CHARACTERIZATION OF THE RIBOSOMAL RNA GENE CLUSTER IN
HALOBACTERIUM CUTIRUBRUM

by

IVY YUK CHEE HUI

B.Sc., The University of British Columbia, 1980
M.Sc., The University of British Columbia, 1982

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

in

THE FACULTY OF GRADUATE STUDIES

Department of Biochemistry

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

August 1984

©Ivy Yuk Chee Hui, 1984
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 \textit{Biochemistry}

The University of British Columbia
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date \textit{Oct 12, 1984}
ABSTRACT

A detailed analysis of the structure and organization of a cloned ribosomal RNA gene cluster from the archaeabacterial species *Halobacterium cutirubrum* was undertaken. The DNA sequence of the entire gene cluster with the exception of a region in the middle of the 23S rRNA gene was determined. The gene organization is found to be identical to that of eubacteria with the 16S, 23S and 5S genes occupying proximal, middle and distal positions respectively. No equivalent to the eucaryotic 5.8S gene could be demonstrated. This ribosomal RNA gene cluster also contains two putative tRNA genes, an alanine tRNA gene in the 16S–23S intergenic space and a cysteine tRNA gene distal to the 5S rRNA gene. The 16S and 23S rRNA genes are surrounded by long nearly perfect inverted repeat sequences which are presumably utilized in the processing of a large primary transcript into mature rRNAs. The 5' sequence flanking the gene cluster contains two imperfect copies, followed by three perfect copies, of a bipartite direct repeat unit which contains the hexanucleotide sequence AAGTAA, believed to an important component of the *Halobacterium* promoter. In the 3' sequence flanking the 5S rRNA gene, there are sequences which may function in transcription termination—an inverted repeat followed by T, upstream from the cysteine tRNA, and a G/C rich region followed by an A/T rich region downstream from the cysteine tRNA. Finally, genomic Southern hybridization experiments indicated that the ribosomal RNA genes are unique single copy DNA within the *Halobacterium cutirubrum* genome.
# TABLE OF CONTENTS

ABSTRACT .......................................................... ii

TABLE OF CONTENTS ................................................ iii

LIST OF FIGURES ................................................... vi

LIST OF TABLES .................................................... viii

ACKNOWLEDGEMENTS ............................................... ix

INTRODUCTION ....................................................... 1

  Cellular evolution of Archaebacteria .......................... 1

  Diversity and unique characteristics of archaebacteria ...... 5

    1. The archaebacterial cell envelopes .......................... 6

    2. Metabolisms and bioenergetics .............................. 7

    3. Transcription and translation machineries ................ 9

    4. Antibiotic sensitivity ....................................... 11

    5. Repeated sequences .......................................... 12

The organization of ribosomal RNA genes ......................... 12

    1. Eubacterial ribosomal RNA gene clusters .................. 12

    2. Eucaryotic ribosomal RNA gene clusters ................... 14

    3. Chloroplast and mitochondrial ribosomal RNA gene clusters .......................... 15

    4. Archaebacterial ribosomal RNA gene clusters .............. 15

The present investigation—molecular characterization of the ribosomal RNA genes in *Halobacterium cutirubrum* .................. 17

ABBREVIATIONS ..................................................... 19

MATERIALS AND METHODS .......................................... 20

Chemicals ......................................................... 20
Bacterial strains and growth conditions ........................................... 21
Purification of Halobacterium cutirubrum chromosomal DNA ............ 21
Construction of λ1059-H. cutirubrum library ................................. 23
Purification of ribosomal RNAs ....................................................... 24
Synthesis of ribosomal cDNA ......................................................... 24
Screening of the library for ribosomal RNA genes ........................... 25
Purification of bacteriophage DNA ................................................... 28
Preparation of a nick-translated probe ............................................. 29
Northern blot analysis ................................................................. 29
Southern blot analysis ................................................................. 30
M13 cloning and DNA sequencing .................................................. 30
Nuclease S1 mapping ................................................................. 31
Restriction enzyme digestion ......................................................... 32
Agarose gel electrophoresis ........................................................... 32
Polyacrylamide gel electrophoresis ................................................ 33
Recovery of DNA from polyacrylamide gels ................................. 33

RESULTS ......................................................................................... 34

Part I: Localization of rRNA gene cluster in genomic
λ1059 library ........................................................................ 34
Construction of λ1059-H. cutirubrum library ................................... 34
Screening for rRNA genes in genomic library ................................... 35
Restriction analysis of λHc4 and λHc9 ............................................ 36
Southern blot analysis of λHc4 and λHc9 DNAs .............................. 38
Northern blot analysis of λHc4 and λHc9 DNAs .............................. 41

Part II: Sequence analysis of the rRNA gene cluster ...................... 43
Subcloning of the rRNA gene into plasmid pBR322 and
M13 vectors ................................................................................. 43
DNA sequences of the rRNA gene cluster ........................................ 46
Part III: Single copy rRNA genes .................................................... 52
DISCUSSION .................................................................................... 58
   Direct bipartite repeat sequence ...................................................... 58
   Ribosomal RNA genes .................................................................. 60
   Inverted repeat sequences .............................................................. 63
   Transfer RNA genes ..................................................................... 67
   Terminator sequences .................................................................. 69
   Unique copy of ribosomal RNA genes .......................................... 71
REFERENCES ................................................................................... 73
| Figure 1: | The three primary kingdoms | 3 |
| Figure 2: | Organization of ribosomal RNA gene clusters in eubacteria, eucaryote, and in the organelles chloroplast and mitochondria | 13 |
| Figure 3: | Organization of ribosomal RNA genes in ten different species of archaebacteria | 16 |
| Figure 4: | Plaque hybridization autoradiogram | 27 |
| Figure 5: | Restriction endonuclease maps of λHc4 and λHc9 | 37 |
| Figure 6: | Southern hybridization of ribosomal cDNA to λHc4 and λHc9 DNAs | 39 |
| Figure 7: | Cross Southern hybridization of λHc4 and λHc9 | 40 |
| Figure 8: | Northern hybridization of λHc4 and λHc9 DNA to ribosomal RNAs | 42 |
| Figure 9: | Plasmid subclones of λHc4 and λHc9 | 44 |
| Figure 10: | Sequencing map of the ribosomal gene cluster | 45 |
| Figure 11: | The 5' flanking sequence of the ribosomal RNA gene cluster | 47 |
| Figure 12: | The 16S ribosomal RNA gene sequence | 49 |
| Figure 13: | Nuclease S1 mapping of the 5' and 3' end of mature 23S rRNA | 50 |
| Figure 14: | The intergenic space between the 16S and 23S rRNA genes | 51 |
| Figure 15: | The 3' flanking sequence of the 23S rRNA gene | 53 |
Figure 16: Location of the fragment used for genomic Southern hybridization with respect to the rRNA gene cluster .............................................. 54

Figure 17: Genomic Southern hybridization with probes from within the rRNA gene cluster .................................................. 55

Figure 18: Genomic Southern hybridization probe with fragment containing a 5' flanking repeat ....................................... 57

Figure 19: Direct repeat units in the 5' flanking sequence of the 16S rRNA gene .......................................................... 59

Figure 20: Secondary structure map of 16S rRNA of *H. cutirubrum* .............................................................. 61

Figure 21: Nucleotide sequence homology between *E. coli*, *H. cutirubrum* and *M. vannielii* 23S rRNA genes ............. 64

Figure 22: Sequences flanking the 16S rRNA gene and the 23S rRNA gene .......................................................... 65

Figure 23: Cloverleaf structure of putative alanine and cysteine tRNAs .............................................................. 68

Figure 24: An inverted repeat termination-like signal ............... 70
LIST OF TABLES

Table 1 Genotypes of *E. coli* and phage λ strains .............. 22
ACKNOWLEDGEMENTS

I am very grateful to my research supervisor, Dr. P.P. Dennis, for his patience, guidance throughout my study and especially for the time he spent on the secondary structural map of 16S rRNA of *H. cutirubrum*.

I would like to thank Dr. R.T.A. McGillivray and Dr. J. McPherson for their continuous encouragement and valuable suggestions. In addition, I would also like to thank our lab. technician Mrs. D. de Jong Wong for her technical help.

I would like to acknowledge the financial support of the Canadian Medical Research Council.
INTRODUCTION

Cellular evolution of Archaebacteria

Before 1977, life on earth was believed to be fundamentally dichotomous (80, 127). All organisms were classified as either eucaryotic (e.g. cells containing a well formed nucleus; including plants, animals and some unicellular organisms) or procaryotic (e.g. cells lacking a well formed nucleus; including mycoplasma, cyanobacteria and true bacteria). This phylogenetic dichotomy was based mainly on the structural and functional aspects of the living cell and had neglected the evolutionary origins of these phenotypic characteristics.

In the past 15 years, it has become apparent that there is an enormous evolutionary gap between eucaryotes and procaryotes with no known intermediates (128). The symbiotic theory, one of the most popular theories for explaining the origin of mitochondria and chloroplasts in eucaryotes, suggested that free-living blue green algae and aerobic bacteria separately invaded protoeucaryotic host cells (79, 118, 126). After becoming symbiotic, they continued to evolve to their current status as organelles. These endosymbiotic associations alone are not a sufficient explanation of the many differences that exist between procaryotes and eucaryotes; the aboriginal line of descent of the cytoplasm of the eucaryotic cell is ill-defined and remains unexplained.

Recent studies on cellular evolution have indicated that similar phenotypes can be achieved in organisms with different evolutionary pathways—this phenomenon is termed convergent evolution. A recent technique for following evolution utilizes comparative amino acid or
nucleotide sequences of macromolecules as chronometers of change (32, 166). Using these molecular chronometers it is possible to estimate evolutionary distances and relationships, and to infer the molecular characteristics of ancestral forms without regard to phenotypic characteristics. These techniques are also extremely useful in classifying examples of convergent evolution. The RNA of the small ribosomal subunit (14-18S rRNA) was found to be the best choice for phylogenetic purposes because of its functional constancy in all living organisms, its universal distribution, its size and its ease of isolation (35, 143). Although it is possible to sequence an entire small subunit ribosomal RNA, it is sufficient and much more rapid to simply characterize its T1 oligonucleotide catalog. Using this technique Woese and his coworkers (34, 158) were able to show that living organisms can be divided into three, not two aboriginal lines of descent:

1. The eubacteria—all typical bacteria;
2. The urcaryotes—the nucleate precursors of the eucaryotic cells; and
3. The archaebacteria—methanogenic, extreme halophilic and thermoacidophilic procaryotes.

The fundamental distinctions in this tripartite division of life are molecular rather than supramolecular or structural (Figure 1).

Woese et al., (155, 157, 159) suggested that all extant life has stemmed from a common, universal ancestor—the progenote, which is a simpler entity than a procaryote (Figure 1). In the progenote stage, complexes were still evolving and trying to establish stable relationships between genotypes and phenotypes. The complexes are visualized as semi-autonomous subcellular entities that somehow group to give ill-defined
Woese et al., (156) proposed that the three kingdoms, urcaryotes, eubacteria and archaebacteria, arose from a universal ancestor—the progenote. Eubacteria and archaebacteria are procaryotes and they do not contain a nuclear membrane. Eubacteria contain cyanobacteria, photosynthetic bacteria, and true bacteria; whereas archaebacteria contain halophiles, methanogens and thermoacidophiles. Urcaryotes are nucleated protoeucaryotic cells. They were separately invaded by cyanobacteria and purple photosynthetic bacteria which became endosymbionts and gradually evolved to organelles—chloroplasts and mitochondria respectively, hence urcaryotes become eucaryotes.
cellular forms with a ready exchange and flow of genetic materials. Presumably, as information processing became more accurate, the level of exchange diminished, and the progenote stabilized to gave rise separately to the eubacteria, urcaryote and archaebacteria. It has been suggested that the urcaryote evolved from the progenote at a somewhat later time than its procaryotic counterpart because of the presence of the nucleus (155).

Although archaebacteria bear no structural or supramolecular resemblance whatsoever to eucaryotic cells, they have certain molecular properties that are characteristically eucaryotic. These include:

1. The light-transducing bacteriorhodopsin in *H. halobium* (7, 96);
2. The non-formylated initiator methionine tRNA (150);
3. The amino acid sequence of ribosomal ‘A’ protein (81);
4. Introns in tRNA genes (53);
5. The histone-like proteins, and actin- and myosin-like proteins in *Thermoplasma* (120, 135);
6. The ADP ribosylation of the archaebacterial translation elongation factor by diphtheria toxin (60, 61);
7. The preferential charging of archaebacterial tRNAs by eucaryotic rather than eubacterial aminoacyl tRNA synthetases (69);
8. The sensitivity towards anisomycin but not chloramphenicol (44, 99).

Indeed, it has been suggested that archaebacterial-like cells evolved to give the precursor of the protoeucaryotes prior to endosymbiotic eubacterial invasion (68, 145).
Diversity and unique characteristics of archaebacteria

The kingdom archaebacteria has been divided into three subgroups (35, 155; Figure 1) according to the \( T_1 \) oligonucleotide catalog of their rRNAs:

1. The methanogens and extreme halophiles;
2. The thermoacidophiles—e.g. *Sulfolobus*;
3. The wall-less thermoacidophiles—*Thermoplasma*.

