions/50 Myr; Ochman and Wilson, 1987).

3.2.5 16S-23S Intergenic Spacer

3.2.5.1 Primary Structural Analysis

The nucleotide sequences of the 16S-23S intergenic spacer regions from the *rrnA* and *rrnB* operons of *Ha. marismortui* have been determined (Figure **3.19**). The two spacer sequences differ in length by 20 nucleotides. The shorter *rrnA* spacer sequence, 385 nucleotides in length, contains a $tRNA^{Ala}$ coding sequence and the longer *rrnB* spacer, 405 nucleotides in length, contains no recognizable universal tRNA-like secondary structure (Figures **3.7 A** and **B**). The 16S-23S spacer sequences from the *rrnA* and *rrnB* operons of *Ha. marismortui* and *Hb. cutirubrum* (Hui and Dennis, 1985) were aligned using the Clustal V method (Des Higgins, European Molecular Biology Laboratory, Germany) and pairwise comparisons were made (data not shown). The *rrnA* and *rrnB* sequences were compared at 381 nucleotide positions and they showed 79 substitutions (i.e., 79.3% identity). The two sequences are identical for 177 nucleotides between the conserved TTAA sequence specific for the Box A motif of the intergenic promoters and the 5'- end of the 23S genes. Most of the substitutions between the *rrnA* and *rrnB* were observed within the region corresponding to the 16S rRNA processing sites 4 and 6 of *rrnB*, and the $tRNA^{Ala}$ sequence in the *rrnA* operon. The substitutions caused by transitions (Ts) out–number transversions (Tv) by 59 to 20. This 3:1 ratio is higher than the Ts/Tv ratio between the 16S rRNA sequences (i.e., about 2:1). The available 254 nucleotides of sequence in front of the 23S rRNA gene of the *rrnC* operon (Brombach *et al.*, 1989) is identical to the sequence in the *rrnB* operon and identical for 177 nucleotides with that of the *rrnA* operon.

Comparison of the 16S-23S spacer sequences from the *rrnA* and *Hb. cutirubrum* operons showed 65% identity within 375 nucleotide positions. The regions comprising the 16S and 23S rRNA processing sites (or the inverted repeat sequences) and the $tRNA^{Ala}$ gene show high levels of sequence conservation. Comparison of the *rrnB* and *Hb. cutirubrum* showed 57.9% identity within 383 nucleotide positions. Except within the 23S inverted repeat sequences and the internal promoter regions, there are no highly conserved regions between the *rrnB* and *Hb. cutirubrum* spacer sequences.

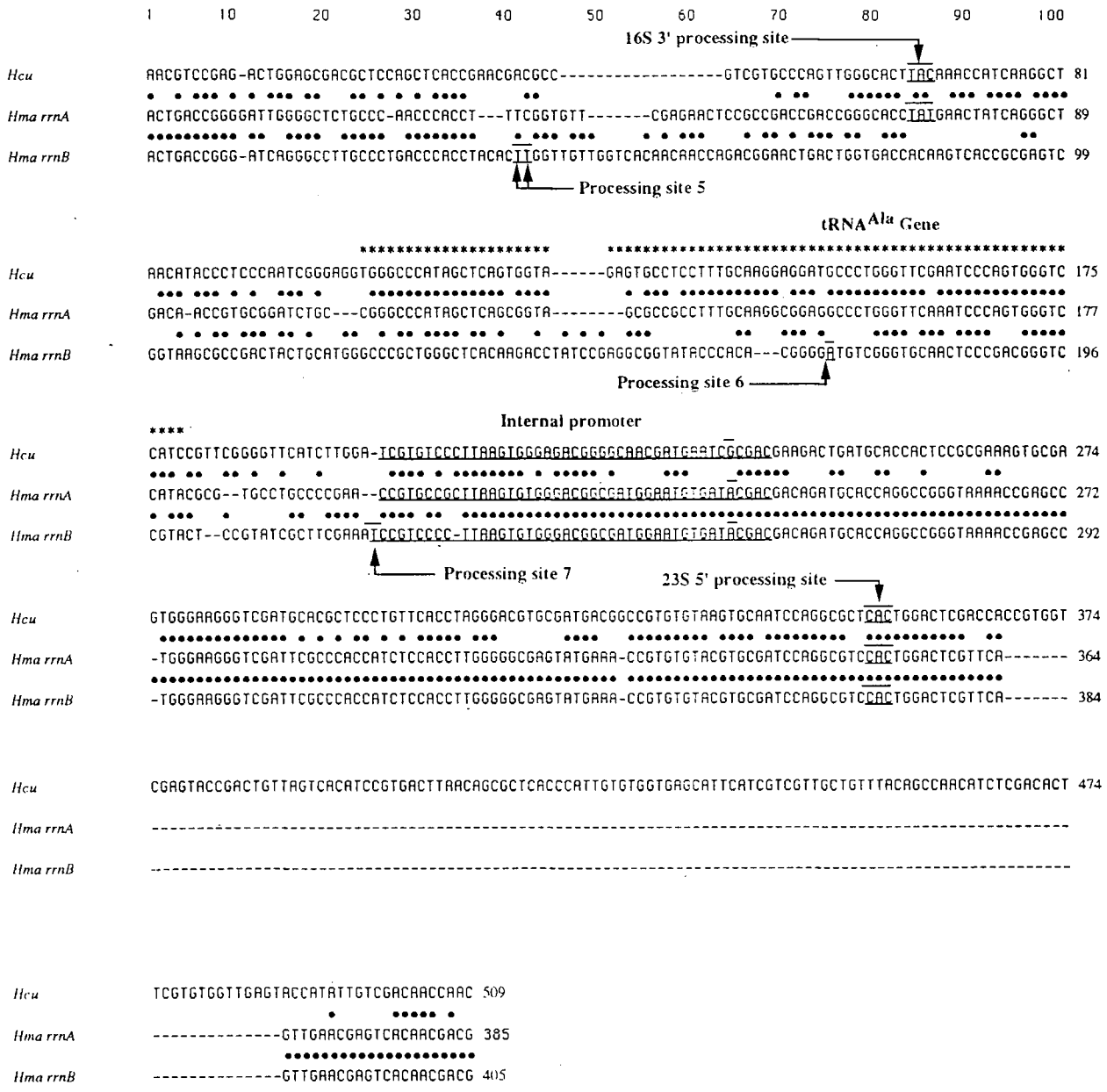

Figure **3.19** Alignment of the 16S-23S intergenic spacer sequences from the *rrnA* and *rrnB* operons of *Ha. marismortui* with the single operon from *Hb. cutirubrum.* The $tRNA^{Ala}$ sequence from the *rrnA* and *Hb. cutirubrum* operons are indicated by asterisks over the sequences. The three nucleotides within the 16S and 23S inverted repeat structures (under and overlined) are involved in the processing of 16S and 23S rRNAs from the *rrnA* and *Hb. cutirubrum* operons and processing of 23S rRNA from the *rrnB* operon. The processing sites 4, 5 and 6 observed in the *rrnB* operon are also under and overlined (indicated by arrows). An internal promoter sequence is underlined and the start site is overlined within the sequence.

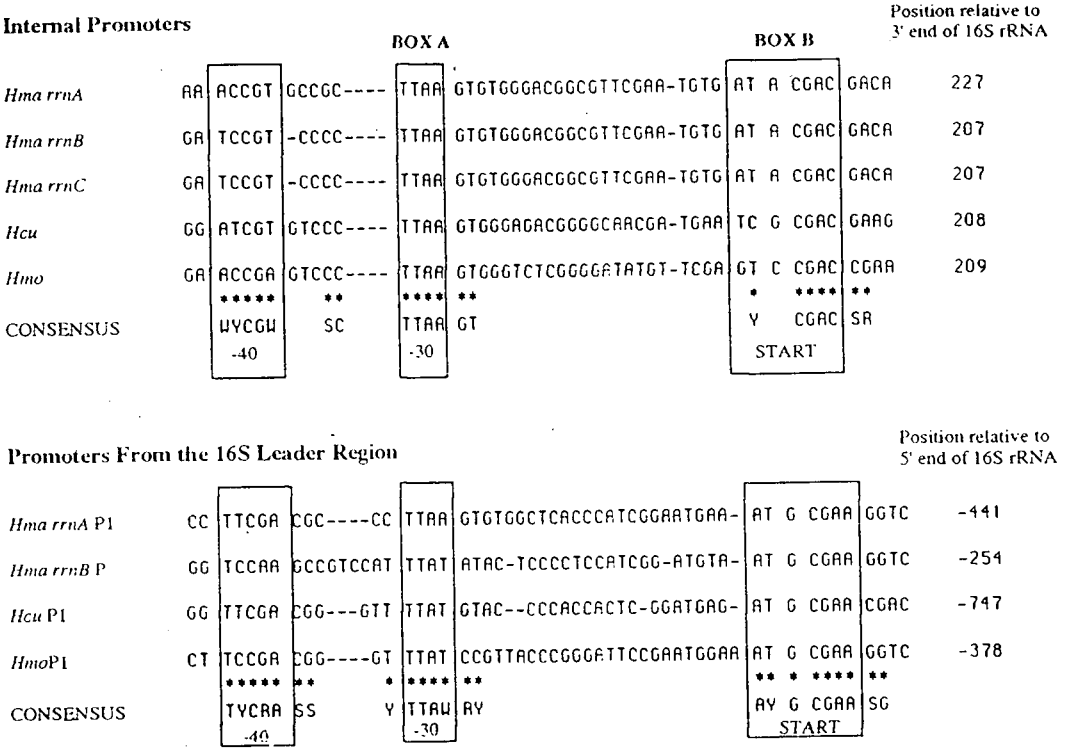

Figure **3.20** Comparison of the internal promoter sequences of the *rrnA*, *rrnB*, *Hb. cutirubrum* and *Hc. morrhuae* operons. The sequences from the intergenic promoters from the halophiles are compared above and also one promoter sequence from the 16S 5'-flanking regions of each organism are compared below. The start sites which have been mapped at the nucleotide position are shown as the third nucleotide within Box B. The position of start sites relative to the 5'- or 3'- end of the 16S sequence are indicated on the right. This figure illustrates that the primary sequences of the Box A, Box B and -40 regions from the intergenic promoters and the 16S leader promoters from halophiles are conserved.

An intergenic promoter sequence (Pi) was found in the 16S-23S spacer region of the *rrnA*, *rrnB* and *rrnC* operons. These Pi sequences show a high degree of identity to the Pi sequences from related halophiles and the consensus motif of the promoter sequences present within the 16S 5'-flanking region (Figure **3.20**).

3.2.5.2 Primary Transcript Analysis of the *rrnA* 16S-23S Spacer

Nuclease S1 digestion experiments were used to investigate transcription through the intergenic spacer. For the analysis of the *rrnA* operon, a 646 nucleotide long *AvaI-AvaI* fragment that contains the last 110-bp of the 16S gene, the 383-bp long 16S-23S intergenic spacer, and the first 153-bp of the 23S gene, was isolated. The fragment was 3'-labelled on the minus strand at position 1362 within the 16S gene. The labelled fragment was hybridized to total *Ha. marismortui* RNA and treated with S1 nuclease. Maxam and Gilbert sequencing reactions for A and A+G were also performed with the same labelled fragment and the reaction products were separated on a 6% polyacrylamide gel, along with the protected fragments from the S1 nuclease assay (Figure **3.21A**).

In addition to full protection, a number of different partial protection products of 110 , 180, 216, 246, 286, 460 and 493 nucleotides in length were observed. These products have 3'-ends near positions 1472 of the 16S rRNA gene, positions 70, 106, 136, 176, and 350 within the spacer region and position 1 of the 23S rRNA gene, respectively. They presumably correspond to protection by RNAs with 3'- ends generated by cutting the primary transcript (i) at the end of 16S rRNA, (ii) at the 16S precursor processing site, (iii) at the 5'- end of the $tRNA^{Ala}$ spacer (iv) at a site within the anticodon loop of the $tRNA^{Ala}$, (iv) at the 3'- end of the $tRNA^{Ala}$, (vi) at the 23S precursor processing site and (vii) at the beginning of the 23S sequence, respectively.

The 646 nucleotide long *AvaI-AvaI* fragment was also 5'- labelled on the plus strand at position 153 within the 23S gene. The labelled fragment was hybridized to total *Ha. marismortui* RNA and treated with S1 nuclease and the reaction products were separated on a 6% polyacrylamide gel, along with the probe and size standard markers (Figure **3.22 A**).

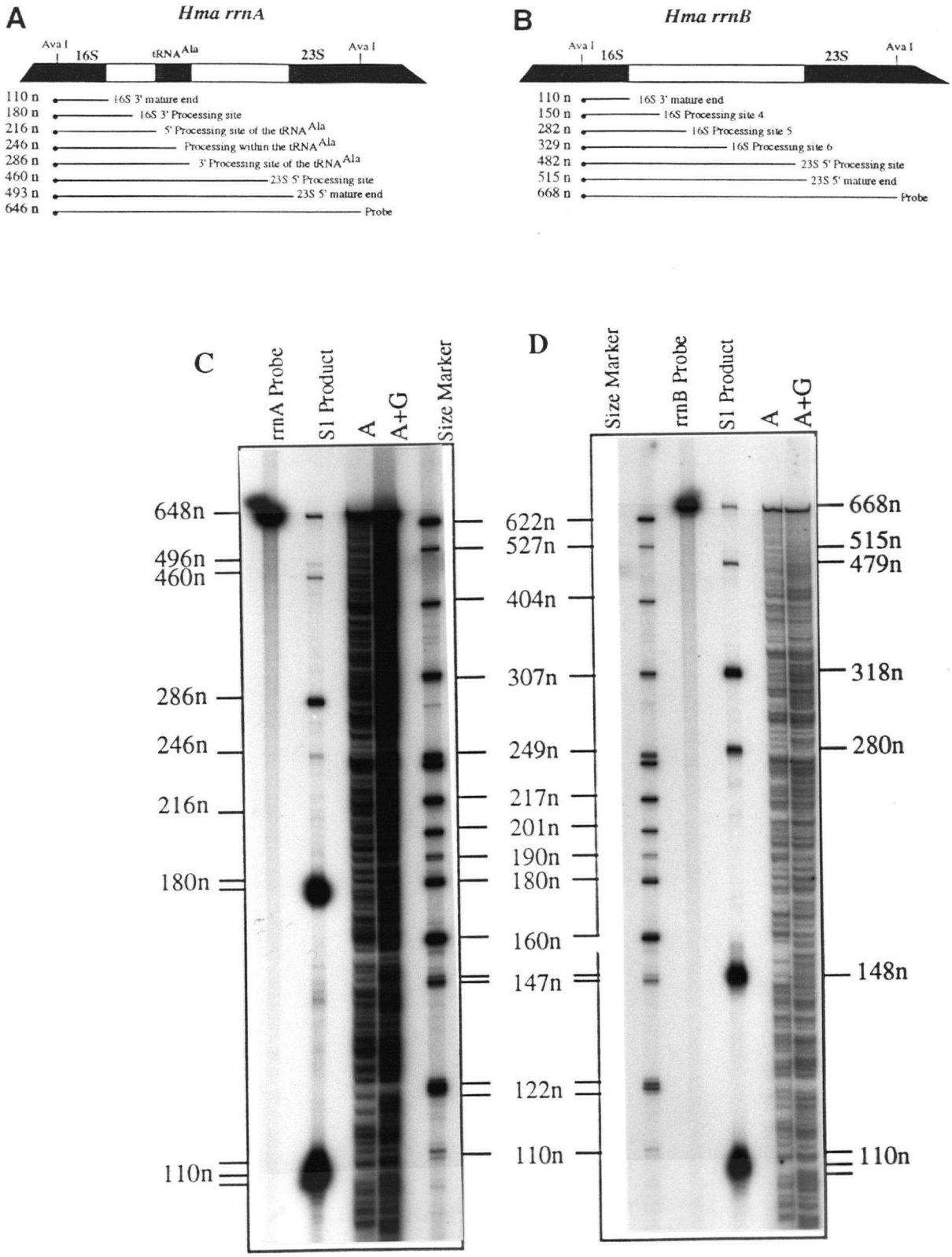

Figure **3.21** Figure caption on pages 101.

Figure **3.21** Nuclease S1 protection assay of the primary transcript products within the 16S-23S intergenic spacer region of the *rrnA* and *rrnB* operons using DNA probes labelled at their 3'- ends.

Figure **3.21A** A line diagram showing the 16S-23S intergenic spacer region of the *rrnA* operon (from plasmid PD 926). The two *AvaI* sites within the 16S rRNA and 23S rRNA were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The position where the probes were labelled with α-^{32}P dCTP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.21B** A line diagram showing the 16S-23S intergenic spacer region of the 16S rRNA gene of the *rrnB* operon (from plasmid PD 929). The two *AvaI* sites within the 16S rRNA and the 23S rRNA were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The position where the probes were labelled with α-^{32}P dCTP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.21C** The autoradiogram showing the nuclease S1 protection assay products from the 16S-23S spacer region of the *rrnA* operon. Above each lane, the *rrnA* probe, S1 product (S1 nuclease digestion product of the annealed probe with total RNA), marker (pBR 322 plasmid digested with *MspI* and labelled at the 3'-end with α-^{32}P dCTP) and the Maxam and Gilbert sequencing products (G and A+G) are indicated. The sizes of the protected fragments from the S1 treated DNA-RNA hybrids and the markers are shown on both sides of the autoradiogram.

Figure **3.21D** The autoradiogram showing the nuclease S1 protection assay products from the 16S-23S spacer region of the *rrnB* operon. Above each lane, the *rrnB probe*, S1 product (S1 nuclease digestion product of the annealed probe with total RNA), marker (pBR 322 plasmid digested with *MspI* and labelled at the 3'- end with α-^{32}P dCTP) and the Maxam and Gilbert sequencing products (G and A+G) are indicated. The sizes of the protected fragments from the S1 treated DNA-RNA hybrids and the markers are shown on both sides of the autoradiogram.

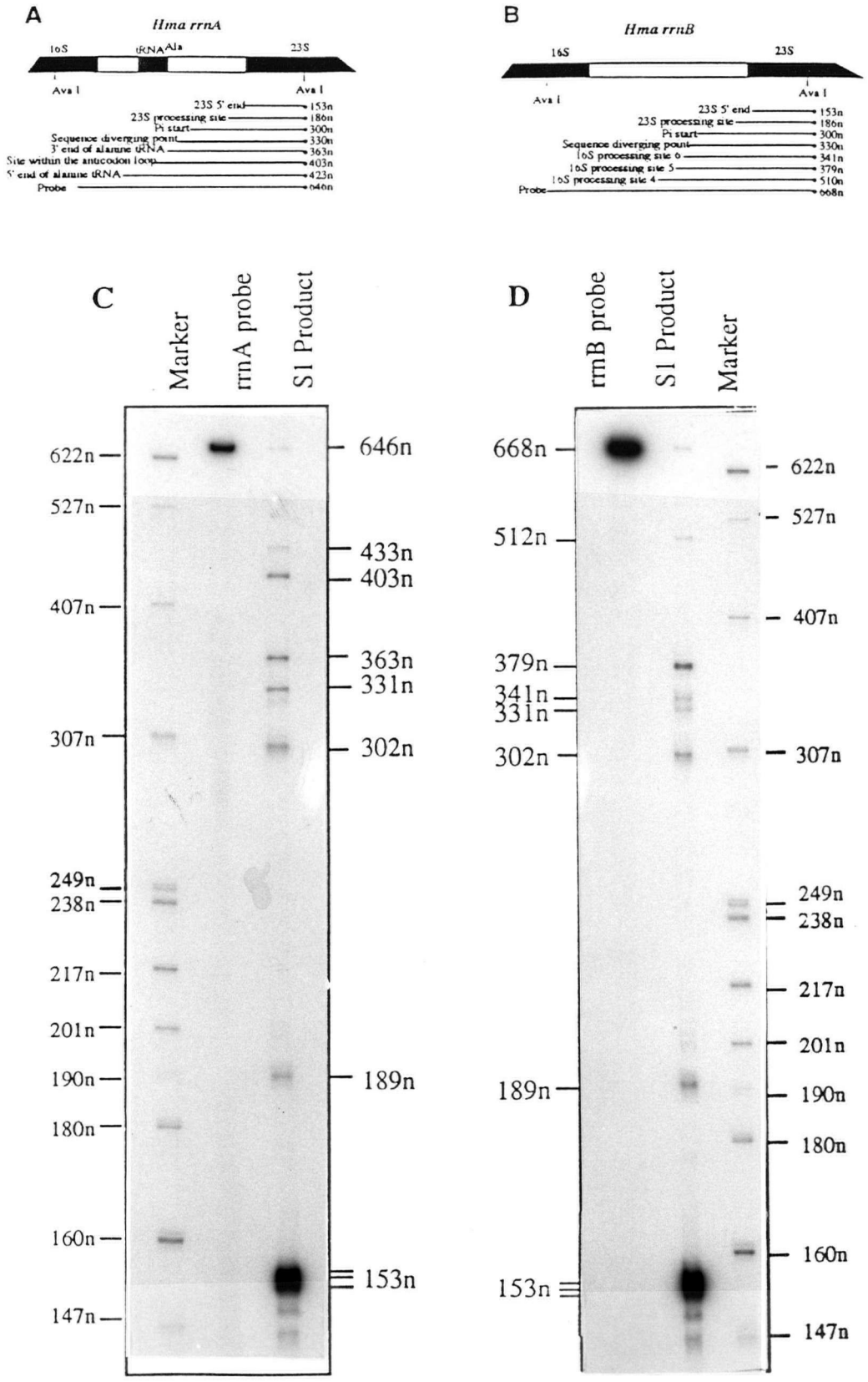

Figure **3.22** Figure caption on pages 103.