Collectively these organisms inhabit extreme and unusual environmental niches similar to conditions prevalent on earth at least 3500 million years ago. Methanogenic bacteria, which were originally described by Barker in 1956 (5) as anaerobic bacteria that metabolize \( \text{H}_2, \text{CO}_2, \text{methanol or acetate} \) to form methane gas, are found in sewage-treatment plants, in bogs or in the bottom of oceans or hot springs. Extreme halophiles are bacteria that require high concentrations of salt to survive; they grow in salty habitats along the ocean borders, the Great Salt Lake and the Dead Sea. *Sulfolobus* is a thermoacidophiles which grows at high temperatures (80°C) in extremely acidic environments (pH=2) and is found in hot sulfur springs; *Thermoplasma* is a thermoacidophiles which possesses no cell wall but merely the limiting cell membrane and is found in the slug of hot acidic coal tailings (20).

In spite of the small number of known species of archaebacteria, their phylogenetic diversity and phenotypic variety appear comparable to that of the true bacteria. This diversity has been confirmed by a wide range of biochemical and molecular techniques which are discussed below.
1. The archaebacterial cell envelopes

The diaminopimelic and muramic acid-containing peptidoglycan, an essential component of eubacterial cell wall, is absent in all archaebacteria (55). In Methanobacteriales, murein is replaced by an unique analogous structure, pseudomurein (63). Although both murein and pseudomurein possess alternating N-acetylated hexosamines and acid-derivatives of hexosamines cross-linked by short peptides, they are different in a number of ways. Unlike murein, pseudomurein contains:

1. D-glucosamine and D-galactosamine instead of only D-glucosamine;
2. L-talosaminuronic acid instead of D-muramic acid;
3. γ-1,3 linkage instead of γ-1,4 linkage; and

Other gram-positive archaebacteria contain heteropolysaccharides (e.g. galactosamine, glucose, mannose, glucuronic or galacturonic acid) and have no mureins in their cell envelopes (54, 55). The gram-negative group of archaebacteria contains no rigid cell wall. Their cell envelopes are composed of glycoproteins arranged in hexagonal arrays (4, 51), or in fibrillar structures which sometimes form a protein sheath encasing several cells (54, 55). Thermoplasma contains no cell wall but rather it has a cytoplasmic membrane containing glycoproteins (162).

Archaebacteria lack the ester-linked straight chain fatty acids characteristically found in the membranes of the eubacterial and eucaryotic cells. Instead they have isoprenoid, isoprenyl glycerol ethers, and non-polar isoprenoid hydrocarbons (58, 71). The glycerolipids of archaebacteria contain identical ether-linked alkyl chains of C_{20}-phytane or C_{40}-biphytane and are called diphytanylglycerol diether or dibiphytanyldiglycerol tetraether.
respectively (58, 59, 72, 73, 139). While diethers are the sole glycerolipid in the halophiles, tetraethers constitute nearly all of the glycolipid residues of *Thermoplasma* and *Sulfolobus*. The tetraethers of thermoacidophiles can also contain from 1 to 4 cyclopentyl rings (23, 24). Increased cyclization in the biphytanyl group is found with strains that grow at extremely high temperature (70). The diethers form a normal lipid bilayer through interaction of opposing phytanyl chains, whereas the tetraethers span the membrane and resemble an amphiphilic lipid "monolayer" (72). These diether and tetraether components are used to assemble polar lipids which are analogous to the glycolipids and phospholipids found in membranes of bacterial and eucaryotic cells (58, 74, 83, 123). Neutral lipids of archaebacteria are characterized by isoprenoid and non-polar isoprenoid hydrocarbons (48, 141). The hydrocarbons are C_{15}–C_{30} derivatives of isoprenoid skeletons with varying degrees of unsaturation, e.g. squalene (C_{30}), pentaisoprenes (C_{25}) and their unsaturated derivatives. Variation in degree of reduction reflects differences in the physiological state of the cell, e.g. the ratio of squalene to hydrosqualene decreases proportionately with decreased aeration rates (140). Therefore, these lipids permit archaebacteria to vary the fluidity and stability of their membranes.

2. Metabolisms and bioenergetics

Not only do archaebacteria appear to use novel pathways in constructing cell envelopes and synthesizing lipids, but the methanogens have evolved a distinct pathway for the acquisition of energy from a restricted range of substrates (H₂, CO₂, trimethylamine, methanol and acetate). Most methanogenic bacteria are autotrophic and can grow on H₂ plus CO₂ as sole source energy, CO₂ being reduced to methane (4, 134).
Studies on the biochemistry of methanogenesis have revealed a new cofactor involved in methyl-transfer reactions, coenzyme M or 2-mercaptoethansulfonic acid (84, 133). The reduction of methyl coenzyme M to methane is an exergonic process and is coupled with the net synthesis of ATP. In addition to coenzyme M, at least six other new cofactors are involved in methanogenesis (109, 146). Neither the mechanism of methane generation nor that of ATP synthesis has been elucidated completely.

Besides the highly specialized energy metabolism, the autotrophic methanogenic bacterium Methanobacterium thermoautotrophicum assimilates CO₂ via a novel non-Calvin-type CO₂ fixation pathway (36, 37). The central intermediate of this pathway is acetyl coenzyme A, which appears to be synthesized from 2CO₂ via bound one-carbon units, and is the starting compound for the synthesis of the carbon skeleton of all cell compounds except for one-carbon units. Further CO₂ fixation proceeds to generate pyruvate, oxaloacetate, succinyl coenzyme A, and α-ketoglutarate, etc. More detailed studies are required to provide additional information on this metabolic pathway.

The halophilic archaeabacteria, e.g. Halobacterium halobium and Halobacterium cutirubrum, possess a purple membrane containing bacteriorhodopsin (7, 131). Bacteriorhodopsin, like the eucaryotic rhodopsins, consists of a protein, bacteriorhodopsin, and a retinal molecule. Bacteriorhodopsin is a single chain polypeptide of 248 amino acids which transverses the membrane seven times, and is covalently linked to the retinal via a Schiff base conformation at the ε-amino group of lysine-216.
residue (6, 62). This protein is involved in the light-dependent membrane translocation of protons and consequently is responsible for generating a transmembrane electrochemical gradient.

The thermoacidophiles oxidize molecular sulfur to sulfate in order to generate ATP and reducing power.

\[
2S + 2H_2O + 3O_2 \rightarrow 2SO_4^{2-} + 4H^+
\]

\[
S + 4H_2O \rightarrow SO_4^{2-} + 8H^+ + 6e^-
\]

In *Thermoplasma acidophilum*, the sulfate anion is extruded by a membrane bound ATPase, leaving a positive electrical potential inside the cell (19). It is this positive electrical potential which enhances the active transport of protons against its concentration gradient to maintain the internal pH of the cell near neutrality. The electron transport chain, which contains cytochrome \( b \) and quinone, in the *Thermoplasma* membrane is responsible for the transport of protons across the membrane.

3. Transcription and translation machineries

The DNA-dependent RNA polymerases from archaebacteria appear to contain as many as 9 to 11 different protein subunits (165). These polymerases appear much more complex than the eubacterial \( \sigma, \beta, \delta, \rho \) holoenzyme and are resistant to rifampicin and streptolidigan. On the other hand, archaebacterial polymerase resembles yeast RNA polymerase I in its complexity, its resistance to \( \alpha \)-amanitin and its stimulation by the alkaloid silybin.

The archaebacterial ribosome subunits (30S, 50S) and rRNAs (16S, 23S, 5S) are similar in sizes to those found in eubacteria (81). However,
recent electron micrographs on the 30S ribosomal subunit from archaeabacteria showed structural features in common with those found in the 40S subunit of eucaryotes but not in the 30S subunit of eubacteria. The archaeabacterial 70S ribosome contains 50 to 65 proteins (115, 116). This value is close to the number found in the *E. coli* ribosome (54 proteins) and less than that found in typical eucaryotic cytoplasmic ribosomes (70-80 proteins). Most archaeabacterial ribosomal proteins are acidic, ranging from 35-66% in methanogens to >90% in the extreme halophiles, whereas in *E. coli* only 16% of the ribosomal proteins are acidic (8, 26). The ribosomal ‘A’ protein (equivalent to the *E. coli* L12 protein) appears to be more closely related by amino acid sequence homology to the eucaryotic than to the eubacterial equivalent. The ‘A’ protein is a multicopy ribosomal protein present in four copies per ribosome, and binds to 23S rRNA as a 4:1 complex with another ribosomal protein. In *E. coli* this protein is L10, whereas in *Halobacterium* and *Sulfolobus* this protein by N-terminal amino acid sequence analysis is related to the *E. coli* L11 protein (81).

The archaeabacterial 5S rRNA in the large ribosomal subunit shows unique secondary structure that does not resemble its eubacterial nor eucaryotic counterparts (33). Archaeabacteria also exhibit several unusual post-transcriptional modifications of tRNAs. Their tRNAs do not contain the uracil to dihydrouracil modification. In addition, the normal ribothymine modification in the universal sequence TψCG is modified to pseudouridine or N-methyl pseudouridine instead, to give ψψCG or ψmψCG (66, 156).
4. Antibiotic sensitivity

The sensitivity of archaebacteria towards antibiotics have been examined using agar diffusion tests and tube dilution assays (44, 99). Archaebacteria were insensitive to many eubacterial and eucaryotic antibiotics. They were resistant to: penicillin and D-cycloserine which inhibit the cross-linking of the murein cell wall in eubacteria; cycloheximide which inhibits protein synthesis on the eucaryotic 80S ribosome; spectinomycin, streptomycin and erythromycin which inhibit protein synthesis on the eubacterial 70S ribosome; rifampicin which inhibits the DNA-dependent RNA polymerase of eubacteria and α-amanitin which blocks the RNA polymerase II and III of eucaryotes. Archaebacteria are sensitive to: bacitracin, gardimycin and nisin which inhibit the lipid cycle in the biosynthesis of cell wall polymers, lasalocid, a K⁺-ionophor, and monensin, a Na⁺-ionophor, which interfere with membrane functions and anisomycin which inhibits protein synthesis on the archaebacterial ribosome. Mixed sensitivity to some antibiotics was also observed. The methanogens were found to be susceptible to gramicidin S, a cyclic decapeptide which increases the permeability of membranes, and chloramphenicol which inhibits peptidyl transferase activity of the eubacterial 70S ribosomes; in contrast, these antibiotics do not seem to interfere with the growth of the halophiles. However, it remains to be shown whether insensitivity to some or all of these antibiotics is due to:

1. The impermeability of the cytoplasmic membrane, or
2. The inactivation of the antibiotics by the cell, or
3. The lack of a particular target for the antibiotic.
5. Repeated sequences

Molecular cloning of DNA fragments and Southern hybridizations indicated that the genomes of the halophiles contain more than 50 families of repeated sequences located on both chromosome and plasmid (112, 113). Most of these repeat sequence families are related to spontaneous genomic rearrangements by means of transposition, deletion or recombination events. At least five different transposon-like elements ranging from 520 bp to 3 kb in size have been characterized in *H. halobium* (100, 160). Repeat sequence-associated rearrangements occur at high frequencies and result in mutation rates due to insertional inactivation in the range of $10^{-4}$ to $10^{-2}$ per generation in the bacteriorhodopsin gene (101). These insertion mutations are unstable and reversible, and provide the cell with an unusual degree of structural and perhaps functional variability with its genome.

The organization of ribosomal RNA genes

1. Eubacterial ribosomal RNA gene clusters

The typical eubacterium, *E. coli*, contains seven operons of ribosomal RNA genes scattered around its genome (rrnA–G; Ref. 3). Each cluster contains three ribosomal RNA genes in the order: 16S–23S–5S (Figure 2). These genes are cotranscribed as a 30S rRNA molecule which is cleaved by RNase III to generate precursor 16S and 23S species (39). They are further processed by a set of site-specific endonucleases and their nucleotides are modified to give mature rRNAs. These transcription units also contain tRNA genes. In the 16S–23S spacer region in all rRNA operons, there is either a gene for glutamate tRNA or genes for isoleucine and alanine tRNAs (15, 164). Similarly, distal to the 5S rRNA gene in some
Figure 2: Organization of ribosomal RNA gene clusters in eubacteria, eucaryote, and in the organelles chloroplast and mitochondria.

Ribosomal RNA genes are shown in open boxes and tRNA genes are shown in black boxes. Intervening sequences within rRNA and tRNA genes are indicated by shaded boxes. There are seven rRNA cistrons (16S-23S-5S) in *E. coli* (3, 90). In the *rrnC* operon, there is a glutamate tRNA gene in the 16S-23S spacer, and aspartate and tryptophan tRNA genes distal to the 5S rRNA gene. In the *rrnD* operon, there are isoleucine and alanine tRNA genes in the 16S-23S spacer and a threonine tRNA distal to the 5S rRNA gene.

In eucaryotes, the nuclear rRNA genes are arranged in two transcription units. In human, the two rRNA transcription units (18S-5.8S-28S, 5S) are located on different chromosomes and are arranged as tandem repeats of hundreds of copies in the genome separated by non-transcribed spacers (30, 149). In *Physarum polycephalum*, the two rRNA transcription units (19S-5.8S-26S, 5S) are located on different chromosomes as hundreds of repeat copies separated by non-transcribed spacers with the large transcription units arranged as palindromic repeat units. The 26S rRNA gene has two intervening sequences (17). In *Saccharomyces cerevisiae*, the two rRNA transcription units (18S-5.8S-25S, 5S) are arranged close together as hundreds of alternating repeats in the genome (10, 144).

In *Zea mays* chloroplast DNA, the two identical rRNA units (16S-23S-4.5S-5S) are arranged in an inverted orientation and separated by 18.5 kb of DNA. In the 16S-23S spacer, there are isoleucine and alanine tRNA genes with large intervening sequences (64). In *Euglena gracilis* chloroplast DNA, the three rRNA transcription units (23S-16S-5S) are arranged in tandem repeats with isoleucine and alanine tRNA genes in the 16S-23S spacer (97).

In mitochondrial DNA, there is only a single copy of each rRNA gene. In human mitochondrial DNA, the arrangement of the ribosomal rRNA gene cluster is: tRNA(phenylalanine)-12S-tRNA(valine)-16S-tRNA(leucine). There appear to be no noncoding sequences between the junctions of the tRNA and rRNA genes (2). In *Zea mays* mitochondria, the rRNAs are: 26S, 18S and 5S. The 5S and 18S rRNA genes are close together and separated from the 26S gene by 16 kb of DNA (129). In *Saccharomyces cerevisiae* mitochondria, the rRNAs are 21S and 15S, and their genes are separated by 25 kb of DNA (98). (Note: The non-transcribed spacer between each repeating unit is not drawn to scale.)
EUBACTERIA

E. coli (rrnC)

E. coli (rrnD)

EUCARYOTE

Human

P. polycephalum

S. cerevisiae

CHLOROPLAST

Z. mays

E. gracilis

MITOCHONDRIA

Human

Z. mays

S. cerevisiae
operons, there is either a gene for threonine tRNA, or aspartate tRNA or genes for aspartate and tryptophan tRNAs or glycine, threonine and tyrosine tRNAs (90). Two tandem promoters which initiate transcription approximately 200 and 300 bases upstream from the coding region of the 16S rRNA gene are present in each of the seven transcription units (22, 163).

2. Eucaryotic ribosomal RNA gene clusters

Eucaryotes have four different types of ribosomal RNA molecules organized into two transcription units: One codes for the precursor of the 17–18S, 5.8S and 25S–28S rRNAs (transcribed in a 5′ to 3′ manner), and the other codes for the 5S rRNA (65, 102; Figure 2). The 5.8S rRNA is functionally and structurally equivalent to the 5′ end of the eubacterial 23S rRNA (147). In yeast and other lower eucaryotes, these two types of transcription units are found clustered in an alternating fashion of more than a hundred repeat copies (rDNA) at a single site in the genome (10, 144). In higher eucaryotes, the large 18S–5.8S–28S rRNA repeating units are located at different chromosomal positions than the 5S rRNA repeating units (30, 149). Between the large repeating units are long stretches of non-transcribed spacer DNA. In Physarum polycephalum and various strains of Tetrahymena, there are extrachromosomal copies of the rDNA within which the rRNA repeating units are present as palindromic dimers (17, 152). In eucaryotes, the 5S rRNA gene is transcribed by RNA polymerase III and the precursor 18S–5.8S–23S rRNA is transcribed by RNA polymerase I (13, 104, 106, 148). The eucarytic ribosomal RNA unit does not contain any tRNA genes; some of the large ribosomal RNA genes of lower eucaryotes contain intervening sequences that are present in the primary transcript but not in the mature rRNA.
3. Chloroplast and mitochondrial ribosomal RNA gene clusters

All chloroplasts and mitochondria code for their own ribosomal RNAs. In the chloroplast genome, e.g. maize, the ribosomal RNA genes are organized in the order 16S, 23S, 4.5S and 5S (64, 97). In most chloroplast genomes, there are only two copies of each ribosomal RNA gene; these are present as inverted repeats and orientated far apart from each other (9; Figure 2). One exception is in *Euglena*, where there are three copies of rRNA genes located very close together as tandem repeats (105). Transfer RNA genes are found in most 16S–23S spacer regions, and introns are found in some 23S rRNA and tRNA genes.

In the mitochondrial genome there generally appears to be only two high molecular weight ribosomal RNAs present in single copies (2, 43, 98; Figure 2). In higher eucaryotes these two genes are closely linked with a short spacer region of 100 to 200 bp, whereas in lower eucaryotes the genes are widely separated. In higher plant mitochondria, e.g. maize, in addition to these two rRNAs there is also a small 5S rRNA (129). Genes encoding tRNAs have been found near the rRNA genes and some rRNA genes contain introns.

4. Archaebacterial ribosomal RNA gene clusters

The organization of ribosomal RNA genes in at least ten different archaebacteria have been determined (50, 93, 142; Figure 3). There appear to be in most species only three types of rRNA genes, closely linked as in eubacteria in the order 16S–23S–5S; no eucaryotic-like 5.8S rRNA genes have been identified. The genome of *Halobacterium halobium*, *Thermoproteus tenax*, *Thermofilum pendens*, *Desulfurococcus mobilis* and *Desulfurococcus*
Figure 3: Organization of ribosomal RNA genes in ten different species of archaebacteria (50, 93, 142).

The rRNA genes are represented by open boxes and all transfer RNAs which are associated with these clusters are represented by black boxes. Broken lines indicated the genes are not linked. The rRNA gene clusters in most archaebacteria are closely linked in the order; 16S-23S-5S. *Halobacterium halobium, Thermofilum pendens, Thermoproteus tenax, Desulfurococcus mobilis* and *Desulfurococcus mucosus* have only one single copy of the rRNA gene cluster. *Sulfolobus acidocaldarius* and *Thermococcus celer* have one copy of the rRNA gene cluster and an extra copy of the 5S gene. *Methanococcus vannielii* has four copies of the rRNA gene cluster and an extra copy of the 5S gene. An alanine tRNA gene was found in one of the four 16S-23S spacers in *M. vannielii*. Sequences distal to the 5S gene also code for a tRNA of unknown sequences. In *Thermoplasma acidophilum*, the three rRNA genes are unlinked, and there is a tRNA gene (sequence not known) closely linked to one end of the 23S gene.
H. halobium

M. thermoautotrophicum

M. vannielii

plus two more clusters not yet mapped

S. acidocaldarius

T. acidophilum

T. celer

T. pendens

T. fenax

D. mobilis

D. mucosus
mucosus contain only single copies of each rRNA gene; the genome of Sulfolobus acidocaldarius and Thermococcus celer contain single copies of closely linked 16S, 23S and 5S genes and an unlinked 5S gene; Methanobacterium thermoautotrophicum contains two sets of rRNA genes; Methanobacterium vannielii contains four sets of rRNA genes and an extra 5S rRNA gene. A unique organization of rRNA genes in archaebacteria is found in Thermoplasma acidophilum, in which the rRNA genes (5S, 16S and 23S) are all unlinked and present once per genome. Since little sequence data is available from archaebacteria, it is not yet known whether the spacer regions of these rRNA genes contain any specific tRNA genes. A recent paper has reported the presence of an alanine tRNA gene in one of the 16S–23S rRNA spacers of Methanococcus vannielii (49). Some archaebacterial tRNA genes apparently not associated with rRNA genes were found to contain introns (53); however, no introns have been found in any of the archaebacterial rRNA genes examined.

The present investigation—molecular characterization of the ribosomal RNA genes in Halobacterium cutirubrum

At present very little is known about archaebacteria at the DNA sequence level. Although some 5S rRNAs and tRNAs sequences are available, they were obtained from RNA sequencing (29, 138); in total there are less than 10 genes with established DNA sequences. To understand the underlying molecular mechanisms responsible for the diverse and unusual features of their protein synthesis and the novel mechanism for regulating gene expression, a comprehensive sequence analysis of the ribosomal RNA gene cluster from the organism Halobacterium cutirubrum was carried out in this study.
Only a few general features of the ribosomal RNA genes in archaebacteria are known. It has been shown that rRNA genes, are closely linked on the *Halobacterium* genome in the order 5' 16S-23S-5S 3', and that the number of copies per genome is probably one (45, 92). The 3' end of the 16S rRNA sequence from *H. halobium* and the complete 16S gene sequence from *H. volcanii* have been published (42, 52). The secondary structure of *H. volcanii* 16S rRNA conforms to the universal structure proposed for small subunit rRNA. Finally, an alanine tRNA gene has recently been identified in a 16S-23S spacer region of the organism *Methanococcus vannielii*.

The results presented here confirm that rRNA gene organization in *Halobacterium cutirubrum* resembles those found in eubacteria, arranged in a 5' to 3' manner 16S-23S-5S. This gene cluster is represented only once in the genome. The flanking regions surrounding the cluster and the spacer regions within the gene cluster have been sequenced. Within the rRNA gene cluster, there are also two putative tRNA genes—an alanine tRNA gene located in the 16S-23S intergenic space and a cysteine tRNA gene located distal to the 5S rRNA gene. The regions surrounding the 16S and 23S rRNA genes are potentially rich in RNA secondary structures; these unusual structural features are probably utilized in the processing and maturation of 16S and 23S rRNAs. Putative promoter-like sequences containing two imperfect and three perfect copies of a precise direct bipartite repeat unit were identified in the 5' flanking sequence of the gene cluster. In addition, two transcription termination-like sequences were observed; an inverted repeat followed by T₅ is found at the end of 5S rRNA genes and a G/C rich region followed by an A/T rich region is found after the cysteine tRNA gene.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMV</td>
<td>Avian myeloblastosis virus</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine-5'-triphosphate (disodium salt)</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pair</td>
</tr>
<tr>
<td>BPB</td>
<td>Bromophenol blue</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine serum albumin</td>
</tr>
<tr>
<td>dATP</td>
<td>2'-deoxy-adenosine-5'-triphosphate (disodium salt)</td>
</tr>
<tr>
<td>dCTP</td>
<td>2'-deoxy-cytidine-5'-triphosphate (sodium salt)</td>
</tr>
<tr>
<td>dGTP</td>
<td>2'-deoxy-guanosine-5'-triphosphate (sodium salt)</td>
</tr>
<tr>
<td>dTTP</td>
<td>2'-deoxy-thymidine-5'-triphosphate (sodium salt)</td>
</tr>
<tr>
<td>ddATP</td>
<td>2',3'-dideoxy-ATP (disodium salt)</td>
</tr>
<tr>
<td>ddCTP</td>
<td>2',3'-dideoxy-CTP (sodium salt)</td>
</tr>
<tr>
<td>ddGTP</td>
<td>2',3'-dideoxy-GTP (sodium salt)</td>
</tr>
<tr>
<td>ddTTP</td>
<td>2',3'-dideoxy-TTP (sodium salt)</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>cDNA</td>
<td>Copy DNA</td>
</tr>
<tr>
<td>ssDNA</td>
<td>Single-stranded DNA</td>
</tr>
<tr>
<td>DNase</td>
<td>Deoxyribonuclease</td>
</tr>
<tr>
<td>DTT</td>
<td>Dithiothreitol</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediamine tetracetic acid disodium salt</td>
</tr>
<tr>
<td>MOPS</td>
<td>Morpholinopropane sulfonic acid</td>
</tr>
<tr>
<td>PEG</td>
<td>Polyethylene glycol</td>
</tr>
<tr>
<td>PIPES</td>
<td>(Piperazine-N,N'-bis[2-ethane sulfonic acid])</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>rRNA</td>
<td>Ribosomal RNA</td>
</tr>
<tr>
<td>tRNA</td>
<td>Transfer RNA</td>
</tr>
<tr>
<td>RNase</td>
<td>Ribonuclease</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium dodecyl sulfate</td>
</tr>
<tr>
<td>TEMED</td>
<td>N, N, N',N'-tetraethylenediamine</td>
</tr>
<tr>
<td>Tris</td>
<td>Tris(hydroxymethyl)aminomethane</td>
</tr>
<tr>
<td>XC</td>
<td>Xylene cyanol</td>
</tr>
<tr>
<td>Xgal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactoside</td>
</tr>
</tbody>
</table>
MATERIALS AND METHODS

Chemicals

All chemicals were obtained from commercial sources and were of reagent grade. Special chemicals were obtained as follows: agarose (Type II, medium EEO), ammonium persulfate, ampicillin, BSA, DNase I, Ficoll (MW 400,000), MOPS (pKa=7.2, pH range 6.5–7.9), PIPES (pKa=6.8, pH range 6.1–7.5), polyvinylpyrrolidone (MW 360,000), proteinase K, RNase A, SDS, tetracycline and Tris from Sigma; AMV reverse transcriptase, BamHI (lyophzyme), EcoRI (lyophzyme), formamide, Klenow fragment (large fragment of E. coli DNA polymerase I), low melting point agarose, Smal and Xgal from Bethesda Research Laboratories; AccI, BglII, HindII, HindIII, HpalI, KpnI, PstI, SalI, Sau3AI, TaqI, E. coli DNA polymerase I and T4 DNA ligase from New England Biolab; dATP, dCTP, dGTP, dTTP, ddATP, ddCTP, ddGTP, ddTTP, M13 single-stranded primers; (15mer: 5'–d[CCCAGTCACGACGTT]–3', 17mer: 5'–d[GTAAAACGACGGCCAGT]–3'), and nuclease S1 (Aspergillus oryzae) from Pharmacia P–L Biochemicals; acrylamide, bis-acrylamide and mixed bed resin (AG501–X8D, 20–50mesh) from Bio–Rad; Gene Screen, T4 polynucleotide kinase and radioisotopes ([α-32P]–dATP, [α-32P]–dCTP, [γ-32P]–ATP; specific activity =3000 Ci/m mole) from New England Nuclear; nitrocellulose sheets (HAHY, 0.45 μm) from Millipore; nitrocellulose circles BA85 (0.45 μm, 82 mm) from Schleicher & Schuell; ultrogel AcA54 (5,000–70,000) and PEG 8000 from Fisher Scientific; dATP from Terochem Laboratories; TEMED from Eastman; urea from Western Scientific.
Bacterial strains and growth conditions

The characteristics of bacterial strains and bacteriophages employed are listed in Table 1. The *E. coli* strain MC1000, used as a general recipient for plasmids in the course of the investigation, was cultured at 37°C in NY broth or M9 minimal medium supplemented with 0.2% glucose, 0.01% vitamin B1 and 0.8% casamino acids (19). *E. coli* strains NS428 and NS433, used for the preparation of packaging extracts, were cultured at 32°C in M9 medium supplemented with 2% casamino acids, 0.2% glucose and 2 mM MgSO₄ as described by Sternberg et al. 1977 (130). *E. coli* strains Q358 and Q359, hosts for wild type phage λ1059 and λ1059-*H. cutirubrum* recombinants respectively, were cultured at 37°C in NZYC medium (57). Bacteriophage M13 hosts, the *E. coli* strains JM101 and JM103, were cultured at 37°C in YT medium for the screening of recombinant phage, in 2xYT medium for the preparation of ssDNA, and in M9 medium supplemented with 0.2% glucose and 0.01% vitamin B1 for storage at 4°C (86).