Figure **3.22** Nuclease S1 protection assay of the primary transcript products within the 16S-23S intergenic spacer region of the *rrnA* and *rrnB* operons using DNA probes labelled at their 5'- ends.

Figure **3.22A** A line diagram showing the 16S-23S intergenic spacer region of the *rrnA* operon (from plasmid PD 926). The two *AvaI* sites within the 16S rRNA and 23S rRNA were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The position where the probes were labelled with γ-^{32}P ATP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.22B** A line diagram showing the 16S-23S intergenic spacer region of the 16S rRNA gene of the *rrnB* operon (from plasmid PD 929). The two *AvaI* sites within the 16S rRNA and the 23S rRNA were used in the isolation of the probe for the nuclease S1 protection assay (see section **2.2.16.1**). The position where the probes were labelled with γ-^{32}P ATP are indicated by dots (•). The sizes of the protected fragments and their corresponding sites are shown below the operon structure.

Figure **3.22C** The autoradiogram showing the nuclease S1 protection assay products from the 16S-23S spacer region of the *rrnA* operon. Above each lane, the *rrnA* probe, S1 product (S1 nuclease digestion product of the annealed probe with total RNA) and markers (pBR 322 plasmid digested with *MspI* and labelled at the 3'-end with α-^{32}P dCTP) are indicated. The sizes of the protected fragments from the S1 treated DNA-RNA hybrids and the markers are shown on both sides of the autoradiogram.

Figure **3.22D** The autoradiogram showing the nuclease S1 protection assay products from the 16S-23S spacer region of the *rrnB* operon. Above each lane, the *rrnB* probe, S1 product (S1 nuclease digestion product of the annealed probe with total RNA) and markers (pBR 322 plasmid digested with *MspI* and labelled at the 3'-end with α-^{32}P dCTP) are indicated. The sizes of the protected fragments from the S1 treated DNA-RNA hybrids and the markers are shown on both sides of the autoradiogram.

In addition to full protection, a number of different partial protection products of 153, 186, 300, 330, 360, 403 and 423 nucleotides in length were observed. These products have 5'- ends near positions 1 of the 23S rRNA gene and near positions 350, 236, 207, 176, 136, 105, and 70 within the spacer region, respectively. They presumably correspond respectively to protection by RNAs with 5'-ends generated by cutting the primary transcript (i) at the beginning of the 23S gene, (ii) at the 23S precursor processing site, (iii) at the site of transcription initiation from the internal promoter, (iv) at the sequence diverging point between the *rrnA* and *rrnB* sequences, (v) at the 3'- end of the $tRNA^{Ala}$ sequence, (vi) at a site within the anticodon loop of the $tRNA^{Ala}$, (vii) at the 5'- end of the $tRNA^{Ala}$, and (viii) at the 16S rRNA precursor processing site. Multiple transcript ends observed near the processing and end sites of the 16S rRNAs (Figures **3.22C** and **3.22D**) and 23S rRNAs (Figures **3.23C** and **3.23D**) may have been produced by endogeneous nuclease activities acting on the ends of the RNAs or alternatively may have been caused by inaccurate S1 trimming at the ends of RNA-DNA hybrids. The presence of other minor bands which are not labelled, could also be due to S1 artefacts or could represent unstable or abnormal products produced during the rRNA maturation process.

Since S1 nuclease analysis are only semi–quantitative, the band intensities do not reflect the exact amount of RNAs present in the cells. These analysis are mainly done in order to detect the transcription initiation, termination or processing sites of the transcripts. In the case of rRNA transcripts, usually we expect only about 1 - 2% of rRNA in the form of very long precursors and processing intermediates; most of the rRNA is in the mature form (King and Schlessinger, 1983). However, this is not what has been observed from the S1 nuclease analysis shown in Figures **3.21**, **3.22** and **3.23**; the intensities of the bands representing the mature 16S and 23S rRNAs are much less than expected (Figures **3.21C, 3.21D, 3.22 C** and **3.22 D**). There are several explanations for this. First, the efficiency of hybridization between rRNA and the DNA probe may depend on the lengths of the rRNAs; longer RNAs hybridize more efficiently than the shorter RNAs and during a three hour hybridization, displacement may occur. Second, the conditions used in the S1 analysis (especially temperature) may not have permitted the mature rRNA (smallest in size) to hybridize efficiently.

Third, majority of the mature rRNAs are folded into a secondary structural conformation and may not be readily available to hybridize with the DNA probe.

Primer extension analysis are performed in order to locate more precisely the end sites of the products identified from S1 nuclease assays (the approximate positions of the end sites are already determined using DNA ladders, as shown in Figures **3.21C, 3.21D, 3.22C** and **3.22D**). Using oPD 43 (positions 106-125 in *rrnA*; Figure **3.19**), the processing site at the 3'-end of the $tRNA^{Ala}$ was mapped at a U residue at position 176 (Figure **3.23B**). There was very little extension product through this site because the tRNA structure represents an efficient block to reverse transcription. Primer extension analysis using oligonucleotide oPD 44 (positions 37-59 within the 23S rRNAs of *rrnA* and *rrnB*; Figure **3.24**) showed that the 5'-ends of the two 23S rRNAs stopped at U residues at position 1 of the 23S rRNAs (Figure **3.23C**). The 23S precursor processing sites were also mapped at a C residue within the bulge motif located at positions 350 and 370 (Figure **3.19**) within the intergenic spacer of *rrnA* and *rrnB* operons, respectively (Figure **3.23C**). Since the first 70 nucleotides in the 23S rRNA genes and the 177 nucleotide upstream regions are identical in both the *rrnA* and *rrnB* operons, the primer extension analysis using oPD 44 gave identical products corresponding to the 5'- end of 23S rRNA and the 5'- processing site.

Nuclease S1 digestion experiments within the 16S-23S spacer regions of the two rRNA operons of *Ha. marismortui* showed that the two internal promoters are active (section **3.2.5.3**). If one considers the possibility of nonspecific premature transcription termination inside the ribosomal operons (i.e. between the 16S and 23S rRNAs or within the 23S rRNA genes), it would lead to a slight increase in the level of 16S rRNA over the 23S and 5S rRNAs. The activity of the Pi promoters could possibly result in excess 23S and 5S rRNAs relative to 16S rRNAs to compensate for such an imbalance and adjust the cellular level of the ribosomal RNAs.

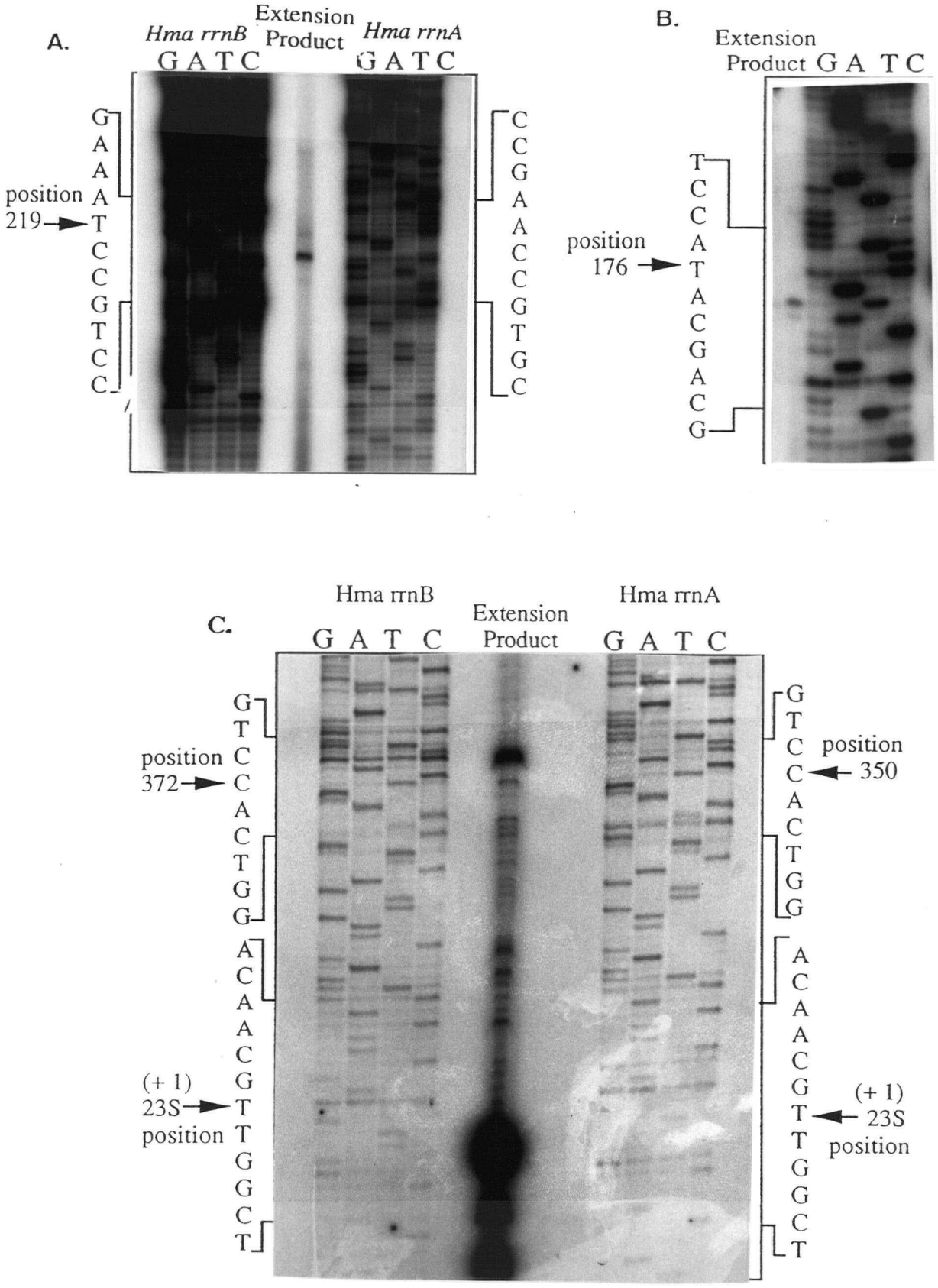

Figure **3.23** Figure caption on next page (page 107).

Figure **3.23** Primer Extension Analysis within the 16S-23S spacer regions of the *rrnA* and *rrnB* operons.

Figure **3.23A** Primer extension analysis to map the processing site 6 (Figure 3.7B) within the spacer of the *rrnB* operon using oligonucleotide oPD 46. Sequencing ladders from the *rrnA* (PD 926) and *rrnB* (PD 929) operons using the same primer, oPD 46, were also generated in order to identify the nucleotide positions at which the extension products terminate.

Figure **3.23B** Primer extension analysis to map the 3'- processing site of the $tRNA^{Ala}$ from the *rrnA* operon using oligonucleotide oPD 43. In order to identify the exact nucleotide position of the extension products, the spacer region of the *rrnA* operon (PD 926) was also sequenced using the same oligonucleotide, oPD 43.

Figure **3.23C** Primer extension analysis to map the 5'-end of 23S rRNA and 5'- processing sites for the *rrnA* and *rrnB* operon transcripts. In order to identify the exact nucleotide position of the extension products, sequencing analysis were also carried out within the spacers of the *rrnA* (PD 926) and *rrnB* operon (PD 927) using oPD-44.

A number of conclusions can be made from the studies involving nuclease protection and primer extension analysis concerning transcription and processing in the intergenic spacer of the *rrnA* operon. First, as evident from the full length protection product, the 16S and 23S rRNA genes are co-transcribed. Second, the internal promoters are active and give rise to transcripts with a 5'- end at position 236 within the intergenic spacer. Third, a 331 nucleotide long product was observed when using the 5'-end labelled probe. This was due to partial protection of the *rrnA* probe with transcripts derived from the *rrnB* operon where the sequences of the two operons start to diverge at positions corresponding to 207 and 228 within the spacer of the *rrnA* and *rrnB* operons, respectively (see Figure **3.19**). Finally, all other end sites appear to be generated by endonuclease cleavage of the primary transcript since both 3'- and 5'-ends can be mapped to each of the positions. Clearly, there appears to be no predetermined ordering of cleavage events since all possible products were observed with both 5'- and 3'-labelled probes. The only exception is the maturation of the 3'-end of 16S rRNA where precursor excision occurs before maturation. It is uncertain if maturation is the result of an endo- or exonuclease activity. In contrast, maturation at the 5'- end of 23S rRNA can

sometimes occur before precursor excision as evident from the presence of a 496 nucleotide protection product (Figure **3.21 A** and **B**)

Although there is no mandatory order for intergenic processing events, some sites appear to be cleaved earlier than others. These include the 16S precursor excision site and the $tRNA^{Ala}$ 3'- end processing site. It is likely that the 23S precursor excision and maturation occur at a later time because the RNA polymerase must traverse the entire 23S gene before the second half of the inverted repeat strcture is available to form the processing stem.

3.2.5.3 Primary Transcript Analysis of the *rrnB* 16S-23S Spacer

For the nuclease S1 protection analysis of the 16S-23S spacer region of the *rrnB* operon, a 668 nucleotide long *AvaI-AvaI* fragment containing the last 110-bp of the 16S gene, the 405-bp long 16S-23S intergenic spacer, and the first 153-bp of the 23S gene, was isolated. The fragment was 3'- labelled on a minus strand at position 1362, located within the 16S gene. The labelled fragment was hybridized with the total RNA of *Ha. marismortui* and then treated with S1 nuclease. The same labelled fragment was also used in the Maxam and Gilbert sequencing reactions A and A+G, and the products were separated on a 6% polyacrylamide gel, along with the protected fragments from the S1 nuclease assay (Figure **3.21D**). In addition to full protection, a number of different partial protection products of about 110, 148-151, 282, 329, 482, and 515 nucleotides were observed. These partial protection products correspond respectively to protection by RNAs with 3'-ends generated by cutting the primary transcript (i) at the end of 16S rRNA, (ii) at the 16S rRNA precursor processing site 4, (iii) at the 16S rRNA precursor processing site 5, (iv) at the 16S rRNA precursor processing site 6, (v) at the 23S rRNA precursor processing site, and (vi) at the beginning of the 23S rRNA.

The 668 nucleotide long *AvaI-AvaI* fragment was also 5'-labelled at position 153 on the plus strand within the 23S gene. The labelled fragment was hybridized with the total RNA of *Ha. marismortui* and then treated with S1 nuclease and the reaction products were separated on a 6% polyacrylamide gel, along with the probe and size standard markers (Figure **3.22 D**). In

addition to full protection, a number of different partial protection products of about 153, 186, 300, 330, 341, 379, and 510 nucleotides in length were observed. These correspond to protection by RNAs with 5'-ends generated by cleavage of the primary transcript; (i) at the beginning of the 23S rRNA gene, (ii) at the 23S precursor processing site, (iii) at the site of transcription initiation from the internal promoter, (iv) at the sequence diverging point between the *rrnA* and *rrnB* sequences, (v) at the 16S rRNA precursor processing site 6, (vi) at the 16S rRNA precursor processing site 5, and (vii) at the 16S rRNA precursor processing site 4.

Primer extension analysis were carried out in order to locate more precisely the positions of some of the end products obtained from S1 nuclease analysis. In the *rrnB* operon, the precise mapping of the 16S precursor processing site (site 6 in Figure **3.7B**) was performed by primer extension analysis using oligonucleotide oPD 46 (Figure **3.23A**) which binds at positions 292-311 within the intergenic spacer of the *rrnB* operon (Figure **3.19**). The extension product stopped at a U residue at position 219 (Figure **3.23A**). Primer extension analysis using oligonucleotide oPD 44 (positions 37-59 within the 23S rRNAs of *rrnA* and *rrnB*; Figure **3.24**) showed that the 5'-ends of the two 23S rRNAs (*rrnA* and *rrnB*) stopped at U residues at position 1 of the 23S rRNAs (Figure **3.23C**). The 23S precursor processing sites were also mapped at a C residue within the bulge motif located at positions 350 and 370 (Figure **3.19**) within the intergenic spacer of *rrnA* and *rrnB* operons, respectively (Figure **3.23C**).

A number of conclusions concerning the transcription and processing of the intergenic spacer of the *rrnB* operon can be reached from these nuclease S1 protection and primer extension analysis. First, as evidenced by the presence of full length protection product, the 16S and the 23S rRNA genes are co-transcribed. Second, the internal promoter is active and gives rise to transcripts with a 5'-end at approximately position 260. Third, a 330 nucleotide long product was observed when using the 5'-end labelled probe. This is likely due to partial protection of the *rrnB* probe with transcripts derived from the *rrnA* operon where the sequences of the two operons start to diverge at positions corresponding to 207 and 228 within the spacer

and *rrnB* operons, respectively. Finally, all other end sites appear to be generated by endonuclease cleavage of the primary transcript, because both the 3'- and the 5'- ends can be mapped to each cleavage position. There is no mandatory order of cleavage events since all products are observed with both 5'- and 3'- probes. The only exception being the maturation of the 3'- end of 16S rRNA where precursor excision precedes maturation.

In the 16S-23S rRNA spacer of the *rrnB* operon, three endonuclease processing sites have been identified, and labelled as sites 4, 5, and 6 located at positions 36 to 41, 172, and 219 (Figure **3.7B**). Since the nucleotide sequences of the *rrnA* and *rrnB* spacer regions are identical within the 177 nucleotides immediately upstream to the 23S rRNA genes, the two 5'-labelled probes detected identical 23S intermediates and maturation products . In addition, the nuclease S1 protection assay showed that the maturation of the 23S 5'- end can occur before the processing of primary transcripts; however, the intermediates are minor when compared to other intermediates from the processing reactions. Since all protection products can be accounted for as either endonuclease processing sites, internal promoter initiation sites or *rrnA-rrnB* sequence divergence sites, it implies that there is apparently no interference with transcripts derived from the *rrnC* operon intergenic space.

3.2.5.4 The Processing Pathways Within the 16S-23S Spacer

It is postulated that the "bulge-helix-bulge" motifs on both the 16S and 23S processing stems are the substrate for the processing endonuclease in archaeal organisms (Thompson and Daniels, 1988). The endonuclease activity is also present in *Ha. marismortui* and is used to excise precursor 16S and 23S rRNAs from the *rrnA* operon transcript and precursor 23S rRNA from the *rrnB* operon transcript. The 16S processing stem in the *rrnB* operon transcript is different in that it lacks the "bulge-helix-bulge" motif and precursor 16S rRNA is presumably excised by one or more alternative endonuclease activities.

The S1 nuclease and primer extension analysis have been used to locate endonuclease cleavage sites in the 5'-flanking region and the 16S-23S intergenic spacer of the *rrnB* operon transcript. Three unique sites were located in the intergenic spacer. These sites exhibit no apparent sequence or structural conservations although the nucleotides involved in the cleavages are always either an A or a U and are associated with either a G • U or an AoG base pair in the universal secondary structure for primary transcript (Figure **3.7B**). Comparison of the processing site from the 5'-leader also shows that the nucleotide involved in the cleavage is an A and is located adjacent to a G • U base pair in the universal secondary structure. These findings suggest that the presence of a G • U or an AoG base pair adjacent to the cleavage site might be an important feature for the processing by endonuclease enzyme. It is uncertain whether all the cleavages observed in the processing of the 16S rRNA from the primary transcript of *rrnB* operon are mediated by a single activity or whether or more than one activity is involved.