*Halobacterium cutirubrum*, a gift from A. Matheson, was cultured at 37°C in low phosphate complex medium as described by Nazar et al. 1978 (92).

Purification of *Halobacterium cutirubrum* chromosomal DNA

*H. cutirubrum* cells, grown in 500 ml low phosphate complex medium at 37°C to stationary phase, were harvested (7,000xg, 15 min., 4°C), resuspended in 20 ml of complex medium and lysed with 2 ml of 70 mM sodium deoxycholate (30 min., 4°C). The viscous lysate was diluted using 70 ml of 0.1 M Tris pH 7.5, 10 mM EDTA, and extracted successively with
Table I: Genotypes of *E. coli* and phage \( \lambda \) strains.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Genotype</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>MC1000</td>
<td><em>araD139</em>, (araBCO/C leu)7697, (lacIOPZY)X74, gaiU, gaiK, strA, thi</td>
<td>(19)</td>
</tr>
<tr>
<td>NS428</td>
<td>N205(( \lambda )Aam11 b2 red3 clts857 Sam7)</td>
<td>(130)</td>
</tr>
<tr>
<td>NS433</td>
<td>N205(( \lambda )Eam4 b2 red3 clts857 Sam7)</td>
<td>(130)</td>
</tr>
<tr>
<td>Q358</td>
<td>( hsdRk^- ), ( hsdMk^- ), supF(^+), ( \Phi80 )</td>
<td>(57)</td>
</tr>
<tr>
<td>Q359</td>
<td>( hsdRk^- ), ( hsdMk^- ), supF(^+), ( \Phi80 ), P2</td>
<td>(57)</td>
</tr>
<tr>
<td>JM101</td>
<td>( \Delta )lacpro, supE, thi, F'traD36, proAB, ( lac^Q ) Z M5</td>
<td>(86)</td>
</tr>
<tr>
<td>JM103</td>
<td>( \Delta )lacpro, supE, thi, F’traD36, proAB, ( lac^Q ) Z M5, strA, endA, hsdR4</td>
<td>(86)</td>
</tr>
<tr>
<td>( \lambda )1059</td>
<td>( h \lambda ) sbam1(^{10} ) b189 &lt;int29 ninL44 clts857 pACL29&gt; ( \Delta ) [int-c111] KH54</td>
<td>(57)</td>
</tr>
<tr>
<td></td>
<td>sRI4(^{10} ) nin5, Chi D</td>
<td></td>
</tr>
</tbody>
</table>
equal volumes of phenol, chloroform/octanol (1:1, v/v), and chloroform. The DNA-containing aqueous layer was dialyzed against 2 liters of 10 mM Tris pH 7.5, 1 mM EDTA (16 hrs., 4°C), and purified through a cesium chloride gradient (117).

Construction of \( \lambda 1059-H. \) cutirubrum library

Partial endonuclease digestion conditions were established for aliquots (10 \( \mu \)g) of \( H. \) cutirubrum DNA for the restriction enzyme Sau3AI. Reactions were performed for 15 min. at 37°C with 0.5, 1 and 2 times the estimated amount of enzyme (0.7, 1.3 and 2.5 units respectively) required to yield the maximum proportion of 15–25 kb fragments (Sau3AI buffer used was 50 mM NaCl, 6 mM Tris pH 7.5, 5 mM MgCl\(_2\), 100 \( \mu \)g/ml BSA). The digests were pooled and electrophoresed on a 0.6% low melting point agarose gel, and DNA fragments of 15–25 kb were isolated (67). The bacteriophage \( \lambda 1059 \), used as a vector for constructing the library (67), was digested with restriction enzyme BamHI. The fragments of \( \lambda 1059 \) were ligated to partial Sau3AI fragments (15–25 kb) of \( H. \) cutirubrum total DNA (2 \( \mu \)g and 1 \( \mu \)g respectively) with 0.1 unit of T4 DNA ligase in 20 \( \mu \)l of ligation buffer (50 mM Tris pH 7.8, 10 mM MgCl\(_2\), 20 mM DTT, 1 mM ATP, 50 \( \mu \)g/ml BSA) for 16 hrs. at 12°C, and ligated DNA was recovered by \textit{in vitro} \( \lambda \) phage packaging.

The packaging extracts made from \( E. \) coli strains NS428 and NS433 (130) were a gift from Rashmi Kothary. \textit{In vitro} packaging reactions were carried out based on the methods of Hohn 1979 (46). To amplify the library, aliquots of the packaged mixture containing \( 5 \times 10^3 \) recombinant bacteriophage were preadsorbed with 0.2 ml of a fresh overnight culture of \( E. \) coli Q359 (10 min., 37°C), mixed with 3 ml of NZYC top agar and
plated on five 100 mm diameter plates (14 hrs., 37°C). Each plate was overlaid with 5 ml of SM buffer (10 mM Tris pH 7.5, 0.1 M NaCl, 0.02% gelatin) for 16 hrs. at 4°C. The SM buffer containing bacteriophage was pooled (25 ml), and 1 ml of chloroform was added. Cell debris and agar were removed by centrifugation (6,000xg, 5 min.). The phage supernatant was recovered and stored over chloroform at 4°C (78).

Purification of ribosomal RNAs

_H. cutirubrum_ cells, grown to stationary phase in 20 ml of complex medium at 37°C, were harvested (7,000xg, 5 min., 4°C) in a SS-34 rotor, and lysed with 10 ml of SDS-lysis buffer (1% SDS, 0.14 M NaCl, 0.5 M NaOAc pH 5.1, 1 mM EDTA). The lysate was extracted with phenol, phenol/chloroform (1:1, v/v), and chloroform. The nucleic acids in the aqueous layer were precipitated in 0.3 M NaOAc pH 7.0 with 3 volumes of ethanol (1 hr., -70°C). The pellet was recovered by centrifugation (10,000xg, 20 min., 4°C) and redissolved in 1.5 ml of 0.1 M NaOAc pH 5.0. Ribosomal RNAs (16S and 23S) were separated from DNAs and small RNAs by preferential precipitation in 1.5 ml of 5 M LiCl (4 hrs., -20°C).

Synthesis of ribosomal cDNA

Oligonucleotides for priming the synthesis of ribosomal cDNA from the 16S and 23S rRNAs were generated by digesting 1 mg of _H. cutirubrum_ DNA with 14 μg of DNase I in 0.2 ml of DNase I buffer (10 mM Tris pH 7.4, 10 mM MgCl₂). After incubation for 2 hrs. at 37°C, the DNase I was inactivated by phenol/chloroform extraction. The DNA was ethanol precipitated, resuspended into 500 μl of 10 mM Tris pH 7.5, 1 mM EDTA, and autoclaved (10 min., 121°C).
Conditions used for the synthesis of ribosomal cDNA were derived from those of cDNA synthesis used in oligo-dT and random oligonucleotide priming (107, 132). Reactions were carried out in 20 μl volume containing: 2 μg rRNA, 2 μg *H. cutirubrum* oligodeoxynucleotides, 50 mM Tris pH 8.3; 50 mM KCl, 10 mM MgCl₂, 20 mM β-mercaptoethanol, 1 mM of each dATP, dGTP and dTTP, 10 μM dCTP, 0.8 μg Actinomycin D, 50 μCi [α-³²P]-dCTP (sp. act. 3000 Ci/mmole), and 50 units of AMV reverse transcriptase. The reaction was incubated first for 15 min. at 15°C, and then for 2 hrs. at 37°C. Ribosomal RNAs were hydrolyzed with 0.3 volumes of 1 M NaOH (16 hrs., 37°C). The solution was neutralized with 0.1 volumes of 3 M NaOAC pH 4.8.

The ribosomal cDNA was separated from radioactive nucleotides by chromatography on an acrylamide/agarose column (AcA54; 0.7cmx18cm) equilibrated with 0.2 M NaCl, 10 mM Tris pH 7.5, 0.25 mM EDTA. The radioactive DNA mixture was layered on the top of the column, eluted with the same column buffer and collected in 0.5 ml fractions. The amount of radioactivity was measured by Cerenkov counting. The radioactive fractions in the excluded volume containing purified DNA were pooled and ethanol precipitated (sp. act. 3x10⁶ cpm/μg).

Screening of the library for ribosomal RNA genes

Approximately 500 plaques were plated onto each 100 mm petri dish of NZYC agar, and were transferred onto nitrocellulose filters as described by Benton and Davis 1977 (11) with some modification. A nitrocellulose filter (0.45 μm, 82 mm), numbered with a soft pencil, was placed carefully onto the surface of the top agar containing the plaques (1 hr., 4°C). The filter was removed, transferred onto a fresh NZYC agar plate with the
plaque side faced upward, and the bacteriophages were amplified by incubating the plate for 16 hrs. at 37°C. The plaques were lysed by immersing the filter, plaque side up, in 1.5 M NaCl, 0.5 M NaOH for 20 min. The filter was neutralized in 1 M Tris pH 7.4 (20 min.), rinsed in 0.5 M Tris pH 7.5, 1.5 M NaCl (20 min.), and air dried. The DNA was immobilized by baking the filter for 2 hrs. at 80°C.

The nitrocellulose filter, wetted in 2xSSC (0.3 M NaCl, 0.03 M sodium citrate pH 7.0), was prehybridized in 3 ml of 6xSSC, 2xDenhardt’s solution (0.04% BSA, 0.04% Ficoll, 0.04% polyvinylpyrrolidone) in a sealed plastic bag (2 hrs., 68°C). Hybridization was performed in 3 ml of 6xSSC, 2xDenhardt’s solution, 1 mM EDTA, 0.5% SDS (16 hrs., 68°C) using radioactively labeled ribosomal cDNA as a probe (10⁶ cpm/filter, sp. act. 3×10⁶ cpm/μg). Washes were carried out in 200 ml volumes: twice in 2xSSC (5 min., 20°C), three times in 1xSSC, 0.5% SDS (30 min., 68°C), and twice in 2xSSC (5 min., 20°C). All prehybridization, hybridization and washes were done in a water bath with constant agitation. The filter was dried and autoradiography was carried out using Kodak XRP-1 X-ray film with an intensifying screen (16 hrs., −70°C).

The bacteriophage plaques, corresponding to positives on the autoradiogram (figure 4), were picked using a sterile pasteur pipette and inoculated into 0.5 ml of SM buffer (2 hrs., 37°C). The phage solution was titred and purified by successive rescreening with ribosomal cDNA probes as described until -90% of the plaques on a plate gave positive signals after screening.
Figure 4: Plaque hybridization autoradiogram.

Approximately 500 plaques were plated onto each plate (100 mm in diameter). The plaques were transferred onto nitrocellulose filter and lysed with NaOH. Hybridizations and washes were performed as described by Benton and Davis (11). The probes used were radioactively $^{32}$P-labeled cDNAs for ribosomal RNAs 16S and 23S. Kodak XRP-1 X-ray film was put over the filter with an intensifying screen for 16 hrs. at $-70^\circ$C.
Purification of bacteriophage DNA

A purified bacteriophage plaque was picked and inoculated into a solution containing 100 μl of a fresh overnight culture of Q359 and 100 μl of SM buffer. Bacteriophages were allowed to adsorb as described above. The infected cells were added to 25 ml of prewarmed NZYC medium, and incubated at 37°C with vigorous shaking until lysis occurred (about 6–9 hrs.). A few drops of chloroform were added and incubation was continued for 15 min. Cellular debris was removed by centrifugation (10,000xg, 10 min., 4°C).

Bacteriophages were precipitated with 0.15 volumes of 5 M NaCl and 0.3 volumes of 50% PEG 8000 (161). The solution was mixed thoroughly and cooled (1 hr., 4°C). Phage particles, recovered by centrifugation (10,000xg, 10 min., 4°C), were resuspended into 0.5 ml DNase buffer (50 mM Tris pH 7.0, 5 mM MgCl₂, 0.05 mM CaCl₂), and were treated with 5 μg of DNase I and 50 μg of RNase A (30 min., 37°C). Nucleases were inactivated and bacteriophages were lysed by adding 100 μg of proteinase K and 50 μl of 10% SDS, 0.05 mM EDTA (1 hr., 37°C). The mixture was centrifuged and the supernatant was extracted with phenol/chloroform (1:1, v/v), and chloroform. Phage DNA was precipitated in 2.5 ml of 0.1 M NaCl with 3 volumes of ethanol (2 hrs., −20°C). DNA recovered by centrifugation (10,000xg, 20 min., 4°C), was washed with 5 ml of ethanol/0.1 M NaCl (2:1), and air dried. The pellet was resuspended in 200 μl of 10 mM Tris pH 7.5, 1 mM EDTA and stored at 4°C.
Preparation of a nick-translated probe

Nick translation was performed according to Rigby et al. 1977 (108) with some modifications. Each reaction volume was 50 µl containing: 0.5 µg DNA, 50 mM Tris pH 7.5, 5 mM MgCl₂, 10 mM β-mercaptoethanol, 0.2 mM CaCl₂, 2 µM of each dCTP and dTTP, 8 µM dGTP, 0.7 µM dATP, and 50 µCi [α-³²P]-dATP (sp. act. 3000 Ci/m mole). After the addition of 5 µl of DNase I (0.05 µg/ml) and 10 units of *E. coli* DNA polymerase I, the mixture was incubated for 1 hr. at 16°C. The reaction was stopped by adding 1 µl of 0.5 M EDTA and extraction with phenol/chloroform. The nick-translated DNA was purified by column chromatography as described for cDNA (sp. act. 20–40×10⁶ cpm/µg).

Northern blot analysis

Ribosomal RNAs were electrophoresed through a 2% agarose-formaldehyde gel (40, 75). The gel was made up in 6% formaldehyde and 1xMOPS electrophoresis buffer (20 mM MOPS, 5 mM NaOAc, 1 mM EDTA pH 7.0). Samples were heated for 15 min. at 60°C in 50% formamide, 6% formaldehyde, 1xMOPS before loading onto the gel. To hydrolyze the rRNAs, the gel was soaked in 50 mM NaOH (30 min., 20°C) and neutralized in 0.1 M Tris pH 7.0 (30 min., 20°C). After equilibrating the gel in 25 mM sodium phosphate pH 6.5 (20 min., 20°C), rRNAs were transferred onto a Gene Screen filter using the same phosphate buffer based on the manufacturer's instructions (New England Nuclear). The filter was air dried and baked (2 hrs., 80°C). Residual rRNAs in the agarose gel were visualized by staining the gel in 30 µg/ml acridine orange (85).
The Gene Screen filter was prehybridized in 3 ml of 50\% formamide, 5xSSC, 2xDenhardt's solution, 1% SDS, 500 \( \mu \)g/ml denatured calf thymus DNA (16 hrs., 42°C). Hybridization was carried out in 3 ml of 50\% formamide, 5xSSC, 1xDenhardt's solution, 1% SDS, 250 \( \mu \)g/ml denatured calf thymus DNA (16 hrs., 42°C). Nick-translated phage DNA, denatured with 0.1 volumes of 1 M NaOH (10 min., 68°C) and neutralized with 0.125 volumes of 2 M NaH\(_2\)PO\(_4\)_ was used as a probe (10\(^6\) cpm/filter; sp. act. 20\(\times\)10\(^6\) cpm/\(\mu\)g). Washes were carried out as follows: twice in 2xSSC (5 min., 20°C), twice in 2xSSC, 0.5% SDS (30 min., 65°C) and twice in 0.1xSSC (5 min., 20°C). The filter was dried and subjected to autoradiography.

**Southern blot analysis**

Restriction digests of phage or genomic DNA electrophoresed on a 0.6\% agarose gel were denatured, neutralized and transferred onto nitrocellulose (HAHY, 0.45 \( \mu \)m) based on the method described by Southern 1975 (124). The filter was baked (2 hrs., 80°C.), prehybridized in 6xSSC, 2xDenhardt's solution (2 hrs., 65°C), and hybridized with a radioactively labeled ribosomal cDNA or nick-translated probe (10\(^6\) cpm/filter) in 6xSSC, 2xDenhardt's solution, 0.5\% SDS, 1 mM EDTA (16 hrs., 60°C). Washes were carried out twice in 2xSSC (5 min., 20°C), three times in 1xSSC, 0.5\% SDS (45 min., 60°C) and twice in 2xSSC (5 min., 20°C). The filter was dried and subjected to autoradiography. For reduced stringency of annealing, hybridizations and washes were carried out at 60°, 52° and 44°C.

**M13 cloning and DNA sequencing**

All techniques concerning M13 cloning were carried out based on Messing (1983; Ref. 86). Fragments to be sequenced were digested with
restriction enzymes (Sau3A, HindII, TaqI and HpaII) and were cloned into M13 vectors (mp8, mp9, mp10 or mp11) digested with the appropriate restriction enzyme (BamHI, HindII, and AccI) (86, 87). Single-stranded phage DNA from each positive clone was isolated and sequenced using Sanger's dideoxy chain-termination method (111). Clones in different orientations and clones containing overlapping sequences in different orientations were identified by "C-Test" as described by Messing (1983; Ref. 86).

Maxam and Gilbert's chemical method for DNA sequencing was also performed on end-labeled fragments (82). Fragments were either 3' end-labeled using an $[\alpha-^{32}P]$-deoxyribonucleoside triphosphate and Klenow fragment of E. coli DNA polymerase I (27) or 5' end-labeled with $[\gamma-^{32}P]$-dATP and T4 polynucleotide kinase in phosphate exchange reactions (12). End-labelings were carried out at some convenient restriction sites, e.g. The BamHI site in the middle of the 16S rRNA gene and the PstI site at the end of the 5S rRNA gene.

**Nuclease S1 mapping**

Approximately 0.2 $\mu$g of the radioactively labeled fragment, 5' end-labeled or 3' end-labeled, was precipitated with ethanol with 5 $\mu$g of rRNAs, and the pellet was resuspended into 100 $\mu$l of hybridization buffer (80% formamide, PIPES pH 6.4, 0.4 M NaCl, 1 mM EDTA). The mixture was incubated for 15 min. at 85°C to denature the double-stranded DNA and then incubated for 3 hrs. at 55°C permitting RNA-DNA hybrids to form. Nuclease S1 (300 or 600 units) was added to the mixture together with 0.3 ml ice-cold S1 buffer (0.28 M NaCl, 50 mM NaOAc pH 4.6, 4.5 mM ZnSO$_4$, 20 $\mu$g/ml carrier ssDNA). The reaction, incubated for 30 min. at 37°C, was terminated by adding 50 $\mu$l of 4.0 M NH$_4$OAc, 0.1 mM EDTA. The solution
was extracted with phenol/chloroform and ethanol precipitated with 20 μg of carrier tRNA. The precipitate was resuspended into 5 μl of formamide stop (90% formamide, 20 mM EDTA, 0.03% XC, 0.03% BPB), boiled for 2 min., and electrophoresed on urea-polyacrylamide sequencing gels (31).

Restriction enzyme digestion

Restriction endonuclease digestion of purified DNA was carried out according to the manufacturer's instructions (New England Biolabs). The DNA samples were taken up in the appropriate enzyme buffer and aliquots of restriction enzymes were added. Digestion was carried out for 4-16 hrs. at 37°C with AccI, BamHI, BglII, EcoRI, HindII, HindIII, HpaII, KpnI, PstI, SalI, Sau3AI and SmaI, and at 65°C with TaqI. The reactions were terminated by the addition of 0.2 volumes of Ficoll stop mix (20% Ficoll, 50 mM EDTA, 0.01% BPB).

Agarose gel electrophoresis

For Southern blot and qualitative analysis of restriction fragments (>500 bp), electrophoresis was carried out on horizontal slab gels of 0.6% agarose (21cmx16cmx0.5cm). The 0.6% agarose gels contained 40 mM Tris pH 8.0, 20 mM NaOAc, 10 mM EDTA, 0.25 μg/ml ethidium bromide and were run at 100 volts until the bromophenol blue marker dye had moved approximately 10cm. The DNA bands, visualized under ultra-violet light, were photographed with a Polaroid camera.
Polyacrylamide gel electrophoresis

For qualitative analysis of small DNA fragments (<500 bp), purification of restriction fragments for DNA sequencing, S1 mapping, and molecular cloning, electrophoresis was carried out on 5% polyacrylamide slab gels (14cmx17cmx0.15cm) containing 89 mM Tris, 89 mM boric acid, 2.5 mM EDTA. The gels were run at 250 volts for an appropriate amount of time and stained with 1 µg/ml ethidium bromide. DNA bands were visualized and photographed as described above.

Recovery of DNA from polyacrylamide gels

DNA bands cut out from polyacrylamide gels were chopped into small pieces and transferred into a tube plugged with glass wool (made from a 10 ml plastic pipette which had the top half cut away and the tip attached to a dialysis bag). With the tube held in a vertical position, electroelution was carried out in 45 mM Tris, 45 mM boric acid, 1.25 mM EDTA for 4-16 hrs. at 2mA/tube. The DNA collected in the dialysis bag was recovered by ethanol precipitation.

Other molecular cloning techniques, such as: ligation of restriction fragments, transformation of *E. coli* with plasmid DNA, and small scale isolation of plasmid DNA were carried out according to the Molecular Cloning Manual by Maniatis *et al.* 1982 (77).
RESULTS

Part I: Localization of rRNA gene cluster in genomic λ1059 library.


The bacteriophage λ1059 is a BamHI substitution vector composed of three BamHI fragments—a 19.6 kb left arm, a 17 kb central fragment and a 9.4 kb right arm (56). The two arms of the vector contain all the essential functions required for λ replication and maturation. Since lambdoid phages require genomic sizes between 70% and 108% of the wild type DNA to fill the phage head properly, the central BamHI fragment of λ1059 can be replaced by DNA inserts in the size range of 6.3–24.4 kb. To avoid cloning of multiple unrelated ligated sequences, partial Sau3A1 digests of H. cutirubrum DNA were fractionated on a 0.5% low melting point agarose gel and fragment sizes corresponding to 15–25 kb were recovered by melting the agarose. The λ1059 DNA was cleaved with BamHI and 2 μg aliquots were religated with 1 μg of partial Sau3A1 fragments (15–25 kb) of H. cutirubrum.

Recombinant λ1059 phages are spi− because of the deletion of the λred and gamma genes and they can be separated from the spi+ vector phage by plating on E. coli strains lysogenic for bacteriophage P2. Nonrecombinant spi+ phage are capable of propagating on Q358, a non-restrictive strain, but they cannot grow on Q359 (a P2 lysogen of Q358), whereas recombinant spi− phage grow on both Q358 and Q359. After in vitro packaging, the total number of phage generated by the ligation reaction was 10^7/μg H. cutirubrum DNA when plated on Q358; the number of phages harboring H. cutirubrum inserts was 10^5/μg H. cutirubrum DNA.
when plated on Q359. The packaged recombinant phages were amplified approximately 1000 fold on Q359 and recovered as a plate lysate. Plating was done at low density (5,000 plaques per 100 mm diameter plate) to minimize the risk of eliminating particular recombinant-phage from the library during amplification.

According to Clarke and Carbon, 1976 (18), the number of independently derived phage recombinants (library size) required to find any given single-copy sequence in a library with a known average size of the DNA insert is represented by the equation:

\[
N = \frac{\ln (1 - p)}{\ln (1 - f)}
\]

where

- \(N\) = number of recombinants
- \(p\) = probability the library contains a given unique sequence
- \(f\) = average fraction of total genome in each insert fragment

The genome size of \(H.\) \textit{cutirubrum} is \(4.1 \times 10^6\) base pairs (88) and the average insert in each clone was 20 kb; therefore, approximately \(940 \times 10^5\) recombinant phage are required to represent 99% of the \(H.\) \textit{cutirubrum} genome.

Screening for rRNA genes in genomic library.

\(H.\) \textit{cutirubrum} 16S and 23S rRNAs were used as templates and fragmented \(H.\) \textit{cutirubrum} DNAs were used as primers for the synthesis of radioactive cDNA probes by AMV reverse transcriptase. The two rRNAs were prepared by precipitation from total cellular nucleic acid with 2.5 M LiCl in which small RNAs and DNA soluble. Although contaminated by a small fraction of messenger RNAs, greater than 95% of the precipitate was
16S and 23S rRNAs. Random primers of 8-12 nucleotides in length were generated by DNase I treatment of *H. cutirubrum* DNA (132) and were heat-denatured (10 min, 121°C) before each cDNA synthesis reaction.

Radioactively \(^{32}\text{P}\)-labeled ribosomal cDNA (specific activity of \(3 \times 10^6\) cpm/µg) was used to screen the genomic library by means of a plaque hybridization technique based on that of Benton and Davis, 1977 (11). Six independent positives were detected by screening 3,000 recombinants. Phage DNA from each positive was purified (16) and analyzed with restriction endonucleases *BamH*I, *EcoR*I and *Hind*III. Two different types of clones, \(\lambda\)Hc4 and \(\lambda\)Hc9, were identified. Four of the positives belonged to the \(\lambda\)Hc4 group and two belonged to the \(\lambda\)Hc9 group.

### Restriction analysis of \(\lambda\)Hc4 and \(\lambda\)Hc9

Restriction maps of \(\lambda\)Hc4 and \(\lambda\)Hc9 were constructed by digesting the phage DNAs with various combinations of restriction endonucleases (*EcoR*I, *Hind*III, and *BamH*I) followed by analysis on a 0.6% agarose gel (Figure 5). These three enzymes were chosen because of the sparsity of their recognition sequences within the arms of \(\lambda\)1059 DNA. There is only one *EcoR*I site and one *Hind*III site located about 3.6 kb and 4.4 kb respectively from the end of the right arm and *BamH*I does not cut within either of the two arms. Restriction sites in the *H. cutirubrum* inserts were then mapped accordingly. The *H. cutirubrum* inserts in each clone were about 17-20 kb in length. Of the four *BamH*I-Sau3A1 junctions from the two clones, only one regenerated a *BamH*I restriction site (in \(\lambda\)Hc4 at the right arm junction; see below). All other *BamH*I sites are within the insert.
Figure 5: Restriction endonuclease maps of λHc4 and λHc9.