3.2.6 23S rRNA

3.2.6.1 Primary Structural Analysis

The complete 2917 nucleotide long sequences of the *rrnA* and *rrnB* 23S rRNA genes have been determined. A third, 2917 nucleotides long 23S rRNA sequence from *Ha. marismortui*, different from the *rrnA* and *rrnB* operon sequences and designated *rrnC*, has been determined by Brombach *et al.*, (1989). It is uncertain if the Brombach sequence is entirely derived from *rrnC* or if it is a composite derived from the three different operons; here we assume that it is not a composite and represents the *rrnC* sequence. Comparisons of the 23S rRNA sequences indicate that 29 substitutions are unique to *rrnA*, 11 substitutions are unique to *rrnB* and only four substitutions are unique to *rrnC* (Figure **3.24**).

The three 23S *Ha. marismortui* sequences were compared to each other and with other related halophilic archaea, *Hb. cutirubrum* (2905 nucleotides) and *Hc. morrhuae* (2927 nucleotides). Alignment of the sequences was obtained by the Clustal V method (Figure **3.25**). Secondary structural features were also taken into consideration to obtain the optimal alignment.

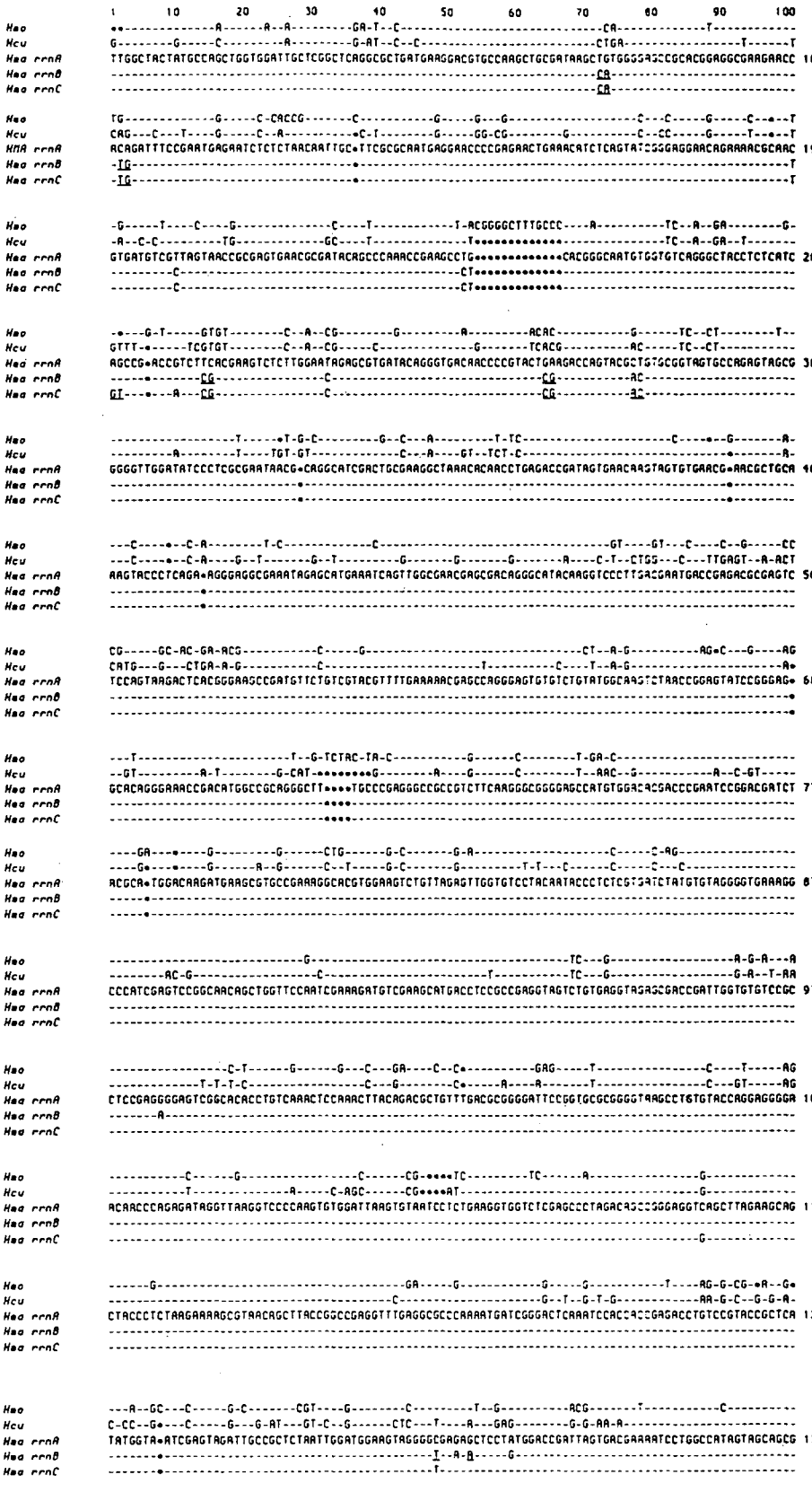

Figure 3.24 Figure caption on page 113

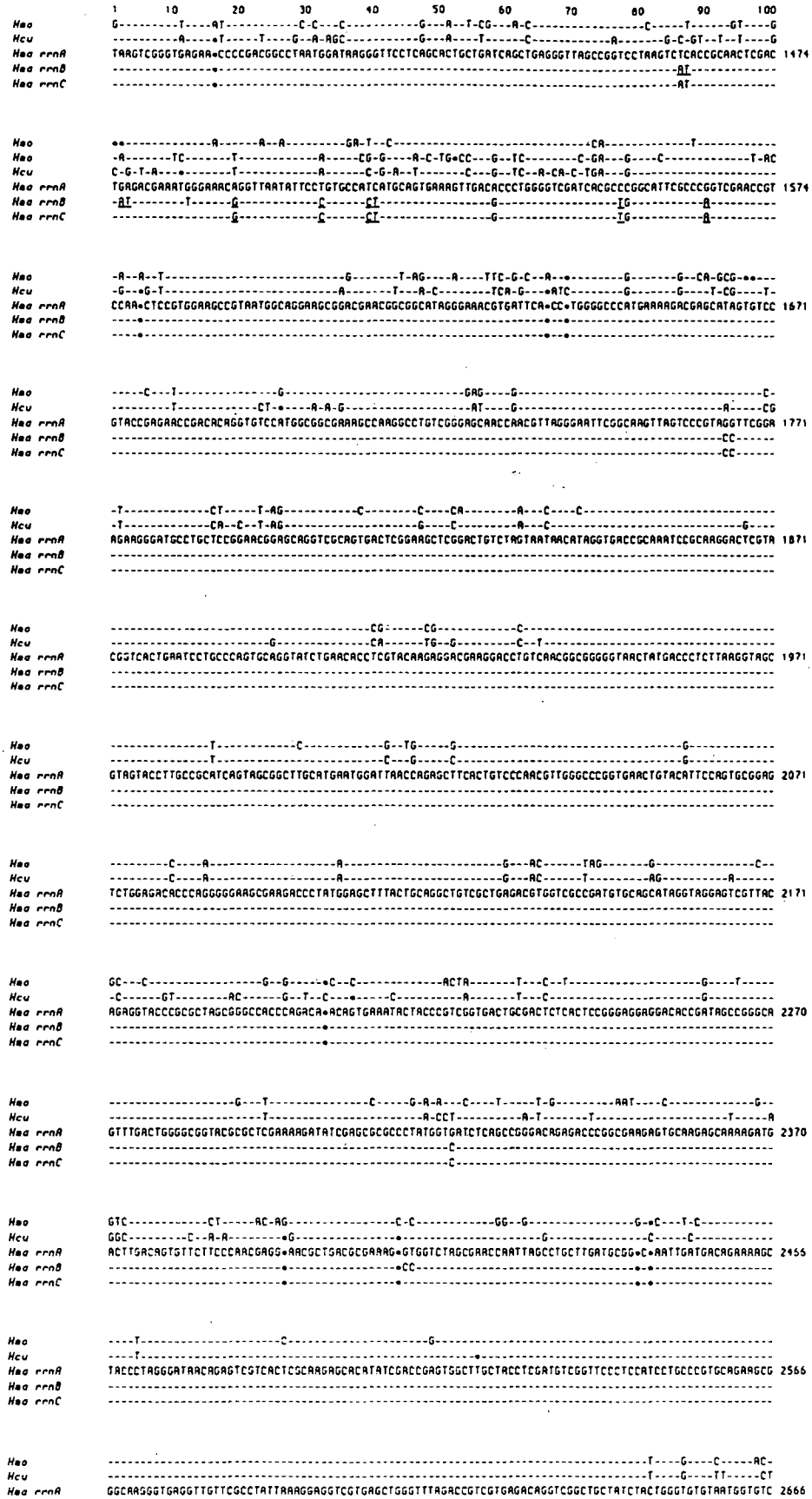

Figure 3.24 Figure caption on page 114

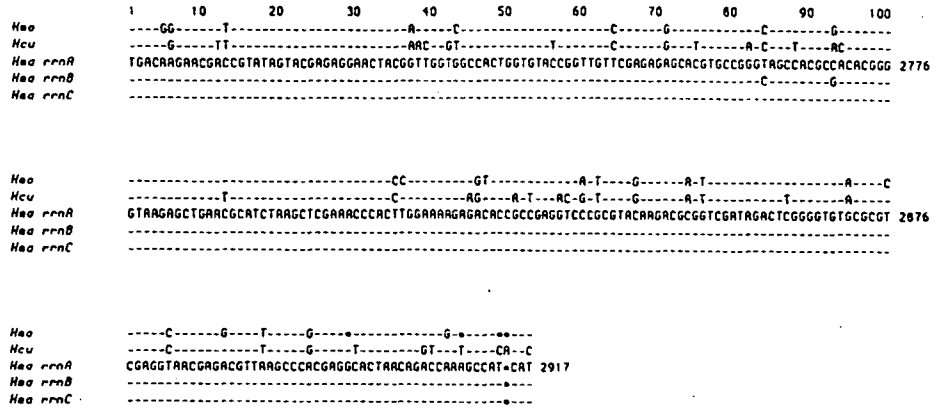

Figure **3.24** Nucleotide Sequences and Alignment of halophilic 23S rRNA encoding genes. The complete nucleotide sequence of the *Ha. marismortui* 23S encoding gene from the *rrnA* operon is depicted (*Hma rrnA*). The sequence from Brombach *et al.*, (1989) is assumed to be from the *rrnC* operon. The 23S rRNA encoding sequences from the *rrnB* and *rrnC* operons are aligned below the *rrnA* sequence; only the nucleotides that differ from the *rrnA* sequence are indicated. Substitutions that are compensatory, affecting both components of a base pair in regions of RNA secondary structure are underlined. For comparison, the entire 23S rRNA sequences from *Hb. cutirubrum* (*Hcu*) and *Hc. morrhuae* (*Hmo*) are also included. Dashes (-) indicate nucleotides identical to the *rrnA* sequence; dots (•) indicate single nucleotide gaps in the sequence(s) required to maintain alignment.

Pairwise analysis of the five sequences in this alignment is summarized in Table **3.4**. The distribution of nucleotide substitutions along the length of the 23S rRNA genes from the *rrnA*, *rrnB* and *rrnC* operons of *Ha. marismortui* and the operons from related halophiles *Hb. cutirubrum* and *Hc. morrhuae* is illustrated by the histograms in Figure **3.25**.

Comparisons of the 23S rRNA gene sequences from the *rrnA* and *rrnB* operons of *Ha. marismortui* indicate that the sequences are 98.7% identical where the differences are concentrated in three intervals defined by positions 72 to 363, 984 to 1765 and 2322 to 2769. The *rrnA-rrnC* comparison indicates that the sequences are 98.8% identical and the differences

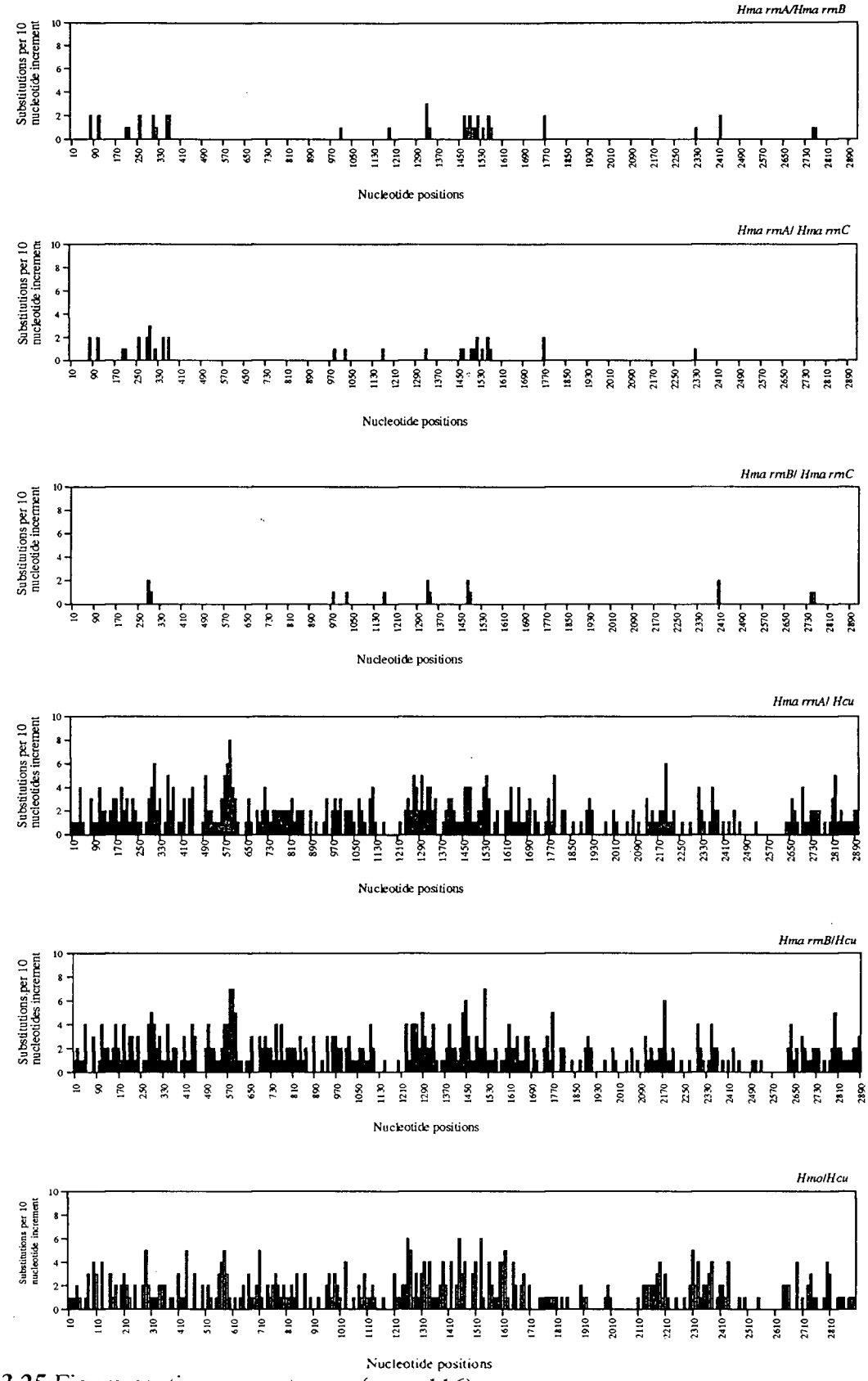

Figure 3.25 Figure caption on next page (page 116).

Figure **3.25** Histogram showing the distribution of nucleotide substitutions within the 23S rRNAs from *Ha. marismortui*, *Hb. cutirubrum* and *Hc. morrhuae*. The number of nucleotide substitutions for every 10 nucleotide increments was determined for six pairwise sequence comparisons; the two sequences in each pair are indicated on the upper right hand corner. The number of nucleotide substitutions were plotted against the position of the nucleotide sequences and the level of heterogeneity is indicated by the height of the vertical bars.

Table **3.4** A comparison of the nucleotide sequences of the 23S rRNA from halophilic archaeal species. The abbreviations used are as follows: L is the length of the nucleotide sequences compared; Ts is the number of transition substitutions; Tv is the number of transversion substitutions; Total, is the number of total substitutions.

Comparisons	Nucleotide Differences				Percentage similarities
	L	Ts	Tv	Total	
Complete 23 S sequence					
rrnA/ rrnB	2917	27	12	39	98.7%
rrnA/ rrnC	2917	24	10	34	98.8%
rrnB/ rrnC	2917	10	2	12	99.6%
rrnA/ Hcu	2903	288	153	441	84.8%
rrnB/ Hcu	2903	283	155	438	84.9%
rrnA/ Hmo	2896	235	152	387	86.6%
rrnB/ Hmo	2896	238	155	393	86.4%
Hcu / rrnC	2903	285	155	440	84.8%
Hmo/ rrnC	2903	240	155	395	86.4%
Hcu / Hmo	2890	275	120	395	86.3%

are concentrated within regions defined by positions 72 to 363, 984 to 1765 and position 2322. The sequences of *rrnB* and *rrnC* are almost identical (99.6% identical); the 15 nucleotide differences are located in regions defined by nucleotide positions 287 to 295, 984 to 1486 and 2413 to 2769. These results indicate that the distribution of nucleotide substitutions in the

pairwise comparison of the 23S rRNA sequences from the *rrnA*, *rrnB* and *rrnC* operons are similar to the comparison of the *rrnA* and *rrnB* 16S sequences, in that the differences are not randomly distributed but rather are concentrated in defined intervals.

Comparison of the *rrnA* and *rrnB* 23S sequences with the *rrnC* sequence indicates that the *rrnC* sequence contains 29 *rrnA* specific substitutions (within nucleotide positions 984-1786 and 2413-2769), 11 *rrnB* specific substitutions (at position 2322 and within regions 72-363 and 1322-1564). Four substitutions (at positions 287, 288, 295 and 1163) are unique in that they are not present in *rrnA* or *rrnB*. These results indicate that in the *rrnC* sequence, except at position 2311, the *rrnB*-like substitutions are concentrated within the first 1786 nucleotides of the sequence and the *rrnA*-like substitutions are located only beyond position 984 of the *rrnC* 23S rRNA sequence. Both the *rrnA*- and *rrnB*-like substitutions are present between positions 984-1786. In addition analysis of the 23S flanking sequences of *rrnC* indicates that the 16S-23S spacer appears identical to *rrnB* whereas the 5S and distal regions appear identical to *rrnA*.

There are three possible explanations for the presence of *rrnA*- or *rrnB*-like sequences within the third (*rrnC*) 23S rRNA gene sequence determined by Brombach *et al.*, (1989). First, the sequence is from the third operon, *rrnC*. Second, the third sequence might be a hybrid of *rrnA* and *rrnB* operons produced by recombination within the identical domains of the rRNA gene sequences from the *rrnA* and *rrnB* operons and present in a small proportion of the bacterial population. The presence of four unique sites in the *rrnC* sequence might be sequencing errors. Third, the sequence obtained by Brombach *et al.*, (1989) might be a combination of the *rrnA*, *rrnB* and *rrnC* operon sequences. The procedure used by Brombach *et al.*, (1989) involved a DNA probe to identify the 23S rRNA gene (the probe binds at identical positions within the the 3'- end of the 23S rRNA genes of all three operons) and then the 23S rRNA gene was sequenced in sections by chromosomal walking. Since the precise strategies used in isolating the DNA fragments for subcloning or how the chromosomal walking was performed were not described, it is not clear if their clones may have been derived from all three

operons. The resolution of this issue awaits the cloning and sequencing of the intact *rrnC* operon.

Pairwise comparisons of the 23S sequences from *Hb. cutirubrum* and *Hc. morhuae* to either one of the three *Ha. marismortui* sequences show that they are only 84.8-86.6% identical. Unlike the intraspecies comparisons (between the three 23S sequences from *Ha. marismortui*), the nucleotide differences in all interspecies pairwise alignments are not concentrated in specific domains but rather are more generally distributed throughout the entire 23S sequence (Figure **3.25**). These distribution patterns are similar to the interspecies comparison of the 16S sequences from halophiles. In both intra- and inter-species comparisons, transitions outnumber transversions by about two to one, and 60% of the substitutions are compensatory affecting both components of a nucleotide base pair in regions of RNA helical structure.