The λHc4 and λHc9 DNAs were digested with restriction enzymes (Ε, EcoRI; H, HindIII; B, BamHI; HE, HindIII+EcoRI; BE, BamHI+EcoRI; BH, BamHI+HindIII; HBE, HindIII+BamHI+EcoRI), and fractionated by electrophoresis on a 0.6% agarose gel with wild type λ DNA cut with HindIII as molecular weight standards (the fragment sizes are 23.7, 9.6, 6.7, 4.3, 2.3 and 2.0 kb). Lane "a" represents DNA from λHc4 and lane "b" represents DNA from λHc9. (Note: The BamHI-EcoRI digest of λHc9 was incomplete). The restriction maps constructed were based on the restricted fragment sizes indicated on the ethidium bromide stained gel. The λ1059 arms, 9.4 and 19.6 kb in length, are indicated by black boxes; Halobacterium cutirubrum inserts, approximately 18 kb in λHc4 and 17 kb in λHc9, are indicated by open boxes. The relative locations of the ribosomal RNA genes, 16S and 23S, are indicated by short and long horizontal arrows respectively. The λHc4 clone contains both 16S and 23S rRNA genes whereas λHc9 contains partial 16S sequences.
Southern blot analysis of $\lambda$Hc4 and $\lambda$Hc9 DNAs

Phage DNAs, $\lambda$Hc4 and $\lambda$Hc9, containing the *H. cutirubrum* rRNA gene cluster were digested with restriction endonucleases (EcoRI, BamHI, HindIII, HindIII + BamHI, BamHI + EcoRI, HindIII + EcoRI) and fractionated on a 0.6% agarose gel (approximately 2–3 µg DNA/lane). The DNAs were transferred onto nitrocellulose filters and probed with radioactively $^{32}$P-labeled ribosomal cDNA (10⁶ cpm/filter; sp. act. 3x10⁶ cpm/µg). A comparison between the ethidium bromide staining and the autoradiograph indicated that the rRNA gene cluster was located on an 8.0 kb HindIII–EcoRI fragment plus some sequences beyond the EcoRI site in the $\lambda$Hc4 clone (Figure 5 and 6). In $\lambda$Hc9, the rRNA genes were located within the 10.9 kb HindIII fragment, of which 5.0 kb belonged to the $\lambda$1059 right arm (Figure 5). Since most weak bands on the autoradiograms are lower in molecular weights than strong bands, this could be the result of limited homology or poor representation of the cDNA probes.

With the knowledge of the relative position of the rRNA gene cluster on each phage clone, a closer examination of the two phage restriction maps suggested that the clones might be carrying overlapping partial Sau3A1 fragments from the *H. cutirubrum* genome. For example; The two 4.9 kb BamHI–HindIII fragments might contain identical sequences. To test this hypothesis, cross-hybridization experiments were carried out. Phage DNA digested with restriction endonucleases was fractionated and transferred onto nitrocellulose filters as indicated above. Hybridizations were carried out with nick-translated phage DNA (10⁶ cpm/filter; sp. act. 20x10⁶ cpm/µg) such that $\lambda$Hc4 (or $\lambda$Hc9) restricted fragments were probed with radioactively $^{32}$P-labeled $\lambda$Hc9 (or $\lambda$Hc4) DNA. The results shown in Figure 7 indicated that the $\lambda$Hc9 probe hybridized to all restriction fragments in $\lambda$Hc4.
Figure 6: Southern hybridization of ribosomal cDNA to \(\lambda\)Hc4 and \(\lambda\)Hc9 DNAs.

The DNAs from \(\lambda\)Hc4 and \(\lambda\)Hc9 were digested with various restriction enzymes (E, EcoRI; B, BamHI; H, HindIII; HB, HindIII+BamHI; BE, BamHI+EcoRI; HE, HindIII+EcoRI) and fractionated by electrophoresis on a 0.6% agarose gel. The fragments were transferred onto nitrocellulose and hybridized to radioactive \(^{32}\)P-labeled cDNA. The ethidium bromide stained gel and the corresponding autoradiogram are shown. Lane "a" corresponds to \(\lambda\)Hc4 DNA and lane "b" corresponds to \(\lambda\)Hc9 DNA.
Figure 7: Cross Southern hybridization of λHc4 and λHc9.

The λHc4 and λHc9 DNAs were digested with various restriction enzymes, electrophoresed on a 0.6% agarose gels, and transferred onto nitrocellulose as described in Figure 6. Panel A depicts λHc4 DNA probed with nick-translated λHc9 DNA; panel B depicts λHc9 DNA probed with nick-translated λHc4 DNA. The ethidium bromide stained gel and the corresponding autoradiogram are shown. The λHc9 DNA hybridized to all restricted fragments of λHc4 DNA, and λHc4 DNA hybridized to all but one 2.0 kb BamHI–EcoRI fragment of λHc9 DNA.
Similarly, the λHc4 probe hybridized to all but one fragment (2.0 kb BamHI–EcoRI) of λHc9. Some of the bands on the autoradiogram were very faint and this might be a result of poor transfer of the fragment or limited homology. Cross-hybridization results imply that the two phages indeed carried overlapping partial Sau3A1 fragments. The BamHI site located near the right arm of λHc4 was in fact the λ1059–H. cutirubrum junction whereas the BamHI site near the right arm in λHc9 is from the H. cutirubrum DNA located within the rRNA gene cluster.

Northern blot analysis of λHc4 and λHc9 DNAs

To identify which rRNA genes were present on each phage clone, Northern hybridization experiments were carried out. Ribosomal RNAs (5 μg of a mixture of 16S and 23S) were fractionated on a 2% formaldehyde-agarose gel, transferred onto “Gene Screen” membrane, and hybridized with nick-translated phage DNA (λHc4 or λHc9; Figure 8). Two rRNA bands, 16S and 23S, were detected when probed with λHc4 DNA and only the 16S band was detected with the λHc9 probe indicating that λHc4 contained both 16S and 23S rRNA genes whereas λHc9 contained the 16S rRNA gene only (Figure 5). The 16S rRNA gene had an internal BamHI restriction site and the 23S rRNA gene was located as least 1.0 kb from this BamHI site.
Figure 8: **Northern hybridization of λHc4 and λHc9 DNA to ribosomal RNAs.**

Ribosomal RNAs, 16S and 23S, were electrophoresed on a 2% agarose gel containing 6% formaldehyde. The RNAs were transferred to "Gene Screen" membranes and probed with nick-translated $^{32}$P-labeled phage DNAs. Panel A represents ribosomal RNAs probed with λHc4 DNA and panel B represents ribosomal RNAs probed with λHc9 DNA. The λHc4 DNA contains both 16S and 23S rRNA genes whereas λHc9 DNA contains only the 16S rRNA gene.
Part II; Sequence analysis of the rRNA gene cluster

Subcloning of the rRNA gene into plasmid pBR322 and M13 vectors

Specific DNA fragments containing the rRNA gene cluster of \textit{H. cutirubrum} from \textsc{XHc4} were subcloned into plasmid pBR322 for further analysis (Figure 9). The 5.0 kb \textit{HindIII-BamHI} fragment which contained the proximal portion of the 16S gene was cloned into the \textit{HindIII-BamHI} sites of pBR322 giving the recombinant plasmid p4S. Similarly, the 7.5 kb \textit{BamHI-KpnI} fragment, which contained the distal portion of the 16S gene, the 23S and 5S genes, and the 7.5 kb \textit{KpnI-BglII} fragment, which contained the complete cluster, were cloned into the \textit{BamHI-EcoRI} sites of pBR322 giving the recombinant plasmids p4L and p4W respectively. Since pBR322 has no \textit{KpnI} restriction site, a 160 bp \textit{KpnI-EcoRI} restriction fragment from \textsc{p701} (19) was utilized as a linker fragment in the constructions of p4L and p4W.

Restriction endonucleases (\textit{Sau3AI, TaqI, HpaII}) were used to subclone small fragments (<500 bp) from the plasmid clones into M13 vectors (mp8, mp9, mp10, mp11; Figure 10). The recombinant M13 DNAs were used to transform \textit{E. coli} strains JM101 or JM103. Positive clones were identified by clear color plaques on Xgal plates as opposed to the wild type blue color. However, in some cases where the inserts were small (<100 bp in length) and the number of nucleotides in the insert was a multiple of three, light blue positive plaques were observed. This was possibly due to the formation of a hybrid protein which was partially functional. These small clones were first identified by hybridization of their ssDNA with a radioactive probe. Clones overlapping each restriction site and clones in different orientations were identified by "C-test" (86). In a "C-test",
Figure 9: Plasmid subclones of λHc4 and λHc9.

The upper bar represents the *Halobacterium cutirubrum* chromosomal DNA segment which contains the rRNA gene cluster. The second and the third bar represent *H. cutirubrum* DNA inserts in λHc9 and λHc4 respectively. As indicated by Northern hybridization experiments, λHc9 contains part of the rRNA gene cluster whereas λHc4 contains the complete gene cluster. The λ DNAs which were subcloned into pBR322 plasmids, p4S, p4L and p4W, are shown by the lower three bars. Plasmid p4S contains a 5.0 kb *HindIII–BamHI* fragment cloned into *HindIII–BamHI* sites of pBR322. Plasmids p4L and p4W contain a 7.5 kb *BamHI–KpnI* fragment and a 7.5 kb *BglII–KpnI* fragment respectively and they were cloned into the *BamHI–EcoRI* sites of pBR322. The *KpnI* site from the 7.5 kb *BamHI–KpnI* fragment in plasmid p4L came from one of the λ1059 arms. Since pBR322 contains no *KpnI* sites, a *KpnI–EcoRI* (160 bp in length) fragment from p701 was used as a linker. Restriction enzyme sites (*B, BamHI*; *Bg, BglII*; *E, EcoRI*; *H, HindIII*; *K, KpnI*; *P, PstI*; *Sa, SalI*; *S, SmaI*) are indicated by arrows.
Figure 10: Sequencing map of the ribosomal gene cluster.

The organization of rRNA genes is shown by the top bar in black boxes from left to right 16S, 23S and 5S. The three 5' flanking direct repeats are indicated by arrows (►) and the two tRNAs are indicated by shaded boxes. The first nucleotide in the 16S gene is numbered +1; sequences preceding the 16S gene have negative values and sequences after the gap in 23S gene are prefixed by D. Arrows underneath indicate the direction and length of sequences obtained from either chemical or dideoxy sequencing. Fragments from this rRNA gene cluster are digested with Sau3A\textsubscript{I}, TaqI, HpaII or HindII (H2) and cloned into M13 vectors (mp8, mp9, mp10, mp11) cut with BamH\textsubscript{I}, AccI or HindII. Clones contain overlapping fragments and clones in opposite orientations were sequenced using Sanger's dideoxy method (111). The Maxam and Gilbert chemical sequencing method (82) was also employed with purified 3' or 5' end-labeled fragments, e.g. at PstI (P) and SalI (Sa) sites.
single-stranded phage DNA from two different clones are allowed to hybridize. Clones containing sequences in opposite orientations could be identified because they hybridize and form a figure-eight like structure which migrates more slowly than individual non-hybridized circles on an agarose gel.

**DNA sequences of the rRNA gene cluster**

The complete rRNA gene cluster except for the -2600 bp gap in the middle of the 23S rRNA gene was sequenced. Both chemical and dideoxy sequencing methods were used. The approximate length of DNA sequences obtained from different strands of M13 clones or from purified fragments are illustrated in Figure 10. The order of the genes, 5' to 3' within the cluster, was found to be 16S, 23S and 5S. The first base in the 16S rRNA gene was numbered +1, the 5' sequence flanking the 16S rRNA gene was assigned negative numbers and the sequences distal to the gap in the 23S rRNA gene were given a D prefix.

More than 900 nucleotides preceding the 16S rRNA gene were sequenced (Figure 11). Three copies of a bipartite direct repeat unit located at nucleotide -551 to -504, -418 to -371, and -285 to -238, were found in this region. These units contain two repeat segments of 27 and 8 nucleotides in length separated by 13 nucleotides. Each repeat unit is separated from the adjacent one by 85 nucleotides. Upstream from these three repeat units, there are two vestigial copies of the repeat units. The copy located at -670 to -646 has an interruption in its long repeat segment and the other copy located at -757 to 750 has only the short repeat segment. The spacing between these vestigial remains is also altered (see Figure 11).
Figure 11: The 5' flanking sequence of the ribosomal RNA gene cluster.

The three highly conserved and the two partially conserved direct repeat units are highlighted (•). Within each repeat unit the putative promoter sequence AAGTAA is indicated (⦁). The 5' portion of the inverted repeat sequence surrounding the 16S gene is identified by over-underlining. The sequence begins at -909 and ends at -1 which is the nucleotide immediately preceding the 16S rRNA gene.
The 5' flanking sequence of the 16S rRNA gene

CGGCCCGACCGAACGCACTCGCGCGGATGA'CCGGCCGACCTCCGCCTACGCAATACGCTGTGGCGTGTGTCCCTGGTGTGGGCCGCCATCACGAAGCGCT

GGT ACC A C T

GGTACCACT

GGTGGCCCTGGTAATCCGGCTTCGTCCGGGTGGCTGATGCATCTCTTCGACGCTCTCCATGGTGTCGGTCTCACTCTCAGTGAGTG TGAT TCGATGCCCTTAAGT
The 16S rRNA gene of *H. cutirubrum*, located by sequence comparison with that of the 16S rRNA from a related organism *Halobacterium volcanii* (42), is 1472 nucleotides length. The comparison of 16S rRNA between the two species of *Halobacterium* illustrated in Figure 12 shows 88% homology with 178 base substitutions, including a single base pair deletion between nucleotide 327–328 and an insertion at nucleotide 1078.

The mature 5' and 3' ends of the 23S rRNA were determined by nuclease S1 mapping (Figure 13). A Sal-I-TaqI fragment of 205 bp in length (nucleotides 1969–2174), 5'-labeled with T4 polynucleotide kinase and [γ-32P]-dATP, was used to map the 5' end of the mature 23S rRNA. A fragment of 193±4 nucleotides was protected from S1 digestion. This placed the 5' end of 23S rRNA gene at nucleotide 1982±4 (Figure 14). To determine whether the sequence preceding the putative mature 5' end of 23S rRNA encodes a 5.8S-like rRNA (29, 147), nuclease S1 analysis was performed on two other 5' end-labeled fragments, a TaqI fragment (nucleotide 1846–1939) and a BamHI-SalI fragment (nucleotide 684–1969); however, no radioactivity was protected by rRNAs (results not shown) indicating that sequences preceding the mature 5' end of 23S rRNA are not preserved in the ribosome. To determine if there is any nucleolytic cleavage within the 5' portion of the 23S downstream from the TaqI site at nucleotide 2174 to liberate a 5.8S-like rRNA, a 1.0 kb SalI fragment extending from nucleotide 1982 and towards the middle of the 23S gene was used in an S1 nuclease protection experiment. The 5' end-labeled 1.0 kb SalI fragment was protected by rRNAs to give a fragment of similar size, whereas the same 3' end-labeled fragment was not protected (results not shown). These results corroborate with the ones described above and
Figure 12: The 16S ribosomal RNA gene sequence.

The sequence is numbered from 1 to 1472 and is compared to the 16S rRNA sequence of the related organism, *H. volcanii*. The base substitutions are indicated by arrows (ψ). The *H. cutirubrum* contains a single base pair deletion between nucleotides 327–328 and an insertion at nucleotide 1078.
Figure 13: Nuclease S1 mapping of the 5' and 3' end of mature 23S rRNA.

The 5' end of mature 23S rRNA (panel A) was located using a 205 bp 5' end-labeled Sal-I-TaqI fragment (nucleotides 1969-2174). The S1 resistant fragment was estimated to be 193±4 nucleotides in length which positions the 5' end of the 23S gene at about nucleotide 1982±4. The 3' end of mature 23S rRNA (panel B) was located using a 76 bp 3' end-labeled TaqI fragment (nucleotides D1-D77). The S1 resistant fragment was estimated to be 49±4 nucleotides in length which positions the 3' end of the 23S gene at nucleotide D49±4. The upper and lower arrows indicate the positions of the full length DNA fragments and the S1 resistant fragments respectively. The three lanes in A and B are the full length fragment and the protected fragment digested with 300 or 600 units of S1 nuclease.
Figure 14: The intergenic space between the 16S and 23S rRNA genes.

The 3' end of the 16S rRNA gene (underlined) is at nucleotide 1472, and the 5' end of the 23S gene (underlined) is at nucleotide 1982±4. The 3' portion of the inverted repeat surrounding the 16S rRNA gene and the 5' portion of the inverted repeat surrounding the 23S gene are over-underlined. The putative alanine tRNA gene (nucleotides 1575-1651) is highlighted ( * ).
The intergenic space between the 16S and 23S rRNA genes

1410 1420 1430 1440 1450 1460 1470 1480 1490 1500
AATCTGGCTCCGCAAGGGGATTAAGTCAACAAGGTAGGGGAATCTGCGGCTGGATCACCTCCT

3' end of the 16S rRNA gene

1510 1520 1530 1540 1550 1560 1570 1580 1590 1600
CTCACCGAACACCGGCTGCGCCAGTGCCGACTTACAAACCATCAAGGCTAACAT

16S inverted repeat

1610 1620 1630 1640 1650 1660 1670 1680 1690 1700
TGCCTCTTTGCAAGGAGATGGCTGATCACTCGGCTGGATCAGTGGGAGACGGGGCAA

 alanine tRNA gene

1710 1720 1730 1740 1750 1760 1770 1780 1790 1800
CGATGAAATCAGCAAGCAGACTGATGCACCACTGTCGATTAGGGGCTACCTGGCCT

1810 1820 1830 1840 1850 1860 1870 1880 1890 1900
GTGTAAGTGGCAATCCAGCGCCCTGACTGCCAGCTGACGTGGCAGTGATGCTGGCTG

23S inverted repeat

GTFAGCGATTGTACGTGCTGGTGTTACAGGCAACATCTGAGACATCTGAGTGGTGAATCGATGACATCTCGAGACTTCGAGAGA

5' end of 23S rRNA gene

2010 2020 2030 2040 2050 2060 2070 2080 2090 2100
GTGGATAGCTGCTGAGGTGGCCCGCAACTCGCTGAGGCAGGTGGCCCGAGCGGCTAAGAACTGAGTCCACCTGCTGCTG

2110 2120 2130 2140 2150 2160 2170
TCCCTATACAAATCTGCTTGGCAATGGGGAACGGCAGCAATTGAAAGCAATTCATGCTGAGAGAAGACGAGAATCGA
indicate that the 5' end of mature 23S is located just beyond the Sa/I site (nucleotide 1969). There is no nucleolytic cleavage within its 5' end to give rise to a 5.8S–like RNA. The mature 3' end of the 23S rRNA was also determined by an S1 experiment using a Taq fragment of 76 bp in length. The fragment was 3' end-labeled with the Klenow fragment of E. coli polymerase I and [α-32P]dCTP. A fragment of 49±4 nucleotides was protected, placing the mature 3' end of 23S rRNA at nucleotide D49±4 (Figure 15).

The 123 nucleotide long 5S rRNA gene was found to start at nucleotide D162, about 112 nucleotides from the mature 3' end of the 23S rRNA. The 5S DNA sequence agrees with the published 5S rRNA sequence (33, 92).

Part III: Single copy rRNA genes

To determine the number of copies of each ribosomal RNA gene in H. cutirubrum, genomic Southern hybridizations were carried out. Radioactively 32P-labeled nick-translated restriction fragments from within the cloned rRNA gene cluster (as shown in Figure 16) were used to probe H. cutirubrum genomic DNA digested with a number of restriction endonucleases (BamH1, BglII, EcoRI, HindIII, KpnI, PstI and Smal). The probes were: a 484 bp Sau3AI fragment from the 5' end of the 16S gene, a 340 bp EcoRI-PstI fragment from the middle of the 23S gene and a 1.4 kb PstI fragment which contained the 3' end of the 23S gene and the 5S gene.

Hybridization results are illustrated in Figure 17. These hybridizations were carried at moderately high stringency: hybridizations were performed at 60°C in 6xSSC, and washes were at 60°C in 1xSSC. All probes
Figure 15: The 3' flanking sequence of the 23S rRNA gene.

The sequences extend from nucleotide D1 to nucleotide D608. The 3' end of the 23S rRNA gene (underlined) is at nucleotide D49±4. The 5S rRNA gene (underlined) is between nucleotides D162–D284. The 3' portion of the inverted repeat surrounding the 23S rRNA gene is over-underlined and the putative cysteine tRNA gene (nucleotides D395–D469) is highlighted (*). In addition, two sequences which may be related to transcript termination, an inverted repeat (nucleotides D303–D332) and a G/C rich region (nucleotides D544–D589) are indicated. The A/T rich region is found between nucleotides D589–D608.
The 3' flanking sequence of the 23S rRNA gene

D10  D20  D30  D40  D50  D60  D70  D80  D90  D100
TCGAGCAACGGAGACGTTAGCGCCCGGAGTACTAACAGTGAATGCCACACGACTCATGCACCTCACCACATAACCTTCGAGTCGGTCCAGGCGGTAT

3' end of the 23S rRNA gene

D110 - D120  D130  D140  D150  D160  D170  D180  D190  D200
GCCATACACAGGCGGATCGCGCCGACGTTGCGGCTACAGGTTTCGATTGCTGATTATAAGGGCAGCATACGGGTGCTGGTTACTCCCGTACCCA

23S inverted repeat

D210  D220  D230  D240  D250  D260  D270  D280  D290  D300
CCCGAACGAGAAAGTAAGGGGCCTGGGTCCGGTCTACAGGATCGCGAGCCCTCAGGGAAAATCGCGGCTTCCGCGCTACTTCTACTCTGAT

D310  D320  D330  D340  D350  D360  D370  D380  D390  D400
ATGCTGGAGAGACGGTGGTGTGCTGCGGCTTGCAGGTTATGAGACCGTACGTTCATGAGCGGCGCGAGTCGTGATGCTGCGCAG

------ -----------------------
inverted repeat

D410  D420  D430  D440  D450  D460  D470  D480  D490  D500
GTGGCGAGTGGCTTCTACAGGGCCTCGAGGCTCATCGCGGCTTCAATCGCGCGCTTTGCTTCAGCATCGCAGAGTCGACTGGATGTGGATGGCGAA

cysteine tRNA gene

D510  D520  D530  D540  D550  D560  D570  D580  D590  D600
CACAGTCGCCAGTGCTGAGGCGGAGATCGAGTGCTCCGCGCTCCGGGACTTACCGGAGGCCTACG

G/C regions

TATACGCA
Figure 16: Location of the fragments used for genomic Southern hybridization with respect to the rRNA gene cluster.

The relative positions of the 16S, 23S and 5S rRNA genes and two putative tRNA genes (T) are indicated. The solid boxes at the end of the 16S and the 23S rRNA genes represent long nearly perfect inverted repeats; the solid box between the 5S gene and the distal tRNA gene represents a short inverted repeat sequence. Identical direct repeat sequences in the 5' flanking region are indicated by a solid box with overhead arrows. The four bars at the bottom represent restriction fragments used to probe genomic DNA for related sequences. The fragments are: HindII–Sau3AI fragment (115 bp) carrying a single copy of the direct repeat unit; Sau3AI fragment (480 bp) from within the 16S gene; an EcoRI–PstI fragment (340 bp) from within the 23S gene and a PstI fragment (1.4 kb) carrying the distal end of the 23S gene, the entire 5S gene and a portion of the tRNA gene.
Figure 17: Genomic Southern hybridization with probes from within the rRNA gene cluster.

Genomic *H. cutirubrum* DNA was digested with a variety of restriction endonucleases (BgE, BglII+EcoRI; BE, BamHI+EcoRI; E, EcoRI; B, BamHI; BH, BamHI+HindIII; S, SalI; P, PstI), fractionated by electrophoresis, transferred onto nitrocellulose and probed with nick-translated $^{32}$P-labeled restriction fragments. Both hybridizations and washes were carried out at 60°C. The location of the 16S, 23S and 5S probes in the rRNA gene cluster is depicted in Figure 16. The panel labeled 16S+23S was probed with a mixture of two restriction fragments; the double bands in this panel indicate the presence of a restriction site between the two probing fragments. The position of size markers ($\lambda$ DNA cut with HindIII) are indicated by arrows and are from top to bottom: 23.7, 9.6, 6.7, 4.3, 2.3, 2.0, and 0.6 kb.
hybridized to only a single fragment of the restricted genomic DNA; for 
*KpnI* and *BglII* the hybridizing genomic fragments are greater than 25 kb in 
size (results not shown). Because of the respective position of the *KpnI*
and *BglII* sites 5' and 3' to the rRNA cluster (Figure 9), these results imply 
that rRNA genes are unique and single copy components of the *H. cutirubrum* genome. If the rRNA genes are multicopy, they must be within 
highly conserved repeat units greater than 50 kb in size.

Genomic Southern hybridizations were also carried out with the 115 
bp *HindII-Sau3Al* fragment which contained a direct bipartite repeat unit 
upstream from the 16S rRNA gene (Figure 18). The location of this probe 
with respect to the rRNA gene cluster is shown on Figure 16. To determine 
if related sequences were present elsewhere on the genome, hybridizations 
and washes were carried out at both high and reduced stringency at 60°C, 
52°C and 44°C. Although there was an increase in non-specific 
hybridization to *H. cutirubrum* DNA as well as λDNA (the molecular weight 
marker), no apparent specific hybridizations were observed. This indicates 
the probe sequence is probably single copy DNA unique to the rRNA gene 
cluster.
Figure 18: Genomic Southern hybridization probe with fragment containing a 5' flanking repeat.

Genomic *H. cutirubrum* DNA was digested with various restriction enzymes, fractionated by electrophoresis and transferred onto nitrocellulose as described in Figure 17. The filters were probed with nick-translated $^{32}$P-labeled 115 bp *HindII–Sau3AI* fragment (location of probe with respect to the rRNA gene cluster is indicated in Figure 16). Hybridizations and washes were carried out at 60°C, 52°C and 44°C. The size markers as described in Figure 17 are indicated by arrows.
DISCUSSION

Direct bipartite repeat sequence

The 5' flanking region of the ribosomal RNA gene cluster contains three perfectly conserved bipartite direct repeat units at nucleotides -551, -418 and -285. The two segments within each unit are 27 and 8 nucleotides in length (Figure 19). The spacing between the repeats is precise; the short spacer within each unit is 13 nucleotides and the long spacer between the units is 85 nucleotides. There is a considerable degree of conserved sequence homology (36-50%) even within the spacer region (2). The region upstream from the three direct repeat sequences appears to contain the vestigial remains of two additional repeat sequences. In these sequences, there has been both a partial loss of sequence homology within the repeat segments and a change in the spacing between the repeat units.