The intra-species nucleotide comparison between the 23S rRNAs of *rrnA-rrnB*, *rrnA-rrnC* and *rrnB-rrnC* sequences of *Ha. marismortui* gave 98.7%, 98.8% and 99.6% identical, respectively. These values are substantially less than the expected values (i.e., 0.1%) for paralogous small or large subunit rRNA genes within the genome of most organisms. However, based on the nucleotide similarity alone, it would appear that the divergence of the *rrnA*, *rrnB* and *rrnC* 23S rRNA sequences from each other is a recent event (i. e., in terms of evolutionary time). All pairwise interspecies comparisons show higher degrees of sequence divergence (between 86.6% and 84.8% identity) and suggest that the orthologous large subunit genes (23S rRNA) from the halophiles diverged from a common ancestral sequence within a short period of evolutionary time (about 600 million years ago). This situation is similar to the divergence of small subunit genes (16S rRNAs) from the same halophilic organisms (see section **3.2.4.1** above)

3.2.6.2 Secondary Structural Analysis

Pairwise comparisons between the 23S rRNA sequences from the *rrnA*, *rrnB* and *rrnC* operons of *Ha. marismortui* indicate that the nucleotide substitutions between each pair are

concentrated in four specific domains (I, III, IV and VI) of the universal secondary structural model for *E. coli* LSU rRNA (see Figure **1.10**; Noller *et al.*, 1981). Most of the substitutions are located within domain I (Figure **3.26**). Detailed analysis of the nucleotide substitution

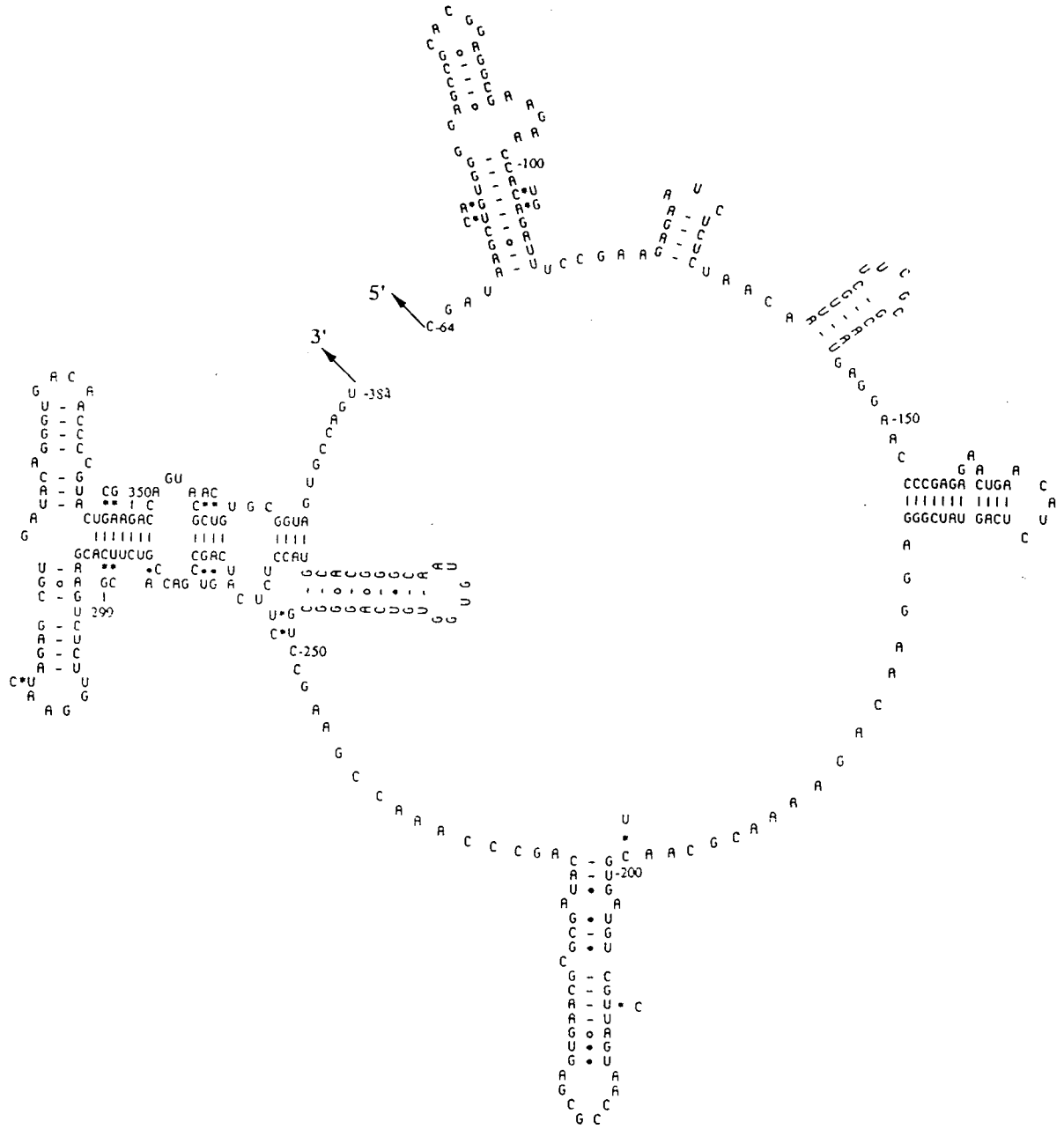

Figure **3.26** The predicted secondary structure of Domain I of 23S rRNA from *Ha. marismortui*. The secondary structure for Domain I of 23S rRNA from the *rrnA* operon, bounded by nucleotide positions 72 to 365, is illustrated. Mutational differences between *rrnA* and *rrnB* (*), *rrnA* or *rrnB* and *rrnC* (•) are indicated.

patterns between the 23S rRNAs from the *rrnA*, *rrnB* and *rrnC* operons using secondary structural models from *E. coli* and *Hb. cutirubrum* also indicate that most of the substitutions are compensatory and are located within the variable regions of the secondary structures. None of the substitutions observed between *rrnA* and *rrnB*, *rrnA* and *rrnC* or *rrnB* and *rrnC* would affect any known tertiary interactions, protein binding sites, or functional domains such as the peptidyl transferase center, the GTPase center or the α-sarcin loop (Leffers *et al.*, 1987; Raue' *et al.*, 1990).

3.2.7 23S Distal Region

3.2.7.1 23S-5S Intergenic Spacer

The 139 nucleotide long 23S-5S intergenic spacer sequences from the *rrnA*, *rrnB*, and *rrnC* operons are identical and are 27 nucleotides longer than the corresponding sequence of *Hb. cutirubrum* . Alignment of the 23S-5S spacer sequence from *Hb. cutirubrum* with any one of the three 23S-5S spacer sequences from *Ha. marismortui* indicate that they are 62.5% identical (Figure **3.27**). The alignment studies also showed that there were two regions of higher sequence conservation; the 23S processing stem region, and a 19 nucleotide long conserved sequence that forms helix G in the universal secondary structure for the primary transcript (Figure **3.7**). The sequences of the bulge motifs of the processing stem that are involved in the excision of 23S rRNAs in *Hb. cutirubrum* and *Ha. marismortui* are different; the former has TTT whereas the three operons of the latter are CAA. The complementary half of this inverted repeat is located in the 16S-23S intergenic spacer (see Figure **3.7**). The conservation of the sequences involve in the formation of helix G within genera of halophilic archaea (Figure **3.27, 3.7A, 3.7B** and **3.7C**) implies that this sequence is of crucial importance (Kjems and Garrett, 1990). The presence of helix G at the 5'-region of the 5S rRNA might be an important element for the processing of 5S rRNA.

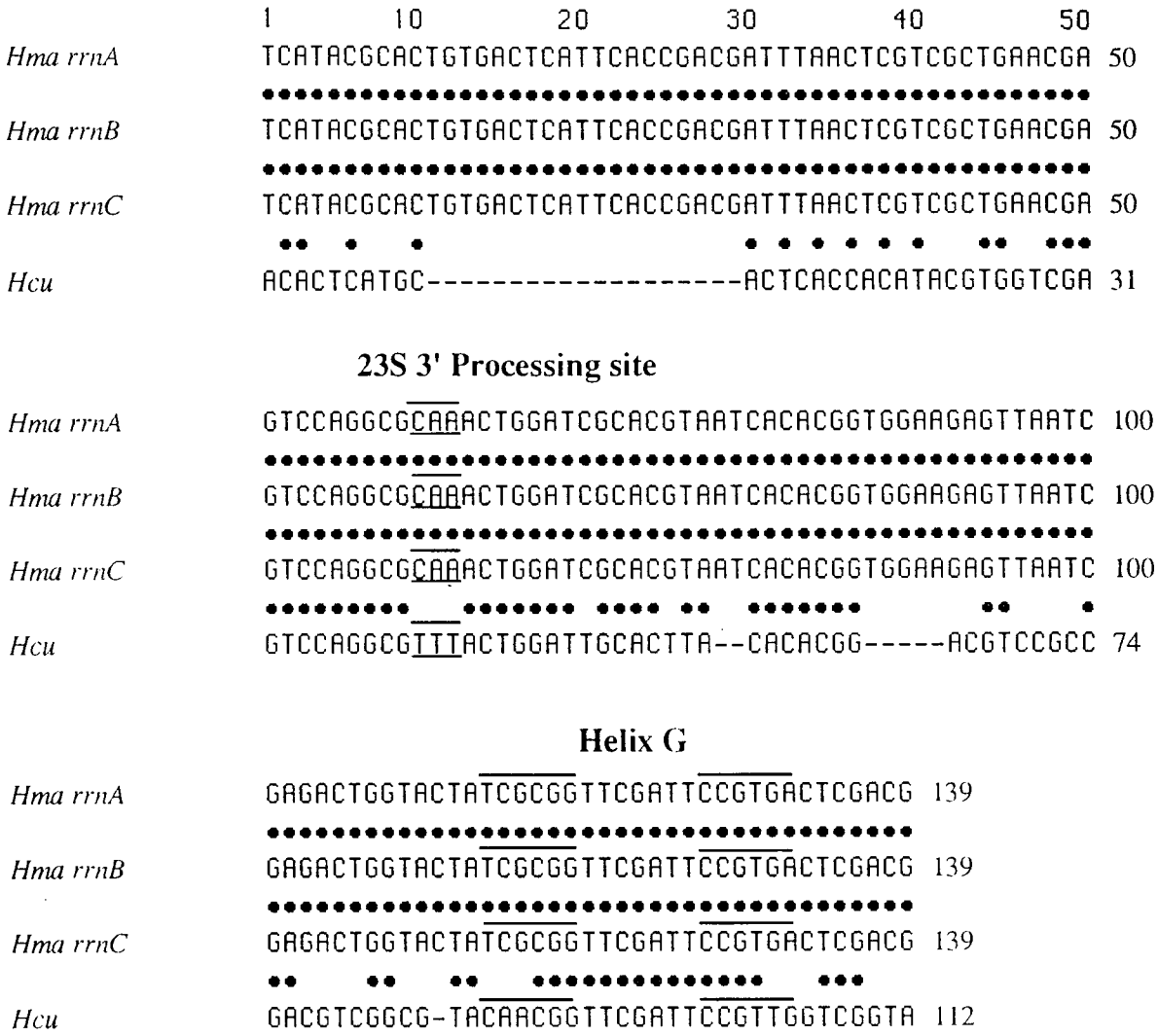

Figure **3.27** Comparison of the 23S-5S intergenic sequences. The sequences from *rrnA*, *rrnB*, and *rrnC* operons from *Ha. marismortui* and the single operon from *Hb. cutirubrum* are compared. The sequence similarity between two sequences are indicated by dots (•) and to maintain the alignment, gaps or deletions (-) were introduced within the sequences. The distal portion of the 23S processing repeat is underlined; excision occurs by cleavage within the three bulge nucleotides (not underlined). Helix G is overlined.

```
                1        10        20        30        40        50
Hma rrnA        TTAGGCGG••CACAGCGGTGGGGTTGCCTCCCGTACCCATCCCGAACACG

Hma rrnB        --------••----------------------------------------

Hma rrnC        --------••----------------------------------------

Hcu             --------GC--T-----------A-•----------•----------


Hma rrnA        GAAGATAAGCCCACCAGCGTTCCAGGGAGTACTGGAGTGCGCGAGCCTCT

Hma rrnB        ----------------------G-------------------------

Hma rrnC        ---------------------------------------------------

Hcu             ------------G--T-------G-TC----------------------


Hma rrnA        GGGAAATCCGGTTCGCCGCC•ACCA 124

Hma rrnB        ------C-------------•----

Hma rrnC        --------------------•----

Hcu             --------------------T--T-
```

Figure 3.28 Comparison of the 5S rRNA genes. The 5S sequences from the *rrnA*, *rrnB* and *rrnC* operons of *Ha. marismortui* are compared with the 5S sequence from *Hb. cutirubrum*. Only nucleotides that differ from the *rrnA* sequence are indicated. Dashes (--) indicate nucleotides identical to the *rrnA* sequence; dots (•) indicate single nucleotide gaps in the sequence(s) required to maintain alignment.

3.2.7.2 5S RNA

The 5S rRNA sequences from the *rrnA*, *rrnB* have been determined and are 120 nucleotides in length. A comparison of the *rrnA* and *rrnB* 5S sequences with the *rrnC* sequence (Brombach *et al.*, 1989) revealed that the 5S rRNA sequences from *rrnA* and *rrnC* are identical whereas the *rrnB* differs at two positions (Figure **3.28**). The 5S rRNA sequence of *Hb. cutirubrum* differs from the *Ha. marismortui* sequence at 10 of 120 nucleotide positions (91.7% similarity).

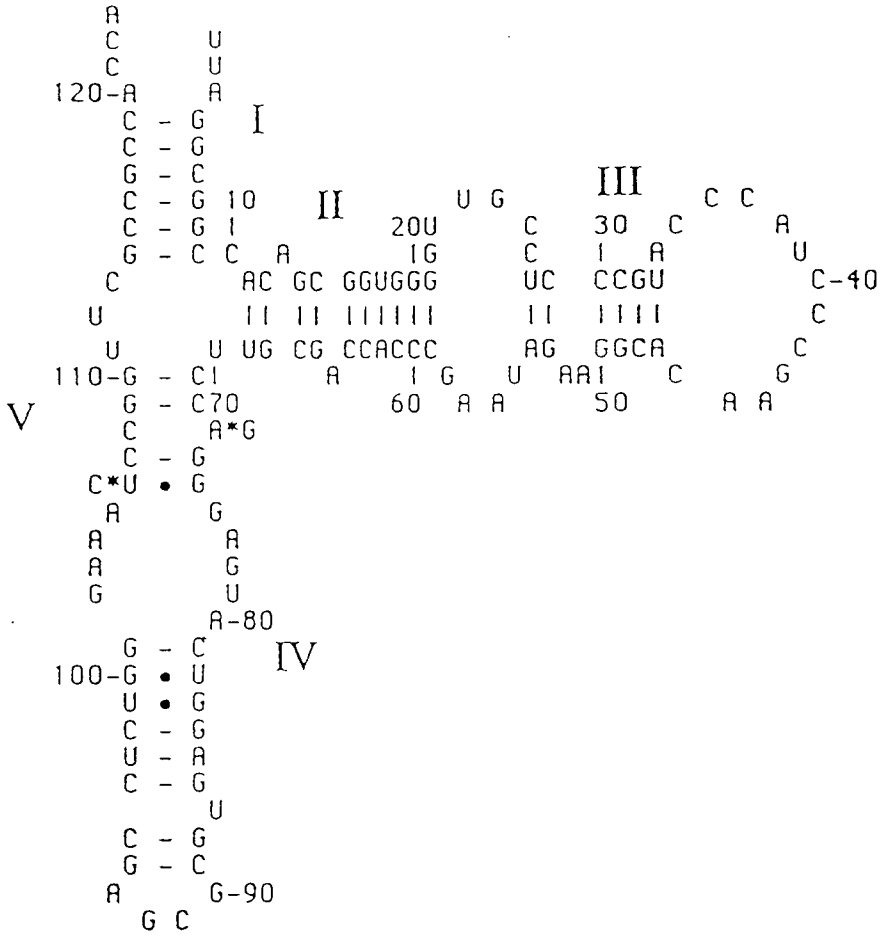

Figure 3.29 The predicted secondary structure for the 5S rRNA from the *rrnA* operon of *Ha. marismortui*. Mutational differences between *rrnA* and *rrnB* (*), normal Watson-Crick (-), and G • U base pairs are indicated. This secondary structure shows the characteristic features for the group II archaeal (representing extremely halophilic) 5S rRNAs (discussed in section **1.7.4**; Fox, 1985). The *rrnC* 5S sequence is identical to that of the *rrnA* 5S sequence.

All three 5S rRNA sequences from *Ha. marismortui* can conform to the expected group II archaeal (halophilic) structure (Figure **1.7**). The two nucleotide substitutions observed between the *rrnB* and *rrnA* (or *rrnC*) 5S sequences are located within the helix V region of the secondary structural model for the 5S rRNA (Figure **3.29**). One substitution, located at position 73 in the 5S rRNAs of the *rrnA* and *rrnC* operons, disrupts a G-C base pair (present in the *rrnB* operon) by a G->A substitution, however, the substitution at position 106 does not disrupt the helix structure (Figure **3.29**). These two substitutions are located within the regions

involved in the binding of L25 protein in *E. coli*. Comparison of the *rrnA*, *rrnB* and *rrnC* 5S rRNAs structures with the reported tertiary interaction sites in *Hb. cutirubrum* revealed that the two substitutions do not affect the nucleotides involved in any tertiary interactions. The characteristic structural features of group II 5S rRNAs include: no loop-out bases in helix I; an extended helix II; a helix III loop length of 13 nucleotides; a -CGAAC- sequence within the helix III loop; 16 nucleotides between the main portion of helix II and beginning of helix IV; and 21 nucleotides in the helix IV stem and loop (Fox, 1985 and Nazar, 1991).

3.2.7.3 The 5S Distal Region

With the exception of the first nucleotide, the sequences of the *rrnA*, *rrnB* and *rrnC* operons are identical for the first 61 nucleotides beyond the 5S gene sequence. At this point, the *rrnB* operon sequence diverges abruptly and exhibits no further similarity to the *rrnA* and *rrnC* sequences (Figures **3.6**, **3.30**). The perfect identity between *rrnA* and *rrnC* continues to position 200 (the *EcoRI* site); there is no sequence available beyond this point for the *rrnC* operon (see section **3.2.2.1**). The 5S-tRNA^{Cys} intergenic spacer in the *rrnA* operon is 239 nucleotides in length. Sequence and Southern hybridization analysis indicate that there is no tRNA^{Cys} gene in the distal region of *rrnB* (Figure **3.4**). The 5S distal region of *Hb. cutirubrum* contains a tRNA^{Cys} gene, 110 nucleotides downstream of the 3'-end of the 5S rRNA gene (Hui and Dennis, 1985). The tRNA^{Cys} sequences from the two organisms are 97.4% identical, whereas, there is very little homology in the immediate 5'-and 3'-flanking sequences.

Predicted secondary structures for the primary transcripts of the *rrnA*, *rrnB*, *rrnC* and that of *Hb. cutirubrum* indicated that helix H structures present in the 5S distal regions, are similar to rho-independent termination signals in *E. coli* (see section 1.6.3). The helix H structures from the *rrnA*, *rrnB* and *rrnC* operons are identical and they are followed by T-rich sequences (Figures **3.7A**, **B** and **C**). The helix H sequence of *Hb. cutirubrum* is also followed by T-rich sequences; however, the primary and secondary structures comprising helix H are not identical to those from the *rrnA*, *rrnB* or *rrnC* operons. In the *rrnA* operon, helix I structure is

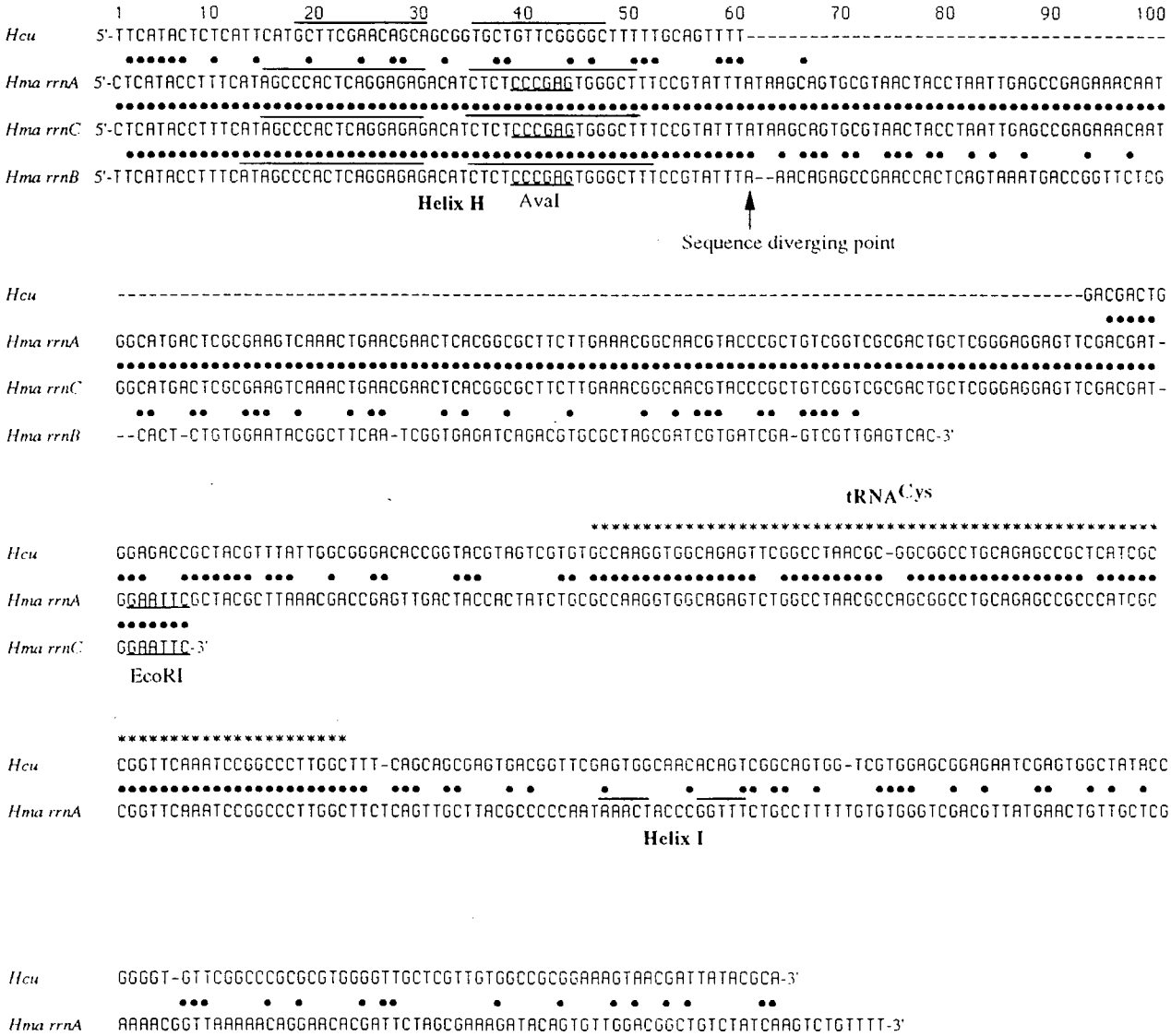

Figure **3.30** Comparison of the 5S distal regions. The 5S distal regions from the *rrnA*, *rrnB*, and *rrnC* operons of *Ha. marismortui* and the single operon from *Hb. cutirubrum*. The lengths of the *rrnA*, *rrnB*, *rrnC* and *Hb. cutirubrum* are 480, 168, 200 and 324 nucleotides, respectively. The nucleotides that are identical between two sequences are indicated by dots (•). The tRNA^{Cys} genes from the *rrnA* and *Hb. cutirubrum* are also indicated (***). Helices H and I and underlined

present within the 3'-flanking region of the tRNA^{Cys}. This structure (helix I) also followed by a T-rich sequence, acts as a termination site in the *rrnA* operon.

3.2.7.4 Primary Transcript Analysis of the 23S Distal Region

Using nuclease S1 digestion experiments, the regions downstream to the 23S rRNAs of the *rrnA* and *rrnB* operons were examined to locate the 3'-processing sites, 3'-maturation sites, 5S maturation sites and termination signals. A 365 nucleotide long *AvaI-AvaI* fragment from the *rrnA* operon was used as a probe for the analysis of the processing and maturation sites of the 23S and 5S rRNAs from the *rrnA* and *rrnB* transcripts. This fragment contains two substitutions (between *rrnA* and *rrnB*)within the 5S genes and one substitution within the 5S distal region, however, none of these single nucleotide substitutions appears to affect the analysis. The 365 nucleotide fragment isolated from the *rrnA* operon contains the last 66-bp of the 23S gene, the 139 nucleotide long 23S-5S spacer, the 120 nucleotide long 5S rRNA gene and a 40 nucleotide long 5S distal region. The fragment was 3'-labelled with α^{32}P dTTP on the minus strand, at position 2852, within the 23S gene. The labelled fragment was hybridized to total *Ha. marismortui* RNA, treated with S1 nuclease and the reaction products were separated on an 8% polyacrylamide gel along with the probe and size markers. In addition to full length protection, a number of partial protection products with nucleotide lengths of approximately 65, 127, 204 and 328 nucleotides were observed (Figure **3.31 A**). These products correspond respectively to protection by RNAs with 3'- ends generated by cleavage of the primary transcripts of the *rrnA* and *rrnB* operons at or near (i) the 3'-end of mature 23S rRNAs at position 2917 of the 23S rRNA genes, (ii) the precursor 23S rRNA processing sites at position 62 within the 23S-5S spacer regions, (iii) the 5'- end of the 5S rRNA gene at position 1 of the 5S rRNA genes, and (iv) the 3'-end of the 5S rRNA genes at position 124 of the 5S rRNA genes. Multiple bands below the 204 band may be due to imprecise 5S processing exonuclease activity at the 3'-transcript end or S1 nibbling at the end of the RNA-DNA hybrid. It is also possible that the 5S rRNAs may be excised as a precursor with a few extra nucleotide at the 5'- or 3'- ends and then trimmed to mature length.

The presence of a full length product from the 368 nucleotide probe suggests that the transcripts exiting the 23S gene read through the distal 5S gene. This probe does not include the

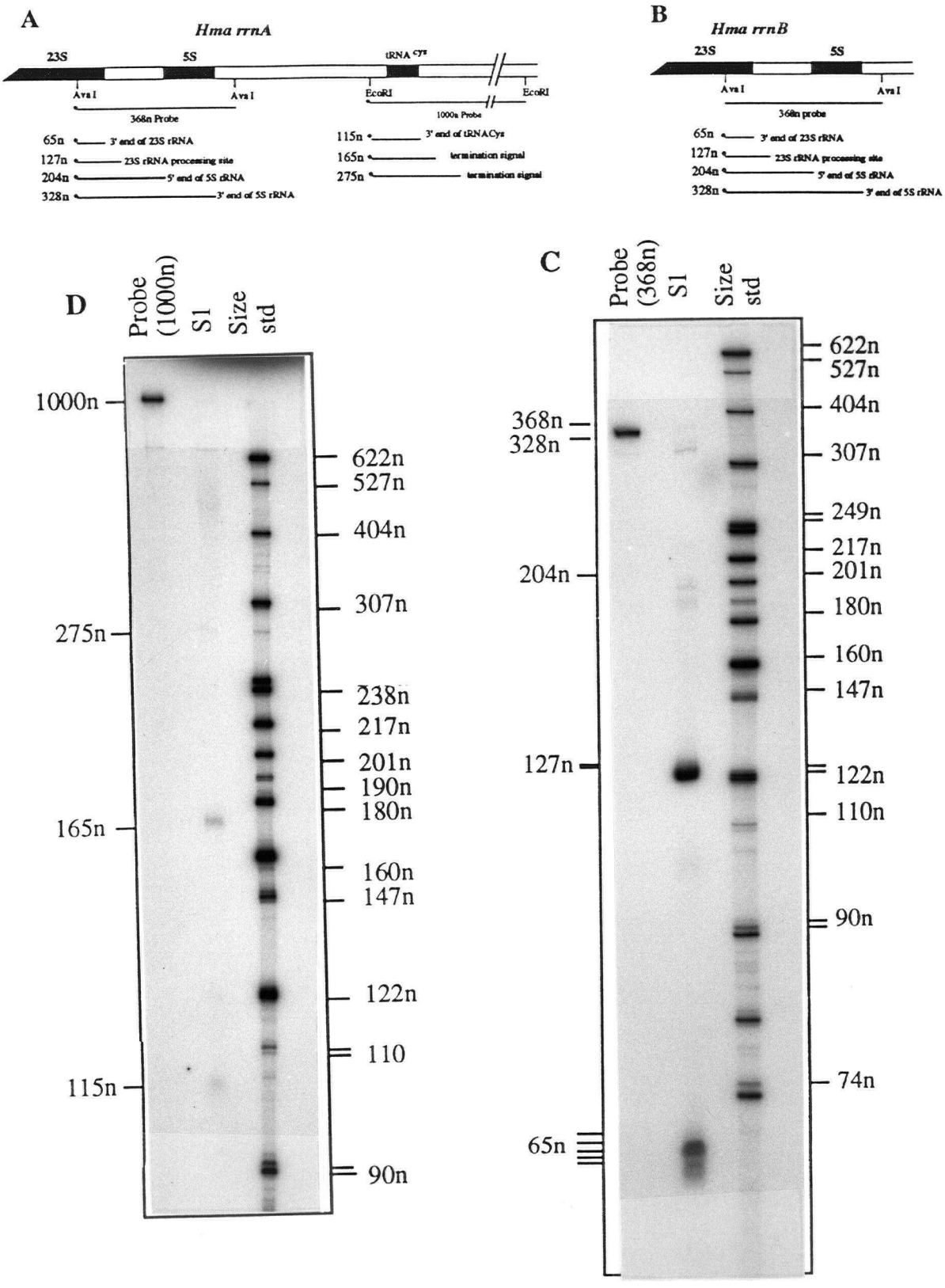

Figure **3.31** Figure caption on next page (page 128).

Figure **3.31** Nuclease S1 mapping analysis of the 5S distal regions from the *rrnA* and *rrnB* operons.

Figure **3.31A** The line diagram showing the 23S distal region of the *rrnA* operon. The 368 nucleotide long *AvaI-AvaI* fragment was used for the S1 nuclease analysis of the 23S distal regions of *rrnA* and *rrnB* transcripts. The lengths of the partial protection products and their corresponding cleavage positions are indicated below. The *AvaI* sites within the 23S rRNA and 5S distal regions are at identical positions in both *rrnA* and *rrnB* operons. The 368 nucleotide *Ava I-AvaI* fragment from the *rrnA* operon was 3'-labelled at the minus strand with α^{32}P dCTP and used as a probe for the S1 nuclease analysis (Figure **3.31C**). A second, 1000 nucleotide long *EcoRI-EcoRI* fragment, containing the $tRNA^{Cys}$ gene, was also isolated from the 5S distal region of the *rrnA* operon. This fragment was 3'-labelled at the minus strand with α^{32}P dATP and was used to study the processing sites from the $tRNA^{Cys}$ and termination sites downstream of the $tRNA^{Cys}$ gene. (Figure **3.31D**). The lengths of the partial protection products and their corresponding cleavage positions are indicated below.

Figure **3.31B** The line diagram showing the 23S distal region of the *rrnB* operon. The two *AvaI* sites shown in the diagram are located at position identical to the *rrnA* operon and this 368 nucleotide long *AvaI-AvaI* fragment from the *rrnB* operon is nearly identical to the *rrnA* operon (with only 3 substitutions). The partial protection products and their corresponding cleavage positions are shown below.

Figure **3.31C** Autoradiogram showing the nuclease S1 protection assay products of the 23S distal regions from the *rrnA* and *rrnB* operons. Above each lane, the probe, S1 products, and size markers are indicated. The sizes of the markers and the S1 products are indicated on the right and left sides, respectively.

Figure **3.31D** Autoradiogram showing the nuclease S1 protection assay products of the 5S distal regions from the *rrnA* operon. Above each lane, the probe, S1 products, and size markers are indicated. The sizes of the markers and the S1 products are indicated on the right and left sides, respectively.

poly pyrimidine sequences located about 50-60 nucleotides distal to the 5S gene of the *rrnA* and *rrnB* operons or the position at which the two sequences start to diverge. Therefore, these sites have not been studied.

Since S1 nuclease analysis are only semi–quantitative, the band intensities do not reflect the exact amount of RNAs present in the cells. These analysis are mainly done in order to detect the transcription initiation, termination or processing sites of the transcripts. In the case of rRNA transcripts, usually we expect about 1 - 2% of rRNA in the form of very long precursors and processing intermediates; most of the rRNA is in mature form (King and Schlessinger, 1983). However, this is not what has been observed from the S1 nuclease analysis shown in Figure **3.31C**; the intensities of the band representing the mature 23S rRNAs (64 nucleotides in length) are much less than expected. There are several explanations for this. First, the efficiency of hybridization between rRNA and the DNA probe may depend on the lengths of the rRNAs; longer RNAs hybridize more efficiently than the shorter ones and during a three hour hybridization, displacement may occur. Second, the conditions used in the assay (especially temperature) permit longer RNAs to hybridize more efficiently. Third, the majority of the mature rRNAs are folded in the form of a secondary structure and may not be readily available to hybridize with the DNA probe.

Another S1 mapping analysis was performed to investigate the processing sites of the $tRNA^{Cys}$ and termination signals within the 5S distal region of the *rrnA* operon. The 1000 nucleotide long *EcoRI-EcoRI* fragment contains a 45 nucleotide region upstream of the 5'-end of the $tRNA^{Cys}$ gene, the 74 nucleotide long $tRNA^{Cys}$ gene and about 881 nucleotides beyond the $tRNA^{Cys}$ gene. This fragment was 3'-labelled at position 203 distal to the 5S sequence, hybridized to total *Ha. marismortui* RNA, treated with S1 nuclease and separated on an 8% polyacrylamide gel along with the probe and size markers (Figure **3.31B**). Three partial protection products of approximately 115, 165 and 280 nucleotides in length were observed. These correspond respectively to protection by RNAs with 3'-ends (i) at the 3'-end of the $tRNA^{Cys}$ sequence at position 321; (ii) at the putative transcription termination sites near position 368 distal to the 5S sequence; and (iii) at a second transcription termination site near position 473 distal to the 5S gene. By carrying out an S1 nuclease analysis using the 1000 nucleotides long *EcoRI-EcoRI* fragment, along with Maxam and Gilbert sequencing, it was

possible to show the 3'-end of the $tRNA^{Cys}$ at nucleotide level (data not shown). The two larger protection products are assumed to result from protection by terminated transcripts. Alternatively, one of them might represent the site of sequence divergence between the *rrnA* and *rrnC* operons.

A number of conclusions concerning the transcription and processing in the 23S distal region can be made from these S1 nuclease analyses. In the *rrnA* operon, the observation that only the 3'-end of the $tRNA^{Cys}$ was apparent and not the 5'-end suggests that the transcript exiting the 5S gene reads through the $tRNA^{Cys}$ gene and terminates downstream of the gene. It also suggests that the processing of the 3'-end of $tRNA^{Cys}$ occurs before that of the 5'-end. In the case of *rrnA* operon, transcription appears to terminate at T-rich residues near positions 368 and 473 distal to the $tRNA^{Cys}$ sequence which, are proceeded by a helix structure (position 368 is located near helix I in Figure **3.7A**). This structure resembles the *rho*-independent termination signals in *E. coli*. The precise role of these conserved secondary structural features and their relationship to transcription termination remains to be established.

3.3 SUMMARY

Ha. marismortui contains two (or possibly three) rRNA operons in its genome. The *rrnA* and *rrnB* operons were initially identified and isolated by Mevarech *et al*, (1989) and cloned into plasmid pBR322. These two clones were fully characterized in this study. In addition, genomic hybridization analysis, using different restriction enzymes suggests that a third operon, designated *rrnC*, might be present. A 23S rRNA gene and flanking sequences from *Ha. marismortui* was published several years ago by Brombach *et al*, (1989). This sequence is different from the 23S rRNA sequences of the *rrnA* and *rrnB* operons determined in this study. Here it is assumed that the sequence published by Brombach *et al*, (1989) corresponds to the *rrnC* 23S rRNA gene and flanking regions, although it may be an artifactual composite of the *rrnA* and *rrnB* and *rrnC* sequences(see section **3.2.6.1** above). Clarification of this issue will require cloning and characterization of the complete *rrnC* operon.

Sequence and Southern hybridization analysis show that the gene order of the *rrnA* operon is 5'-16S-tRNA^{Ala}-23S-5S-tRNA^{Cys}-3' and the gene order of the *rrnB* operon is 5'-16S-23S-5S-3'. Comparison of the nucleotide sequences from the two operons shows that most of the nucleotide substitutions are located within the noncoding flanking and spacer regions, although long portions of the 5'-flanking, the 16S-23S spacer and 5S distal regions are virtually identical. It is possible that these conserved regions contain important motifs used for the regulation or processing of the RNA products. The two putative primary transcripts can be folded to form secondary structures, containing processing stems that surround the 16S and 23S rRNAs as well as seven other conserved regions of potential secondary structures (designated as helices A to G; Kjems and Garrett 1990).

The 16S 5'-flanking sequences of *rrnA* and *rrnB* differ in a number of substantive ways. The *rrnA* operon contains four promoters; primary transcript analysis show that all are active in transcript initiation. The *rrnB* operon contains two promoter-like sequences but only one has been shown to be active. In the *rrnA* transcript, the processing helices surrounding the 16S and 23S gene sequences contain the consensus "bulge-helix-bulge" motif that is recognized by an endonuclease and used to excise pre 16S and pre 23S from the primary transcripts. The transcript from *rrnB* contain the consensus 23S processing motif but the 16S stem lacks the "bulge-helix-bulge" motif and processing presumably proceeds by a different pathway.

The 16S rRNAs from the *rrnA* and *rrnB* operons differ at 74 positions of the 1472 nucleotide sequence. The substitutions are not spread throughout the molecule but rather are confined within three different domains, defined by nucleotide positions 58-321, 508-823, and 986-1158. About two thirds of the substitutions are located within the 508-823 domain and most of them are compensatory, affecting both components of a nucleotide base pair within the molecule. None of the substitutions is believed to affect functionally important nucleotide positions. Using stringent and non stringent nuclease S1 protection assays, it was shown that both of the 16S rRNAs are expressed and present in the active 70S ribosomes.

Within the 16S-23S spacer region, the *rrnA* operon contains a $tRNA^{Ala}$ gene whereas the *rrnB* operon does not. Three unique processing sites were identified within the spacer region of the *rrnB* operon which are probably important for 16S excision and maturation. The intergenic spacers in both operons contains an active internal promoter. This promoter is located at the beginning of a 177 nucleotide long region immediately preceding the 23S genes, which is identical in the two operons.

The pairwise comparison of the 2917 nucleotide long 23S rRNA sequences showed 39 differences between *rrnA* and *rrnB*, 28 between *rrnA* and *rrnC*, and 16 between *rrnB* and *rrnC*. Most of the differences are compensatory. None of the substitutions appear to affect functionally important positions. The presence of identical flanking sequences suggests that excision and processing of 23S rRNA from the three operons proceeds by a single common pathway. The intermediates in this pathway were identified using S1 nuclease protection assays.

The 139 nucleotide long 23S-5S spacer regions are identical in the *rrnA*, *rrnB* and *rrnC* operons. The 5S gene from *rrnB* differs from the *rrnA* and *rrnC* sequence at two positions, but both positions appear to be functionally unimportant. The *rrnA* operon contains a $tRNA^{Cys}$ gene located 239 nucleotides beyond the 5S gene. Transcripts that extend through the tRNA appear to terminate at T-rich pyrimidine tracts located in the 3'-flanking region. No termination signals have been identified for the *rrnB* operon.

CHAPTER 4

Phylogenetic Implications of Sequence Diversity Between the Two Ribosomal RNA operons, *rrnA* and *rrnB* from the *Haloarcula marismortui*.

4.1 Introduction

Studies on ribosomal RNAs from the *rrnA* and *rrnB* operons of *Ha. marismortuii* are of particular interest because this is the first archaeal organism in which rRNA sequence heterogeneity has been documented. It is expected that some of this sequence will help us understand the molecular adaptation of halophilic organisms to an extreme saline environment. These studies will also provide information for establishing the phylogenetic relationship between *Ha. marismortui* rRNA operons and the other related halophilic archaeal organisms. By using the rRNA gene sequence data from two non-identical rRNA operons of *Ha. marismortui*, the phylogenetic implications of sequence diversity between the two operons and how they are related to other halophiles can be studied.

To investigate the phylogenetic implications, the sequence divergences from the *rrnA* and *rrnB* operons of *Ha. marismortui*, the homologous sequences from related halophiles, and sequences from two outgroup species, were aligned. Four different comparisons were carried out for this phylogenetic analysis; the 508-823 domain of 16S rRNA, the entire 16S rRNA, the entire 23S rRNA and the combined 16S-23-5S rRNA sequences. In principle, the trees generated from each of these sequences should be congruent and should reflect the phylogenetic relationships between the organisms from which the sequences were obtained. In practice, the trees are essentially congruent except for the 508-823 domain tree.

4.2 Results and Discussion

4.2.1 Phylogeny of the 508-823 Domains of the 16S rRNA Genes

The 508-823 domains of the 16S rRNA genes from five halophiles, *Ha. marismortui* (both the *rrnA* and *rrnB* operons), *Hf. volcanii*, *Hb. cutirubrum*, and *Hc. morrhua*, and two out-group species, *M. vanielli* and *E. coli*, were used in this study. The tree obtained is shown in Figure **4.1A**. It indicates that all the halophiles grouped together with a consistency value of 1.00. In contrast to expectation, the *rrnA* and *rrnB* 508-823 domains do not group together; the *rrnA* branches with *Hb. cutirubrum* and *Hc. morrhuae*, whereas *rrnB* branches before *Hf. volcanii*. It is also interesting to point out that the *Hb. cutirubrum* and *Hc. morrhuae* prefer high salt concentration (3-5 M) whereas *Hf. volcanii* prefers moderate salt concentration (2.14 M).

Salinity of the hypersaline lakes (e.g. the Dead Sea) can vary due to seasonal changes (wet or dry periods; Edgerton and Brimacombe, 1981). Organisms which live in this environment must be able to adapt to fluctuating salinity conditions. *Ha. marismortui* has been able to adapt to a variety of lake with different salt compositions. The mutational changes observed within the 508-823 domain of the 16S rRNA genes from the *rrnA* and *rrnB* operons may be important for the organism to survive in different saline environment. In this scenario, the *rrnB* operon, because of its similarity to *Hf. volcanii*, is used to survive in moderate salt concentration whereas the *rrnA* operon, because of its similarity to *Hb. cutirubum* and *Hc. morrhuae*, is used to survive in high salt concentration.

4.2.2 Phylogeny of the Complete 16S rRNA Genes

The entire 16S rRNA sequences were used to analyse the phylogenetic implications of the sequence divergence observed between the *rrnA* and *rrnB* operons of *Ha. marismortui*. Again, the five halophilic 16S sequences form a coherent group with a consistency value of 1.00. In this tree, the *rrnA* and *rrnB* sequences form a coherent subgroup because the differences in the 508-823 region have been masked by the higher degree of similarity in the rest of the molecule. The *rrnB* sequence appears to be evolving more rapidly than the *rrnA* sequence. Presumably, this

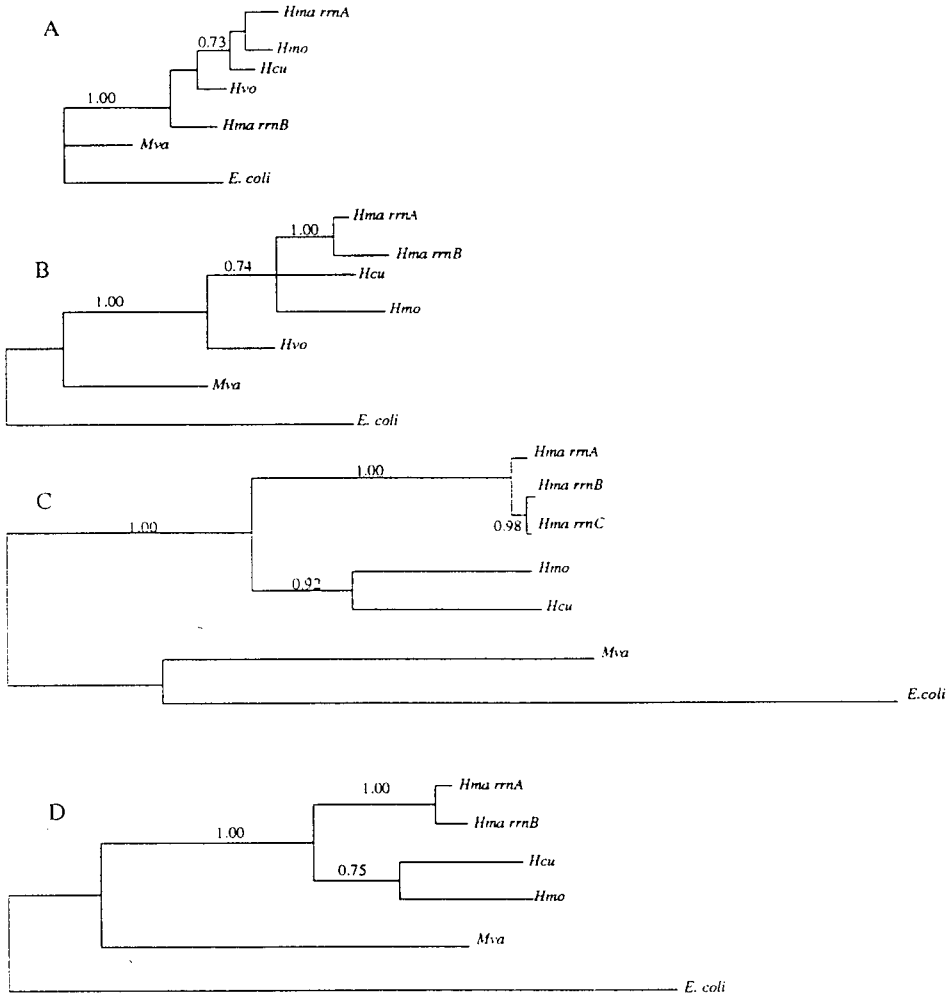

Figure **4.1** Phylogeny obtained from the 508-823 domain of the 16S rRNA genes (Tree A), the complete 16S rRNA genes (Tree B), the complete 23S rRNA genes (Tree C), and the combination of 16S-23S-5S rRNAs (Tree D). A parsimony analysis of aligned 16S and 23S sequences and 16S-23S-5S sequences was carried out using PAUP with the heuristic search option. The most parsimonious tree is illustrated along with the consistency index for each branch. The numbers preceding each branch are bootstrap consistency values and indicate the proportion of replications that group all the taxonomic units within the branch. The consistency index values for Trees A, B, C, and D are, 0.833, 0.824, 0.909, and 0.91, respectively. The branches with consistency values less than 0.5 are not indicated.

represents the accumulation of nucleotide substitutions predominantly in the 508-823 domain that make the domain more like that of *Hf. volcanii*. The two identical 16S genes from *Hf. volcanii* branch early and the single 16S genes from *Hb. cutirubrum* and *Hc. morrhuae* branch later from the lineage leading to the paralogous 16S rRNA genes from *Ha. marismortui*. It is not known whether the 16S gene from the *rrnC* operon is identical to one of the other two or represents a third type of 16S gene that had formed by a second paralogous sequence divergence event.

4.1.3 Phylogeny of the 23 rRNA Genes

The 23S rRNA gene sequence from *Hf. volcanii* is not available whereas a third 23S rRNA gene sequence (*rrnC*) from *Ha. marismortui* is available in this comparison. The available 23S gene sequences were analysed by PAUP and the resulting tree is illustrated in Figure **4.1C**. As with the 16S tree, the halophilic 23S sequences form a coherent group and the *rrnA*, *rrnB*, and *rrnC* sequences of *Ha. marismortui* form a coherent subgroup with *rrnB* and *rrnC* on one branch and *rrnA* on the other. The *Hb. cutirubrum* (*halobium*) and *Hc. morrhuae* are together on a separate branch.

4.1.4. Phylogeny of the 16S-23-5S rRNA Genes

For statistical reasons, the resolving power of the PAUP analysis is related to the length of the sequences under comparison. Therefore, the available 16S, 23S and 5S sequences were combined from the *rrnA* and *rrnB* operons of *Ha. marismortui* and the single operons from *Hb. cutirubrum* (*halobium*) and *Hc. morrhuae*. The outgroup sequences were again from *M. vanielli* and *E. coli*. The topology of this tree is identical to the 23S tree. Within the tree, the *Ha. marismortui* operons are branching later than the *Hb. cutirubrum* and *Hc. morrhuae* operons; this may indicate that the *Ha. marismortui* rRNA operons are evolving more rapidly than the *Hb. cutirubrum* and *Hc. morrhuae* operons. Within the two operons from *Ha. marismortui, rrnB* appears to be evolving more rapidly than the *rrnA*.

The genera of halophilic archaea considered here appear to have diverged from each other within a relatively short period of time (some 600 million years ago; Mylvaganam and Dennis,

1990; Ochman and Wilson, 1987). This calculation is based on the assumption that in archaea, the mean rate of substitution is about 1% per 50 Million years. Extensive analysis of both of the rRNA sequences and protein encoding sequences have failed to produce congruent and reliable trees describing precise phylogenetic relationships and branching orders. Nonetheless, the result presented here indicates that the *rrnA*, *rrnB* and *rrnC* operons of *Ha. marismortui*, in spite of their sequence divergence, are more closely related to each other than they are to the sequences from three other halophilic species. The only exception to this is the 508-823 regions of the *rrnA* and *rrnB* (see section **3.2.4.1**).

4.1.5. Evolution of Ribosomal RNA Operons in *Ha. marismortui.*

In the case of multicopy ribosomal RNA operons, it is highly unusual to observe nucleotide sequence differences above 0.1% within a genome of an organism. Sequence homogenization of these duplicated rRNA genes is believed to involve processes of gene conversion and unequal crossing over in a process termed concerted evolution (see section **1.7**; Li and Graur, 1991).

In the extreme halophilic archaeon *Ha. marismortui*, the 16S, 23S and 5S rRNAs representing the *rrnA* and *rrnB* operons are 95%, 98.7% and 98.4% identical. The number of nucleotide differences between these sequences are between 10 and 50 fold higher than what is normally observed for duplicated ribosomal RNA genes. There are at least three explanations or causes for this unexpectedly high level of nucleotide sequence divergence: (i) a failure in the recombination system used to maintain sequence homogeneity, (ii) fusion to produce a chimeric genome containing a divergent rRNA operon, and (iii) selection-induced divergence within rRNA operons.

The analysis of the positions of nucleotide difference between the 16S, 23S and 5S sequences from the *rrnA*, *rrnB* and *rrnC* operons indicate that all of the substitutions occur at positions that are normally phylogenetically variable. In other words, none of the substitutions affects positions known to be important in the essential structure or functions of the RNA in the

ribosome. It is therefore possible that the substitutions represent neutral mutations that have been fixed by random processes and that they have no physiological or biochemical significance. This does not explain why the differences are confined to certain regions and absent from other regions. However, one could imagine that gene conversion continues to operate over the regions of those genes that are homologous in sequence but no longer operates over the regions that have accumulated a high proportion of substitutions.

A second possible explanation is that there was a fusion between two organisms with divergent rRNA sequences and that their separate genomes were able to form a genetic chimera. One can imagine that gene conversion and recombination events might have homogenized parts of the ribosomal RNA operons whereas other parts have remained distinct. There is no direct evidence or precedent to support this suggestion. By sequencing the genome of *Ha. marismortui* (or the regions that are associated with the rRNA operons), may be possible to identify more examples of duplicate genes with high sequence divergence. If there are such genes, this would support the theory of a cell fusion event.

A third possible explanation is that *Ha. marismortui* duplicated (into *rrnA* and *rrnB*) or triplicated (if *rrnC* operon exists) its rRNA operons and suppressed homogenization in order to allowed each operon to evolve independently. It is known that halophilic archaea balance the intracellular salinity with that of the external environment (which can vary between 1.5-5M NaCl; Christian and Waltho, 1962; Ginzburg et al., 1970; Lanyi and Silverman, 1972). A single homogeneous ribosome population likely will not be able to maintain optimum function over this entire salinity range. It may be that the observed sequence divergence produces ribosome populations with different salt optima. So far, no significant difference in the level of expression of the *rrnA* and *rrnB* operons have been detected under standard laboratory conditions (see section **2.2.1**).

In summary, there clearly exists sequence heterogeneity between the three rRNA operons of *Ha. marismortui*. None of the differences appear to affect residues known to be critical to rRNA structure or function. At least two of the operons are known to be expressed under normal

laboratory conditions. Three possible explanations for this divergence have been suggested; none of the three are compelling or completely satisfying.

Future Research Prospects

The studies involved in this thesis demonstrate that there are at least two and possibly three ribosomal RNA operons (*rrnA*, *rrnB*, and possibly *rrnC*) in the genome of *Ha. marismortui*. Only *rrnA* and *rrnB* have been characterized extensively. In order to understand the features of the *rrnC* operon (whether it exists, whether it is *rrnA*-like, *rrnB*-like, a hybrid of *rrnA* and *rrnB* or unique), the operon must be cloned and its entire sequence must be determined. First, using different 16S and 23S rRNA probes, the existence of this operon in the genomic DNA of *Ha. marismortui* must be determined. If exists, cloning and characterization of the entire *rrnC* operon must be performed in order to understand the origin and propagation of the rRNA operons from *Ha. marismortui*. In order to determine whether the rRNA genes from the *rrnC* operon is active or not and to understand the metabolic pathways involved in the formation of functional rRNA molecules, Nuclease S1 protection assays must be performed (see section **2.2.16**).

It is known that halophilic archaea balance the salinity of the intracellular medium with that of the external environment (which can vary between 1.5 - 5M NaCl; Christian and Waltho, 1962; Ginzburg et al., 1970; Lanyi and Silverman, 1972). It is possible that under different external environments the expression of rRNA genes can be switched between different rRNA operons. In order to determine whether these operons are expressed under different extreme conditions, such as those encountered in halophilic biotopes, the organism should be grown under different salt concentrations (NaCl, KCl and $MgCl_2$) and the expression of the *rrnA*, *rrnB* and if present, *rrnC* rRNA operons under these varying conditions could be studied by nuclease S1 protection assays and/or primer extension analysis. Since efficient transformation systems for *Haloarcula* species are now available (Cline and Doolittle, 1992), the shuttle vectors can be used to investigate the molecular biology of *Ha. marismortui*. One or two of the rRNA operons of *Ha. marismortui* can be altered or removed from the genome using these vectors and their effects on the survival of the organism can be studied. With the altered genomic DNA of *Ha. marismortui* (the presence of only one rRNA operon), the salt effects also can be studied.

REFERENCES

Achenbach-Richter, L., K. O. Stetter and C. R. Woese. 1987. A possible missing link among archaebacteria. *Nature*. **327**:348 - 349.

Achenbach-Richter, L. and C. R. Woese. 1988. The ribosomal gene spacer region in archaebacteria. *System. Appl. Microbiol.* **10**:211 - 214.

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. d. Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, P. H. Schreier, A. J. H. Smith, R. Staden and I. G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature*. **290**:457 - 464.

Atkinson, T. and M. Smith. 1984. Solid phase synthesis of oligoribonucleotides by the phosphite-triester method. *In* Oligonucleotide Synthesis. A Practical Approach. *Edited by* Gait. M. J., IRL Press, Oxford/Washington, pp. 35 - 81.

Bayley, S. T. and D. J. Kushner. 1964. The ribosomes of the extremely halophilic bacterium *Halobacterium cutirubrum*. *J. Mol. Biol.* **9**:654 - 669.

Bayley, S. T. and E. Griffiths. 1968. A cell-free amino acid incorporation system from an extremely halophilic bacterium. *Biochem.* **7**:2249 - 2256.

Berghofer, B., L. Krockel, C. Kortner, M. Truss, J. Schallenberg and A. Klein. 1988. Relatedness of archaebacterial RNA polymerase core subunits to their eubacterial and eukaryotic equivalents. *Nucleic Acids Res.* **16**:8113 - 8128.

Boone, D. R. and R. A. Mah. 1990. Family II. Methanosarcinaceae, genus. *Methanosarcina*. *In* Bergey's Manual. of Systematic Bacteriology. *Edited by* Staley, J. T., Brylant, M. P., Pfennig, N., and Holt, J. G., Williams and Wilkins, Baltimore, pp. 2198 - 2205.

Böck, A., H. Hummel, M. Jarsch and G. Wich. 1986. *In* Biology of Anaerobic Bacteria.

Edited by Dubourgiuer H. D. *et al.* Elsevier Science Publishers, Amsterdam, pp. 206 - 226.

Branlant, C. A., M. A. Krol, A. Machatt, J. Pouyet, J.-P. Ebel, K. Edwards and H. Kossel. 1981. *E.coli* 23S rRNA heterogeneity. *Nucleic Acids Res.* **9**:4303.

Brimacombe, R., B. Greur, P. Mitchell, M. Osswald, J. Rinkeappel, D. Schuler and K. Stude. 1990. Three-dimensional structure and Function of *Escherichia coli* 16S and 23S rRNA as studied by cross-linking techniques. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 93 - 106.

Brombach, M., T. Specht, V. A. Erdmann and N. Ulbrich. 1989. Complete nucleotide sequence of a 23S ribosomal RNA gene from *Halobacterium marismortui. Nucleic Acids. Res.* **17**:8.

Brosius, J., T. J. Dull, D. D. Sleeter and H. F. Noller. 1981. Gene organization and primary structure of a rRNA operon from *Escherichia coli. J. Mol. Biol.* **148**:107 - 127.

Brown, A. D. 1976. Microbial water stress. *Bact. Rev.* **40**:803 - 846.

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

Chang, S., A. Majumdar, R. Dunn, O. Makobe, U. R. Bhandary, H. G. Khorana, E. Ohtsuka, T. Tanaka, Y. Taniyama and M. Ikehara. 1981. Bacteriorhodopsin: partial sequence of mRNA provides amino acid sequence in the precursor region. *Proc. Natl. Acad. Sci. USA.* **78**:3398 - 3402.

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

Christian, J. H. B. and J. A. Waltho. 1962. Solute concentrations within cells of halophilic and non halophilic bacteria. *Biochem. Biophys. Acta.* **65**:506 - 508.

Christiansen, J. and R. A. Garrett. 1986. How do protein L18 and 5S RNA interact? *In* Structure, Function and Genetics of Ribosomes. *Edited by* Hardesty, B., and Kramer, G., Springerverlag, New York, pp. 733 - 748.

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

Cline, S. W. and W. F. Doolittle. 1992. Transformation of members of the genus *Haloarcula* with shuttle vectors based on *Halobium* and *Haloferax volcanii* plasmid replicons. *J. Bacteriol.* **174**:1076 - 1080.

Condon, C., S. French, C. Squires and C. L. Squires. 1993. Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased expression of the remaining intact copies. *EMBO J.* **12**: 4305 -4315.

Court, D. 1993. RNase III: a double strand RNA processing enzyme. *In* Control of mRNA Stability, *Edited by* Brawerman, G., and Belasco, J. New York: Academic Press.

Daniels, C. J., S. E. Douglas, H. Z. McKee and W. F. Doolittle. 1985. Archaebacterial tRNA genes: structure and intron processing. *In* Microbiology. Edited by L. Leive. American Society for Microbiology Publications, Washington, DC. pp. 349-355.

Darnell, J. E. and W. F. Doolittle. 1986. Speculations on the early course of evolution. *Proc. Natl. Acad. Sci. USA.* **83**:1271 - 1275.

Darr, S. C., J. W. Brown and N. R. Pace. 1992. The varieties of ribonuclease P. *TIBS.* **17**:178 - 182.

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

USA. **81**:125 - 129.

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

Dennis, P. P. 1985. Multiple promoters for the transcription of ribosomal RNA gene cluster in *Halobacterium cutirubrum. J. Mol. Biol.* **186**: 457 - 461.

Dennis, P. P. 1991. The ribosomal RNA operons of halophilic archaebacteria. *In* General and Applied Aspects of Halophilic Microorganisms. *Edited by* Rodriguez-Valera, F., Plenum Press, New York, pp. 251 - 257.

Dente, L. and R. Cortese. 1983. pEMBL: a new family of single–stranded plasmids for sequencing DNA. *Methods Enzymol.* **155**:111 - 118.

Doolittle, R. F. 1985. The geneology of some recently evolved vertebrate proteins. *Trends Biochem. Sci.* **10**: 233 -237.

Dover, G. A. 1982. Molecular drive: a cohesive mode of species evolution. *Nature*. **299**:111 - 117.

Dover, G. A. and R. B. Flavell. 1984. Molecular coevolution: DNA divergence and the maintanence of function. *Cell.* **38**:622 - 623.

Dover, G. A. 1987. DNA turnover and the molecular clock. *J. Mol. Evol.* **26**:47 - 58.

Dryden, S. and S. Kaplan. 1990. Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter* spheroides. *Nuclaic Acids Res.* **18**:7267 - 7277.

Dunn, T., S. Hahn, S. Ogden and R. Schleif. 1984. An operator at – 280 base pairs that is required for repression of *araBAD* operon promoter: addition of DNA helical turns between the operator and promoter cyclically hinders repression. *Proc. Natl.Acad. Sci. USA.* **81**:5017 - 5020.

Durovic, P. and P.P. Dennis. 1994. Separate pathways for excision and processing of 16S and 23S rRNA from the primary rRNA operon transcript from the hyperthermophilic archaebacterium Sulfolobus acidocaldarius: similarity to eukaryotic processing. *Mol. Microbiol.* **13**: 229-242

Edgerton, M. E. and P. Brimacombe. 1981. Thermodynamics of halobacterial environments. *Can. J. Microbiol.* **27**:899 - 909.

Ellwood, M. and M. Nomura. 1982. Chromosomal locations of the genes for rRNA in *Escherichia coli* K-12. *J. Bacteriol.* **149**:458 - 468.

Ernst, W. G. 1983. The early earth and the archaean rock record. *In* Earth's Earliest Biosphere: its origin and evolution. *Edited by* Schopf.J. W. University Press, Princeton, New Jersy, pp. 41 - 52

Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and reliability. *Annu. Rev. Genet.* **22**:521 - 565.

Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specified tree topology. *Syst. Zool.* **20**:406 - 416.

Fitch, W. M. and J. S. Farris. 1974. Evolutionary trees with minimum nucleotide replacements from amino acid sequences *J. Mol. Evol.* **3**:263 - 278.

Fitch, W. 1977. On the problem of generating the most parsimonious tree. *Am. Nat.* **111**:223 - 257.

Fitch, W. M. 1981. A non sequential method for constructing trees and hierarchial classifications. *J. Mol. Evol.* **18**:30 - 37.

Fox, G. E., E. Stakebrandt, R. b. Hespell, J. Gibson, J. Maniloff, T. A. Dyer, R. S. Wolfe, W. E. Balch, R. Tanner, L. Magnum, L. B. Zablen, R. Blakemore, R. Gupta, L. Bonen, B. J. Lewis, D. A. Stahl, K. R. Luehrsen, K. N. Chen and C. R. Woese. 1980.

The phylogeny of prokaryotes. *Science*. pp. 261 - 266.

Fox, G. E. 1985. The structure and evolution of archaebacterial rRNA. *In* The Bacteria VIII. *Edited by* Woese, C. R. and R. Wolfe, Academic Press, New York, pp. 257 - 310.

Garrett, R. A., S. Douthwaite and H. F. Noller. 1981. Structure and role of 5S RNA-protein complexes in protein biosynthesis. *Trends in Biochem. Sci.* **6**:137 - 139.

Garrett, R. A., J. Dalgaard, N. Larsen, J. Kjems and A. S. Mankin. 1991. Archaeal rRNA operons. *Trends. Biochem. Sci.* **16**:22 - 26.

Garrett, R. A., M. Aagaard, M. Anderson, J. Z. Dalgaard, J. Lykke-Andersen, H. N. Phan, S. Trevisanato, L. Østergaard, N. Larsen and H. Leffers. 1993. Archaeal rRNA operons. Intron splicing and homing endonucleases, RNA polymerase operons and phylogeny. *In* Molecular Biology of Archaea. *Edited by* Pfeifer, F., P. Palm and K. H. Schleifer, Gustav Fisher Verlag Gmbh and Co, Stuttgart, pp. 180 - 191.

Gerbi, S. A. 1985. Evolution of ribosomal DNA. *In* Molecular Evolutionary Genetics. *Edited by* McIntyre, R. J., Plenum Press, New York, pp. 419 - 517.

Gilbert, W. 1978. Why Genes in Pieces? *Nature*. **271**:501.

Gilbert, W. 1986. The RNA World. *Nature*. **319**:618.

Ginzburg, M., L. Sachs and B. Z. Ginzburg. 1970. Ion metabolism in a *Halobacterium* I. Influence of age of culture on intracellular concentrations. *J. Gen. Physiol.* **55**:187 - 207.

Gogarten, J. P., H. Kibak, Dittrich, L. Taiz, E. J. Bowman, B. J. Bowman, M. F. Manolson, R. J. Poole, T. E. Date, T. Oshima, J. Konishi, K. Denda and M. Yoshida. 1989. Evolution of the vacuolar H^+-ATPase: implication for the origin of eukaryotes. *Proc. Natl. Acad. Sci. USA*. **86**:9355 - 9359.

Green, R. and Szostak. 1992. Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science*. **258**:1910 - 1915.

Griffiths, E. and S. T. Bayley. 1969. Properties of transfer ribonucleic acid and aminoacyl transfer ribonucleic acid synthetases from the extreme halophile, *Halobacterium cutirubrum*. *Biochem*. **8**:541 - 551.

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

Gunderson, J. H., M. L. Sogin, G. Walters and T. F. McCutchan. 1987. Structurally distinct, stage specific ribosomes occur in *Plasmodium*. *Science*. **238**:933 - 937.

Gupta, R., J. M. Lanter and C. R. Woese. 1983. Sequence of the 16S ribosomal RNA from *Halobacterium volcanii*, an archaebacterium. *Science*. **221**:656 - 659.

Haldane, J. B. S. 1932. The Causes of Evolution. Longmans and Green, London.

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

Hancock, J. M. and G. A. Dover. 1988. Molecular coevolution among cryptically simple expansion segments in eukaryotic 26S/28S rRNA. *Mol. Biol. Evol.* **5**:377 - 392.

Hancock, J. M. and G. A. Dover. 1990. Compensatory slippage in the evolution of ribosomal RNA genes. *Nucleic Acids Res.* **18**:5949 - 5954.

Heinonin, T. V. K., M. N. Schnare and M. W. Gray. 1990. Sequence heterogeneity in the duplicated large subunit ribosomal RNA genes of *Tetrahymena pyriformis* mitochondrial DNA. *J. Biol. Chem.* **265**:22336 - 22341.

Henikoff, S. 1987. Exonuclease III generated deletions for DNA sequence analysis. *Promega Notes*. Promega Corp. No. **8**. pp. 1–3.

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

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

Hüber, P. W. and I. G. Wool. 1984. Nuclease protection analysis of ribonucleoprotein complexes: use of the cytotoxic ribonuclease alpha-sarcin to determine the binding sites for *E. coli* ribosomal proteins L5, L18 and L25 on 5S RNA. *Proc. Natl. Acad. Sci. USA*. **8**:322 - 326.

Irani, M. H., L. Orosz and S. Adhya. 1983. A control element within a structural gene: the *gal* operon of *Escherichia coli*. *Cell*. **32**: 783-788.

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

Jarsch, M. and A. Böck. 1985. Sequence of the 16S ribosomal RNA gene from *Methanococcus vanielli*: evolutionary implications. *System. Appl. Microbiol.* **6**:54 - 59.

Joshi, P. and P. P. Dennis. 1993. Structure, function and evolution of the family of superoxide dismutase proteins from halophilic archaebacteria. *J. Bacteriol.* **175**:1572 - 1579.

King, T. C. and D. Schlessinger. 1983. S1 nuclease mapping analysis of ribosomal RNA processing in wild-type and processing deficient *Escherichia coli*. *J. Biol. Chem.* **258**: 12034 - 12042.

Kjems, J. and R. A. Garrett. 1987. Novel expression of the rRNA genes in the extreme

thermophile and archaebacterium *Desulfurococcus mobilis*. *EMBO*. **6**:3521 - 3530.

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

Kjems, J., H. Leffers, T. Olesen, I. Holz and R. A. Garrett. 1990. Sequence, organization and transcription of the ribosomal RNA operon and the downstream tRNA and protein genes in the archaebacterium *Thermophilum pendens*. *Syst. Appl. Microbiol.* **13**:117 - 127.

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

Klein, A., R. Allmansberger, R. Bokranz, M. Knaub, S. B. Müller and E. Muth. 1988. Comparative analysis of genes encoding methyl coenzyme M reductase in methanogenic bacteria. *Mol. Gen. Genet.* **213**:409 - 420.

Klenk, H. P., P. Palm, Loltspeich and W. Zillig. 1992. Component H of the DNA-dependent RNA polymerases of Archaea is homologous to a subunit started by the three eukaryal nuclear RNA polymerases. *Proc. Natl. Acad. Sci. USA*. **1989**:407 -410.

Lake, J. A. 1987a. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. *Mol. Biol. Evol.* **4**:167.

Lake, J. A. 1987b. Determining evolutionary distances from highly diverged nucleic acid sequences. *J. Mol. Evol.* **26**:59 - 73.

Lake, J. A. 1988. Origin of the eukaryotic nucleus determined by rate-invariant analysis of rRNA sequences. *Nature*. **331**:184 - 186.

Lanyi, J. K. and M. P. Silverman. 1972. The state of binding of intracellular K^+ in

Halobacterium cutirubrum. *Can. J. Microbiol.* **18**:993 - 995.

Lanyi, J. K. 1974. Salt-dependent properties of proteins from extremely halophilic bacteria. *Bact. Rev.* **38**:272 - 290.

Lanyi, J. 1979. Physicochemical aspects of salt dependence in halobacteria. *In* Strategies of microbes in extreme environments. *Edited by* Shilo, M. Verlag Chemie: Weinheim, FRG. pp. 93 - 107.

Larsen, A. and H. Weintraub. 1982. An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chicken globin chromatin. *Cell.* **29**:609 - 622.

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

Lechner, K., G. Wich and A. Böck. 1985. The nucleotide sequence of the 16S rRNA gene and flanking regions from *Methanobacterium formicum*: the phylogenetic relationship between methanogenic and halophilic archaebacteria. *System. Appl. Microbiol.* **6**:157 - 163.

Leffers, H. and R. A. Garrett. 1984. The nucleotide sequence of the 16S rRNA gene of the archaebacterium *Halococcus morrhuae*. *EMBO J.* **3**:1613 - 1619.

Leffers, H., J. Kjems, L. Østergaard, N. Larsen and R. A. Garrett. 1987. Evolutionary relationships amongst archaebacteria: a comparative study of 23S rRNAs of a sulphur-dependent extreme thermophile, an extreme halophile and a thermophilic methanogen. *J. Mol. Biol.* **195**:43 - 61.

Leffers, H., J. Egebjerg, A. Anderson, T. Christiensen and R. A. Garrett. 1988. Domain VI of *Escherichia coli* 23S ribosomal RNA-structure, assembly and function. *J. Mol. Biol.* **204**:507 - 522.

Leffers, H., F. Gropp, F. Lottspeich, W. Zillig and R. A. Garrett. 1989. Sequence, organization, transcription and evolution of RNA polymerase subunit genes from the archaebacterial extreme halophiles *Halobacterium halobium* and *Halococcus morrhuae*. *J. Mol. Biol.* **207**:1 - 19.

Lewin, B. 1990. Genes for rRNA are repeated and are transcribed as a tandem unit. *In* Genes IV, Oxford University Press, New York and Cell Press, Cambridge, Mass. pp. 511 - 517.

Lewin, R. 1986. RNA catalysis gives fresh perspective on the origin of life. *Science.* **231**:545 - 546.

Li, W and D. Graur, 1991. Concerted evolution of multigene families. *In* Fundamentals of Molecular Evolution, Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, pp. 162 - 169.

Louis, B. G. and P. S. Fitt. 1971a. Nucleic acid enzymology of extremely halophilic bacteria: *Halobacterium cutirubrum* deoxyribonucleic acid polymerase. *Biochem. J.* **121**:621 - 627.

Louis, B. G. and P. S. Fitt. 1971b. *Halobacterium cutirubrum* RNA polymerase: subunit composition and salt-dependent template specificity. *FEBS Lett.* **14**:143 - 145.

Louis, B. G. and P. S. Fitt. 1972a. Isolation and properties of highly purified *Halobacterium cutirubrum* deoxyribonucleic acid-dependent ribonucleic acid polymerase. *Biochem. J.* **127**:69 - 80.

Louis, B. G. and P. S. Fitt. 1972b. The role of *Halobacterium cutirubrum* deoxyribonucleic acid polymerase subunits in initiation and polymerization. *Biochem. J.* **127**:81 - 86.

Maden, B. E., C. L. Dent, T. E. Farrell, J. Garde, F. McCallum and J. A. Wakeman. 1987. Human rRNA heterogeneity. *Biochem. J.* **246**:519 - 527.

Majumdar, A. and S. Adhya. 1984. Demonstration of two operator elements in *gal*: *in vitro* repressor binding studies. *Proc. Natl. Acad. Sci. USA.* **81**:6100 - 6104.

Maly, P. and R. Brimacombe. 1983. Refined secondary structure models for the 16S and 23S ribosomal RNAs of *Escherichia coli. Nucl. Acids Res.* **11**:7263 - 7286.

Maniatis, T., E. F. Fritsch and J. Sambrook. 1982. Molecular cloning: a laboratory manual, Cold Spring Harbour Laboratories, Cold Spring Harbour, New York.

Mankin, A. S. and A. M. Kopylov. 1981. Secondary structure model for mitochondrial 12S rRNA: an example of economy in rRNA structure. *Biochem Int.* **3**:587 - 593.

Mankin, A. S., N. L. Teterina, P. M. Rubtsov, L. A. Baratova and V. K. Kagramanova. 1984. Putative promoter region of rRNA operon from archaebacterium *Halobacterium halobium. Nucleic Acids Res.* **12**:6537 - 6546.

Mankin, A. S. and V. K. Kagramanova. 1986. Complete nucleotide sequence of the single ribosomal RNA operon of *Halobacterium halobium*: secondary structure of the archaebacterial 23S rRNA. *Mol Gen. Genet.* **202**:152 - 161.

Mankin, A. S., E. A. Skripkin and V. K. Kagramanova. 1987. A putative internal promoter in the 16S/23S intergenic spacer of the rRNA operon of archaebacteria and eubacteria. *FEBS.* **219**:269 - 273.

Mason, S. W., J. Li and J. Greenblatt. 1992. Host factor requirements for processive antitermination of transcription and suppression of pausing by the N protein of bacteriophage λ. *J. Biol. Chem.* **267**:19418 - 19426.

Maxam, A. and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499 - 560.

May, B. P., P. Tam and P. P. Dennis. 1987. Superoxide dismutase from the extremely halophilic archaebacterium *Halobacterium cutirubrum. J. Bacteriol.* **169**:1417 - 1422.

McCutchan, T. F., V. F. d. l. Cruz, A. A. Lai, J. H. Gunderson, H. J. Ellwood and M. L. Sogin. 1988. Primary sequence of two small subunit ribosomal RNA genes from *Plasmodium falciparum. Mol. Biochem. Parasitol.* **28**:63 - 68.

Meier, N., H. U. Göringer, B. Kleuvers, U. Scheibe, J. Eberle, C. Szymkowiak, M. Zacharias and R. Wager. 1986. The importance of individual nucleotides for the structure and function of rRNA molecules in *E. coli*: a mutagenesis study. *FEBS Lett.* **204**:89 - 95.

Mevarech, M. and R. Wercyberger. 1985. Genetic transfer in *Halobacterium volcanii. J. Bacteriol.* **162**:461 - 463.

Mevarech, M., S. Hirsch-Twizer, S. Goldman, S. Yakobson, H. Eisenberg and P. P. Dennis. 1989. Isolation and characterization of the rRNA gene clusters of *Halobacterium marismortui. J. Bacteriol.* **171**:3479 - 3485.

Moazed, D. and H. Noller. 1989. Intermediate states in the movement of transfer RNA in the ribosome. *Nature.* **342**:142 - 148.

Mullakhanbhai, M. F. and H. Larsen. 1975. *Halobacterium volcanii* spec. nov.; a dead sea *Halobacterium* with moderate salt requirement. *Arch. Microbiol.* **104**:207 - 214.

Müller, H. J. 1935. The origination of chromatin deficiencies as minute deletions subject to insertion elsewhere. *Genetics.* **17**:237 - 252.

Mylvaganam S., and P. P. Dennis. 1992. Sequence heterogeneity between the two genes encoding 16S rRNA from the halophilic archaebacterium *Haloarcula marismortui.* Genetics **130**: 399-410.

Nazar, R. N. 1991. Higher order structure of the ribosomal 5S RNA. *J. Biol. Chem.* **266**:4562 - 4567.

Neumann, H., A. Gierl, J. J. Leibrock, D. Straiger and W. Zillig. 1983. Organizations for the genes for ribosomal RNA in archaebacteria. *Mol. Gen. Genet.* **192**:66 - 72.

Nodwell, J. R. and J. Greenblatt. 1993. Recognition of *box A* antiterminator RNA by the *E. coli* antitermination factors Nus B and ribosomal protein S10. *Cell.* **72**:261 - 268.

Noller, H. F., J. Kop, V. Wheaton, J. Brosius, R. R. Gutell, A. M. Kopylov, F. Dohme, W. Heir, D. A. Stahl, R. Gupta and C. R. Woese. 1981. Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res.* **9**:6167 - 6189.

Noller, H. F., D. Moazed, S. Stern, T. Powers, P. Allen, J. Robertson, B. Weiser and K. Triman. 1990. Structure of rRNA and its functional interactions during translation. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 73 - 92.

Noller, H. F. 1991. Ribosomal RNA and translation. *Annu. Rev. Biochem.* **60**:191 - 227.

Noller, H. F., V. Hoffarth and L. Zimniak. 1992. Unusual resistance of peptidyl transferase to protein extraction procedures. *Science.* **256**:1416 - 1419.

Oakes, M. I., A. Scheinman, T. Acha, G. Shankwieler and J. Lake. 1990. Ribosome structure: three dimensional locations of rRNA and proteins. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 180 - 193.

Ochman, H. and A. Wilson. 1987. Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J. Mol. Evol.* **26**:74 - 86.

Ohno, S. 1970. Evolution by Gene Duplication. Springer-verlag, Berlin.

Ohta, T. 1980. Evolution and Variation of Multigene Families. Springer-Verlag, Berlin.

Ohta, T. 1991. Multigene families and the evolution of complexity. *J. Mol. Evol.* **33**:34 - 41.

Oren, A., P. P. Lau and G. E. Fox. 1988. The taxonomic status of *Halobacterium marismortui* from the dead sea; a comparison with *Halobacterium vallismortis*. *System. Appl. Microbiol.* **10**:251 - 258.

Østergaard, L., N. Larsen, H. Leffers, J. Kjems and R. A. Garrett. 1987. A rRNA operon and its flanking region from the archaebacterium *Methanobacterium thermoautotrophicum*,. *Syst. Appl. Microbiol.* **9**:199 - 209.

Pace, N. and T. L. Marsh. 1985. RNA Catalysis and the origin of life. Origins of Life, **16**. 97 - 116.

Pace, N. R., G. J. Olsen and C. R. Woese. 1986. Ribosomal RNA phylogeny and the primary lines of evolutionary descent. *Cell.* **45**:325 - 326.

Pace, N. R. 1991. Origin of life-facing up to the new physical setting. *Cell.* **65**:531 - 533.

Patthy, L. 1985. Evolution of the proteases of blood coagulation and fibrinolysis by assembly from modules. Cell **41**: 657 - 663.

Perlman, P. S. and R. A.Butow. 1989. Mobile introns and intron-encoded proteins. *Science.* **246**:1106 - 1109.

Piatigorsky, J., W. E. O'Brien, B. L. Norman, K. Kalumuck, G. J. Wistow, T. Borras, J. 1988. Gene sharing by delta-crystallin and arginosuccinate lyase. *Proc. Natl. Acad. Sci. USA.* **85**:3479 - 3483.

Piccirilli, J. A., T. S. McConnell, A. J. Zaug, H. F. Noller and T. R. Cech. 1992. Aminoacyl esterase activity of *Tetrahymena* ribozyme. *Science.* **256**:1420 - 1424.

Pieler, T. and J. Hamm. 1987. The 5S gene internal control region is composed of three distinct sequence elements, organized as two functional domains of the variable spacing. *Cell.* **48**:91 - 100.

Pilipenko, E. V., S. V. Maslova, A. N. Sinyakov and V. I. Agol. 1992. Towards identification of *cis*-acting elements involved in the replication of enterovirus and rhinovirus RNAs: a proposal for the existence of tRNA-like terminal structures. *Nucleic Acids Res.* **20**:1739 - 1745.

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

Pürler, G., F. Lottspeich and W. Zillig. 1989b. Organization and nucleotide sequence of the genes encoding the large subunits A, B and C of the DNA-dependent RNA polymerase of the archaebacterium *Sulfolobus acidocaldarius*. *Nucleic Acids Res.* **17**:4517 - 4523.

Ramagopal, S. 1992. Are eukaryotic ribosomes heterogeneous? Affirmations on the horizon. *Biochem. Cell. Biol.* **70**:269 - 272.

Rao, A. L. N., T. W. Dreher, L. E. Marsh and T. C. Hall. 1989. Telomeric function of the tRNA-like structure of brome mosaic virus RNA. *Proc. Natl. Acad. Sci.* **86**:5335 - 5339.

Raue', H. A., J. Klootwijk and W. Musters. 1988. Evolution conservation of structure and function of high molecular weight rRNA. *Prog. Biophys. Mol. Biol.* **51**:77 - 129.

Raue', H. A., W. Musters, C. A. Rutgers, J. V. Riet and R. J. Planta. 1990. rRNA: from structure to function. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 217 - 235.

Ree, H. K., K. Cas, D. Thurlow and R. A. Zimmermann. 1989. Structure and organization of the 16S rRNA gene from the archaebacterium *Thermoplasma acidophilum*. *Can. J. Microbiol.* **35**:124 - 133.

Reiter, W., P. Palm, W. Voos, J. Kanieki, B. Grampp, W. Schulz and W. Zillig. 1987. Putative promoter elements for the ribosomal RNA genes of the thermoacidophilic archaebacterium *Sulfolobus sp.* strain B12. *Nucleic Acids Res.* **15**:5581 - 5595.

Reiter, W. A., P. Palm and W. Zillig. 1988. Analysis of transcription in the archaebacterium *Sulfolobus* indicates that archaebacterial promoters are homologous to eukaryotic Pol II promoters. *Nucl. Acids Res.* **16**:1 - 10.

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

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

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

Sanz, J. L., I. Marin, L. Ramirez, J. P.Abad, C. L. Smith and R. Amils. 1988. Variable rRNA gene copies in extreme halobacteria. *Nucleic. Aci. Res.* **16**:7827 - 7832.

Schopf, J. W. 1993. Microfossils of the early archean apex chert: new evidence of the antiquity of life. *Science.* **260**:640 - 646.

Schwartz, R. M. and M. O. Dayhoff. 1978. Origins of prokaryotes, eukaryotes, mitochondria and chloroplasts. *Science.* **199**:395 - 405.

Sharp, P. A. 1985. On the origin of RNA splicing and introns. *Cell.* **42**:397 - 400.

Shevaick, A., H. S. Gewity, B. Hennemann, A. Yonath and H. G. Wittmann. 1985. Characterization and crystallization of ribosomal particles from *Halobacterium marismortui*. *FEBS Lett.* **184**:68 - 71.

Shimmin, L. and P. P. Dennis. 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.

Shine, J. and L. Dalgarno. 1974. The 3'-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. *Proc. Natl. Acad. Sci. USA.* **71**:1342 - 1346.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503 - 517.

Stern, S., B. Weiser and H. F. Noller. 1988. Model for the three dimensional folding of 16S ribosomal RNA. *J. Mol. Biol.* **204**:447 - 481.

Stockenius, W., R. H. Lozier and R. A. Bogomolini. 1979. Bacteriorhodopsin and the purple membrane of halobacteria. *Biochem. Biophys. Acta.* **505**:215 - 298.

Stockenius, W. and R. A. Bogomolni. 1982. Bacteriorhodopsin and related pigments of halobacteria. *Annual Rev. Biochem.* **52**:587 - 616.

Swofford, D. L. 1993. PAUP: Phylogenetic Analysis Using Parsimony version 3.1. Laboratory of Molecular Systematics, Smithsonian Institution, Washington, D.C.

Tapprich, W., H. U. Goringer, E. d. Slasio, C. Prescott and D. A. E. 1990. Studies of ribosomal function by mutagenesis of *Escherichia coli* rRNA. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 236 - 242.

Thompson, L. and C. Daniels. 1988. A tRNA *trp* intron endonuclease from *H. volcanii:* Unique substrate recognition properties. *J. Biol. Chem.* **263**:17951 - 17956.

Thompson, L. D., L. D. Brandon, D. T. Nieuwlandt and C. J. Daniels. 1989. Transfer

RNA intron processing in the halophilic archaebacterium *Halobacterium cutirubrum*. *J. Microbiol.* **35**:36 - 42.

Thurlow, D. L. and R. A. Zimmerman. 1982. Evolution of protein binding regions of archaebacteria, eubacteria and eukaryotic ribosomal RNAs. *In* Archaebacteria. *Edited by* Kandler. O. Gustav Fisher Verlag, Shuttgart. pp. 347 - 357.

Vogeli, G., H. Ohkubu, M. E. Sobel, Y. Yamada, I. Pastan and B. d. Crombrugghe. 1981. Structure of the promoter for chicken type I collagen gene. *Proc. Natl. Acad. Sci. USA*. **78**:5334 - 5438.

Wais, A. S. 1985. Cellular morphogenesis in a halophilic archaebacterium. *Curr. Microbiol.* **12**:191 - 196.

Walter, M. R. 1983. Archaean stromatolites: evidence of the earth's earliest benthos. *In* Earth's Earliest Biosphere: its origin and evolution. *Edited by* Schopf, J. W. Princeton University Press, Princeton, New Jersy. pp. 187 - 213.

Walters, A. P., C. Syin and T. F. McCutchan. 1989. Developmental regulation of stage-specific ribosome population in *Plasmodium*. *Nature*. **342**:438 - 440.

Watson, J. D., N. H. Hopkins, J. W. Roberts, J. A. Steitz and A. M. Weiner. 1987. *In* Molecular Biology of the Gene. The Benjamin Cummings Publishing Company, Inc., Toronto, 4th edition, Chapter 28, pp. 1095 - 1130.

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

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

Weiner, A. M. and N. Maizels. 1987. 3' Terminal tRNA-like structures tag genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc. Natl.*

Acad. Sci. USA. **84**: 7383 -7387.

Weiner, A. M. and N. Maizels. 1991. The genomic tag model for the origin of protein synthesis: further evidence from the molecular fossil record. *In* Evolution of Life: fossils, molecules and culture. *Edited by* Osawa. S. and Honjo, T., Springer-Verlag, Tokyo, pp. 51 - 66.

Wilson, A., S. Cartson and T. White. 1977. Biochemical evolution. *Annu. Rev.Biochem.* **46**:573 - 639.

Wittmann-Liebold, B., A. Kopke, E. Ardnt, W. Kromer, T. Hetakeyhama and H. G. Wittmann. 1990. Sequence comparison and evolution of ribosomal proteins and their genes. *In* The Ribosome: Structure, Function and Evolution. *Edited by* Hill, W. E., Darlberg, A., Garrett R. A., Moore P. B., Schlessinger, D. and Warner J. R., American Society for Microbiology, Washington, D.C., pp. 598 - 616.

Woese, C. R., R. Gutell, R. Gupta and H. F. Noller. 1983. Detailed analysis of the higher-order structure of 16S-like ribosomal nucleic acids. *Microbiol. Rev.* **47**:621 - 669.

Woese, C. R., R. Gupta, C. M. Hahn, W. Zillig and J. Tu. 1984. The phylogenetic relationships of three sulfur-dependent archaebacteria. *System. Appl. Microbiol.* **5**:97 - 106.

Woese, C. R. and G. Olsen. 1986. Archaebacterial phylogeny: perspectives on the urkingdoms. *System. Appl. Microbiol.* **7**:161 - 177.

Woese, C. R. 1987. Bacterial Evolution. *Microbiol. Rev.*. **51**:221 - 271.

Woese, C. R., O. Kandler and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains archara, bacteria, and eucarya. *Proc. Natl. Acad. Sci. USA*.**87**:4576 - 4580.

Woese, C. R. and N. R. Pace. 1993. Probing RNA structure, function and history by

comparative analysis. *In* The RNA World. *Edited by* Gesteland, R. F. and Atkins, J. F., Cold Spring Harbour Laboratory Press, Cold Spring Harbour. pp. 91 - 117.

Wolffe, A. P. and D. D. Brown. 1988. Developmental regulation of two 5S ribosomal RNA genes.*Science*. **241**:1626 - 1632.

Zhang, H., R. Scholl, J. Browse and C. Somerville. 1988. Double stranded DNA sequencing as a choice for DNA sequencing. *Nucleic Acids Res*. **16**:1220.

Zillig, W. and K. O. Stetter. 1980. Genetics and evolution of RNA polymerase, tRNA and ribosomes. *Edited by* Osawa, S., Ozeki, H., Uchida, H and Yura, T., University of Tokyo Press, Tokyo. pp. 101 -126.

Zillig, W., K. O. Stetter, R. Schnabel, J. Madon and A. Gierl. 1982a. Transcription in archaebacterium. *Zentralbl. Bakteriol.Microbiol. Hyg. Abstr. 1. Orig. C*. **3**:218 - 227.

Zillig, W., R. Schnabel, J. Tu and K. O. Stetter. 1982b. The phylogeny of archaebacterium, including novel anaerobic thermoacidophiles, in the light of RNA polymerase structure. *Naturwissenschaften*. **69**:197 - 204.

Zillig, W., K. O. Stetler, R. Schnabel and M. Thomm. 1985. DNA-dependent RNA polymerases of the archaebacteria. *In* The Bacteria VIII. *Edited by* Woese, C. R. and Wolfe, R., New York Academic Press, New York, pp. 499 - 523.

Zimmerman, R. A., O. L. Thurlow, R. S. Finn, T. L. Marsh and L. K. Ferrett. 1980. Conservation of specific protein-rRNA interaction in ribosome evolution. *In* Genetics and Evolution of RNA polymerase, tRNA and Ribosomes. *Edited by* Osawa, S., Ozeki, H., Uchida, H., and Yura T., University of Tokyo Press, Tokyo, pp. 569 - 580.

Zweib, C., D. K. Jemiolo, W. F. Jacob, R. Wagner and A. E. Dahlberg. 1986. Charecterization of a collection of deletion mutants at the 3'-end of 16S ribosomal RNA of *Escherichia coli*. *Mol. Gen. Genet*. **203**:256 - 264.

APPENDIX

Table A.1 Showing the endonuclease digestion fragments, which are less than 3.0kb, obtained from the operons *rrnA* and *rrnB* of *Ha. marismortui* (see Figure **3.1**). A comparison of these fragments to the restriction maps published by Mevarech *et al.* (1989) revealed discrepancies. The tabulation given below provides a comparison between what was published by Mevarech *et al.* (1989) and what was observed in the gel shown in Figure **3.1**. The following points are considered, the fragments which appear on the gel but not shown on the map (§), the fragments shown on the map but was not observed on the gel (*), the restriction sites that are not indicated at the correct positions on the map (¶), and the restriction site present in the vector (#).

Restriction Enzymes	Fragments from the *rrnA* operon (A) in base pairs	Fragments from the *rrnB* operon (A) in base pairs
ClaI-EcoRI	870	2200
	880	2400
	1600	
	2700 (partial digestion product)	
ClaI-KpnI	2300	450 (§)
	2800 (*)	1700
		1900
ClaI-PstI	400	1850 (¶)
	420	2200 (¶, #)
	600 (§)	3000 (#)
	1500 (§)	
	2100	
ClaI-SmaI	580	2300 (#)

Table A.1 continued.

Restriction Enzymes	Fragments from the *rrnA* operon (A) in base pairs	Fragments from the *rrnB* operon (A) in base pairs
ClaI-XhoI	900 (§)	<450 (*)
	920 (§)	450
	1150 (§)	780
	1900 (§)	1600 (#)
		2600
HindIII-EcoRI	900	2800 (#)
	1500	
	2800 (#)	
HindIII-KpnI	2800 (#)	300
		2800 (#)
HindIII-PstI	400	2800 (#)
	1200 (§)	
	1500 (§)	
	2100	
	2800 (#)	
HindIII-XhoI	900 (§)	450
	920 (§)	780
	1150 (§)	1600 (#)
	1900 (§)	2800 (#)
	2800 (#)	
HindIII-SphI	2800(#)	2800 (#)

Table A.2 Oligonucleotide used for sequencing the *Ha. marismortui rrnA* and *rrnB* operons and primer extention analysis on the *Ha. marismortui* ribosomal RNAs.

Name	Sequence (5'-3')	Size	Description
oPD33	GTCCGATTTAGCCATGCTAG	20mer	*Hma rrnA* and *rrnB* operons, forword, 16S rRNA at position 39-58.
oPD34	CTAGCATGGCTAAATCGGAC	20mer	*Hma rrnA* and *rrnB* operons, reverse, 16S rRNA at position 39-58.
oPD36	ATGCGGGGTTAGGCGGG	17mer	*Hma rrnA* operon, 5'-flanking of 16S rRNA at position -14 to -30.
oPD37	AGCGGCGGAAGTGGTTGC	18mer	*Hma rrnB* operon, 5'-flanking of 16S rRNA at position -10 to -30.
oPD38	CTG(CT)GGCTGGATCACCTCCT	20mer	*Hma rrnA* and *rrnB* operons, 16S rRNA at position 1453-1472.
oPD39	TTAAGTGTGGGACGGCG	17mer	*Hma rrnA* and *rrnB* operons within 16S-23S intergenic spacer at position 106-123 in *rrnA* and position 128-145 in *rrnB*.
oPD40	CCGCCATGTTCAGGAA	17mer	*Hma rrnB* operon, 5'-flanking of 16S rRNA, position -301 to -318.
oPD41	CTGTTACACTTAAGGGC	17mer	*Hma rrnA* operon, 5'-flanking of 16S rRNA, position -301 to -318.
oPD43	CCATCGCCGTCCCACACTTAA	21mer	*Hma rrnA* and *rrnB* intergenic spacer at position 292-311 of *rrnB* and position 106-125 *rrnA*.

Table A.2 continued.

Name	Sequence (5'-3')	Size	Description
oPD44	GCTTGGCACGTCCTTCATCAG	21mer	*Hma rrnA* and *rrnB* 23S rRNA at positions 39-60.
oPD45	CCAAGGGCCGGATTTGAACC	20mer	$tRNA^{Cys}$ reverse primer for *Ha. marismortui* and *Hb. cutirubrum*, position 73-53.
oPD46	GGCGAATCGACCCTTCCCAG	20mer	16S-23S intergenic spacer primer for mapping Pi and processing sites, reverse, position 290-271 in *rrnA* and 291-311 in *rrnB*
oPD48	CTCACATCAGATTCCATCTT	20mer	*Hma rrnB* operon at 5'-flanking of 16S rRNA at position 98-118.
oPD49	CGCTCAGGTCGGCACAATCC	20mer	*Hma rrnB* operon at 5'-flanking of 16S rRNA at position -464 to -444.
oSHM1	GGGTGTGCGCGTCGAGG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 2855-2862.
oSHM2	GAACCTCCAACTCCGT	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 1568-1585.
oSHM3	GGCGGGGGTAACTATGA	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 1942-1959.
oSHM4	AGCATAGGTAGGAGTCG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 2151-2168.
oSHM5	AGGGAGTACTGGAGTGC	17mer	*Hma rrnA* and *rrnB*.5S rRNA at position 74-81.
oSHM6	GGTCGAAGGGGTTGGCG	17mer	*Hma rrnA* and *rrnB* 3'-flanking region of 5S RNAs at position 210-237.

Table A.2 continued.

Name	Sequence (5'-3')	Size	Description
oSHM7	TGGGTGTGTAATGGTGTCTG	20mer	*Hma rrnA* and *rrnB* 23S rRNA at position 2649-2668.
oSHM8	AACGAGGAACGCTGACG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 2390-2407.
oSHM9	GAAAATCCTGGCCATAG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 1347-1354.
oSHM10	GAACAACCCAGAGATAG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 1075-1182.
oSHM11	GAAAGGCACGTGGAAGT	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position 803-820.
oSHM12	GTAACCGCGAGTGAACG	17mer	*Hma rrnA* and *rrnB* 23S rRNA at position212-229.
oCW37	GGTGGTGCATGGCCG	15mer	16S-like primer, forward, *E. coli* 1047-1061.
oCW38	GCATGGC(CT)G(CT)CGTCAG	15mer	16S-like primer, forward, *E. coli* 1053-1068.
oCW40	TGGGTCTCGCTCGTTG	16mer	16S-like primer, reverse, *E. coli* 1115-1100.
oCW40	TCTTAAGGTAGCGAA	15mer	23S-like primer, forward, *E. coli* 1922-1937.
oCW42	CCATTGTAGC(GC)CGCGTG	17mer	16S-like primer, reverse, *E. coli* 1242-1226.
oCW43	CGACCGCCCCAGTCAAACTG	20mer	23S-like primer, reverse, *E. coli* 2260-2280.

Table A.2 continued.

Name	Sequence (5'-3')	Size	Description
oCW45	ACACGCGTGCTACAAT	16mer	16S-like primer, forward, *E. coli* 1225-1240.
oCW46	ACGGGCGGTGTGT(GA)C	15mer	16S-like primer, reverse, *E. coli* 1406-1392.
oCW48	GGTTACCTTGTTACGACTT	19mer	16S-like primer, reverse, *E. coli* 1510-1492.
oCW54	GCTGAAAGCATCTAAG	16mer	23S-like primer, forward, *E. coli* 2744-2759.
oCW59	CGCCGGAAGGGCAAGGGTTCC	21mer	23S-like primer, forward, *E. coli* 1315-1335.
oCW70	GCTTTTCACGGGCCCC	16mer	23S-like primer, reverse, *E. coli* 1575-1560.
oCW76	GCCCAGTGCCGGTATGTG	18mer	23S-like primer, forward, *E. coli* 1835-1852.
oCW78	AGAGGGTGAAAGCCCCGT	18mer	23S-like primer, forward, *E. coli* 322-339.

Table A.3 Plasmids and strains used for the characterization of the *rrnA* and *rrnB* operons in *Ha. marismortui.*

Strain number	Size of the insert	Vector	Host	Description
pD926	2.0kb	pGEM 7+	JM109	A SmaI-HindIII insert consists of a part of 16S rRNA, 16S-23S spacer and a part of 23S rRNA genes from the *rrnA* operon
pD927	1.6 kb	pGEM 7+	JM109	A HindIII-SmaI insert consists of the 5'-flanking of 16S rRNA and a part of 16S rRNA gene from the *rrnB* operon
pD928	2.6 kb	pGEM 7+	JM109	A HindIII-SmaI insert consists of the 5'-flanking of 16S rRNA and a part of 16S rRNA gene from the *rrnA* operon
pD929	2.8 kb	pGEM 7+	JM109	A SmaI-HindIII insert consists of a part of 16S rRNA, 16S-23S spacer and a part of 23S rRNA genes from the *rrnB* operon
pD1097	654 bp	pGEM 3+	DH5α	An AvaI-AvaI insert consists of a part of 16S rRNA, 16S-23S spacer and a part of 23S rRNA genes from the *rrnA* operon.
pD1098	669 bp	pGEM 3+	DH5α	An AvaI-AvaI insert consists of a part of 16S rRNA, 16S-23S spacer and a part of 23S rRNA genes from the *rrnB* operon.
pD1099	1.0 kb	pGEM 7+	DH5α	An EcoR1-EcoR1 fragment consists of the 3'-flanking of 5S and Cystein tRNA gene from the *rrnA* operon
pD1100	1.7 kb	pGEM 7+	DH5α	An EcoR1-EcoR1 fragment consists of a part of 23S rRNA gene and 23S-5S spacer region of the *rrnA* operon.
pD1021	8 kb	pBR322	JM101	A HindIII-ClaI fragment consists of the entire *rrnA* operon.
pD1022	10 kb	pBR 322	JM101	A HindIII-HindIII fragment consists of the entire *rrnB* operon.