Typical eucaryotic or eubacterial consensus promoter sequences are not found in regions believed to act as promoter sequences in the Halobacterium genome (21). Comparison of the well characterized promoter region of the bacteriorhodopsin gene and the 5' sequences flanking the large open reading frames in two separate transposable sequences from H. halobium has revealed a hexanucleotide sequence, AAGTTA, positioned between 28-42 nucleotides upstream from the ATG translation start codon. This hexanucleotide is conserved in at least five of the six positions in all three promoters (25, 28, 122, 160). The related sequence, AAGTAA, also conserved at five of the six positions, is present in the long segment of the three direct repeat units preceeding the rRNA gene cluster. In addition, the long segment of the three units is AT rich (66% AT) relative to the genomic DNA (34% AT; Ref. 89) and this may be an important component
Figure 19: Direct repeat units in the 5' flanking sequence of the 16S rRNA gene.

A homology comparison of sequences within and surrounding the direct repeat units is presented. Homologous bases (•) and the two highly conserved segments of each bipartite repeat (over-underline) are indicated. The line length within the three highly conserved repeat was 133 nucleotides; within the two upstream imperfect units the line length was decreased to achieve maximum homology. The nucleotide numbers at the beginning and end of each line are indicated.
THE 5' FLANKING SEQUENCE OF THE 16S rRNA GENE

-909  GGTACACTCGGCGACCAGCAGCAGCTCGCGGGATGACCAGCGGACCTCAGCTAAGCTGGTGTCCTCCG -825
-824  TGGGCGCCCACATACGAGCCTGGCTGTTATGACCCAATCGCGGGGATGACCGGCCGACCTCCGCCTACGCAATACGCTGTGGCGTGTGTCCCTGG  -824
-700  TGTGGGCCGCCATCACGAAGCGCTGCTGGTTCGACGGTGTTTTATGTACCCCACCACTCGGATGAGA
-699  TGCGAACGACGTGAGGTGGCTCGGTGCACCCGACGCCACTGATTGACGCCCCCTCGTC
-579  CCGTTCGGACGGAACCCGACTGGGTTCAGTC
-578  TGATCGGTTCGGCGTTCGGCCGAACTGATTCGATGCCCTTAAGTAATAACGGG
-445  GGTGTCGGTCTCACTCTGAGGTGATTACGTGATCCCTTAAGTAATAACGGG
-312  GGCAAACGTCACGCTCGATTCGAGCGT
-179  TCCGAAGGAATGAGGATTCCACACCTGCGGTCCGCCGTAAAGATGGAATCTGATGTTAGCCTTGATGGTTTGGTG -105
in regulating transcription of the rRNA gene cluster.

In *E. coli* and in eucaryotic cell culture, strong selection for elevated expression of a gene results in tandem gene duplication (138). The multiple copy direct repeats preceding the rRNA gene cluster might also be the result of selection for elevated expression of the genes encoding rRNAs. In eubacteria, eucaryotes and some archaebacteria the selection has resulted in duplication of the entire gene cluster, whereas in *H. cutirubrum* no gene duplication occurred. Instead, the promoter region has been triplicated and presumably with this triplication are elevated levels of gene expression. It is interesting to note that many genes including the rRNA genes in *E. coli* have tandem multiple promoters. It is unclear how this triplication has taken place; it may be related to the transposable sequences and the high frequency genomic rearrangements known to take place in *Halobacterium* (112).

**Ribosomal RNA genes**

The 16S rRNA sequence as shown in Figure 20 is compatible with the universal secondary structure for small ribosome subunit RNA (94, 154). The nucleotides which are different from that of a rather distantly related *Halobacterium* strain, *H. volcanii*, are indicated (42). About 75% of the base substitution differences occur in regions of secondary duplex structure and are compensatory, maintaining base pairing within the RNA. This suggests that although the nucleotide sequence has evolved considerably, the structural features of the molecule are strongly preserved. Two rapidly evolving segments between nucleotides 115-210 and between 525-590 are clearly apparent. These two regions, representing about 10% of 16S rRNA sequence, contain 35% of the overall base substitutions. The counterpart of
Figure 20: Secondary structure map of 16S rRNA of *H. cutirubrum*.

The secondary structure map of the 16S rRNA is based with minor modifications upon the proposed universal secondary structure map of Woese *et al.*, 1983 (154). The sites of mutational differences between the *H. cutirubrum* and *H. volcanii* are indicated (>). An alternate structure for the 5' end of the 16S rRNA is illustrated. The large number of base substitutions in the region between nucleotides 110–200 allows for additional base interactions in this region which are not found in *H. volcanii* structure. A substitution mutation in the loop at nucleotides 652–671 allows for a potential four base pair stem within this large loop.
the sequences between nucleotides 525–590 in *E. coli* is believed to represent the binding site for the ribosomal protein S8 (91, 136). Surprisingly, *E. coli* S8 protein will bind to archaebacterial 16S rRNA; this suggests that structural features of the binding site are probably more important than its nucleotide sequence (137).

The 16S rRNA in *H. cutirubrum* like that of eubacteria has a pyrimidine rich sequence CCUCCU at its 3' end. In eubacteria this sequence is used in initiation of protein synthesis to identify a complementary purine rich Shine and Dalgarno sequence, AGGAGG, located about 6–9 nucleotides upstream from AUG translation codons on mRNA (121). However, at present there is no evidence supporting the idea that this pyrimidine sequence functions in an analogous way in the selection of translation initiation sites on archaebacterial mRNA. A report on the sequence of the bacteriorhodopsin gene by Dunn *et al.* (28) showed that the bacteriorhodopsin mRNA is different from most eubacterial mRNAs. It contains only two nucleotides at the 5' terminus preceding the AUG initiation codon. Therefore, during bacteriorhodopsin synthesis, interactions between leader sequences of its mRNA and the 16S rRNA cannot occur. However, within the bacteriorhodopsin mRNA beyond the AUG codon there are purine rich sequences, GGAG at nucleotides 6–9 and GGAGG at nucleotides 27–31, which are complementary to the 16S CCUCC sequence. Within the long open reading frames of the *H. halobium* transposable elements, ISH1 and ISH50, there are also purine rich sequences, GGA and GGAG at nucleotides 26–28 and 27–30 respectively downstream from the AUG start codon. These internal purine rich sequences may interact with 16S rRNA during translation.
In Figure 21, the 5' and 3' ends of the *H. cutirubrum* 23S rRNA sequence are compared with that of *E. coli* 23S rRNA sequence (16, 94), and a sequence believed to represent the proximal 100 nucleotides of 23S rRNA from another archaeabacterial species, *Methanococcus vannielii* (49). Several well defined regions of sequence homology were observed. With the sequence available, the *H. cutirubrum* 23S rRNA shows 50% homology with *E. coli* and 68% homology with *M. vannielii*.

The respective 16S-23S and 23S-5S intergenic spaces are about 509 and 112 nucleotides in length. These lengths are much shorter than the estimates of 900 bp and 400 bp for the very closely related species *H. halobium* (45). Fox *et al.* (34) showed that *H. cutirubrum* and *H. halobium* have the same 16S T1 oligonucleotide catalog suggesting that these two organisms are either very closely related or are of the same species.

**Inverted repeat sequences**

Flanking the 16S and 23S rRNA genes are sequences of long nearly perfect inverted repeats located between 23 and 141 nucleotides from their mature 5' and 3' ends. These duplex structures can be extended after brief interruptions to include additional residues from the flanking regions (Figure 22). In *E. coli*, similar long nearly perfect inverted repeat sequences surrounding the 16S and 23S genes form duplex structure in the primary ribosomal RNA transcript with the 16S and 23S mature sequences protruding as large loops (1, 14, 39, 94). These duplex structures are substrates for RNase III, a double-strand specific ribonuclease which liberates the precursor 16S and 23S rRNAs from the primary transcript. By analogy with *E. coli*, the nearly perfect inverted repeats in *H. cutirubrum* probably also form duplex structures with the 16S and 23S rRNA sequences protruding as
Figure 21: Nucleotide sequence homology between *E. coli*, *H. cutirubrum* and *M. vannielii* 23S rRNA genes.

The 5' and 3' sequences of *E. coli* (top), *H. cutirubrum* (middle), and *M. vannielii* (bottom) 23S rRNA genes are aligned for maximum homology (*, conserved bases). The gap in the *E. coli* sequence is 2642 nucleotides and in the *H. cutirubrum* sequence is about 2600 nucleotides. Only the proximal 100 nucleotides of the *M. vannielii* sequence are available. The 5' and 3' ends of the *H. cutirubrum* 23S rRNA gene were estimated by S1 nuclease mapping.
A comparison of 5' and 3' end of 23S rRNA gene among *E. coli* (top), *H. cutirubrum* (middle) and *M. vanneili* (bottom)

<table>
<thead>
<tr>
<th></th>
<th>10</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>GGTAAAGCGGACTACAGTGAGTGGCTGGCAGCTGGAAGGACGTGCTAAATCTGCCGAAGGGCTAGGTATATGACCGTACGGTGTTAGGGTGGGCTGCAAGGGAGGCTGACGGAGGCTACAAAGCCTT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. cutirubrum</td>
<td>CAACGCTACGGCTACCTGGATAAGCGTACACGGTTGGTGACCTGGTCAGTCAGAGGCGATGAAGGACGTGCTAATCTGCGATAAGCGTCGGTAAGGTGATATGAACCGTTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. vanneili</td>
<td>AATTATCTATTACCTACCTGGGGAATGGCTTGCTGAAAACGCCTCGATGAGGACGTATAGGCGAGGGAGCCGCAGGAGGCTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>110</th>
<th>120</th>
<th>130</th>
<th>140</th>
<th>150</th>
<th>160</th>
<th>170</th>
<th>180</th>
<th>190</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>ATAACCGGCGATTTCCGATGAGGAAACCCAGTGGTTTCGACACATCATATTAACTGAAATCCATAGGTTATAGGAACCCGG-GGGGAACCGAATCCATGAAACAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H. cutirubrum</td>
<td>GAACTCAGGATCTCTATGGGAAATCCCTATAACATGCTTTGGAATAGGGAAACCTGCCGAAGGGAGGCTGACGGAGGCTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. vanneili</td>
<td>TAAACCGGAGAGCAGATGAGGAAACCCAGTGGTTTCGACACATCATATTAACTGAAATCCATAGGTTATAGGAACCCGG-GGGGAACCGAATCCATGAAACAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2850</th>
<th>2860</th>
<th>2870</th>
<th>2880</th>
<th>2890</th>
<th>2900</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>C...........TAAGCGGAGAGCAGATGAGGAAACCCAGTGGTTTCGACACATCATATTAACTGAAATCCATAGGTTATAGGAACCCGG-GGGGAACCGAATCCATGAAACAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. vanneili</td>
<td>A ...........TCGAGGCGGAGAGCAGATGAGGAAACCCAGTGGTTTCGACACATCATATTAACTGAAATCCATAGGTTATAGGAACCCGG-GGGGAACCGAATCCATGAAACAT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 22: Sequences flanking the 16S rRNA gene and the 23S rRNA gene.

The sequences surrounding the 16S rRNA gene (left) and the 23S rRNA gene (right) are potentially rich in secondary structure. The extended helical duplex region at the base of each of these structures results from the long nearly perfect inverted repeat sequences which flank both the 16S rRNA gene and 23S rRNA gene. The position of the 5' and 3' ends of mature 16S or 23S rRNA are indicated ( ); the mature 16S and 23S sequences are within an extended loop protruding from the top of these structures. The mature 23S rRNA sequence may be up to four nucleotides longer than indicated on either or both of the 5' and 3' ends. The positions corresponding to the 5' \textsuperscript{32}P-labeled at the TaqI site (nucleotide 1941) and the SalI site (nucleotide 1973), which were not protected in S1 experiments and therefore not part of mature rRNA, are indicated by •.
large loops; the initial processing event is presumably carried out by an RNase III like enzyme which liberates the precursor 16S and 23S rRNAs.

In *E. coli* the mature ends of the 23S gene are located within the inverted repeat duplex structure (and are therefore complementary) and RNase III processing occurs very close to these mature ends. In contrast, the corresponding mature ends of *H. cutirubrum* rRNAs are well removed from the inverted repeat duplex structure. The sequence between the inverted repeat duplex and the mature ends of the 23S rRNA contains some secondary structures reminiscent of eucaryotic 5.8S rRNA (147). In eucaryotes, this RNA is derived by an endonuclease cut from the 5' end of the large ribosomal subunit RNA. In eubacteria, there is no endonuclease cut so that the 5' end of the 23S rRNA is analogous in sequence and structure to eucaryotic 5.8S rRNA. Nuclease S1 protection experiments utilizing end-labeled restriction fragments and total RNA from *H. cutirubrum* were carried out to determine if the 5' sequence flanking the mature end of the 23S rRNA is degraded during processing or if it is preserved as a 5.8S like component in the ribosome. The result indicated that neither 5' nor 3' labeling at the *TaqI* site (nucleotide 1939) or the *SalI* site (nucleotide 1969) was protected by total RNAs, and therefore it is concluded that:

1. The mature 5' end of 23S rRNA was located downstream from the *SalI* site, and
2. The 5' flanking sequence is not preserved as a 5.8S like component of the large ribosomal subunit.
Transfer RNA genes

Within the ribosomal RNA gene cluster there are two sequences which, when transcribed, are capable of forming a universal clover leaf structure containing the highly conserved features of tRNAs (Figure 23). A tRNA sequence, located within the 509 nucleotide long 16S-23S spacer (nucleotide 1575-1651), contains a UGC alanine anticodon. This *H. cutirubrum* alanine tRNA shares with all other alanine tRNAs the feature of contiguous G-C, G-C and G-U base pairs in the terminal positions of the acceptor stem; no tRNA of any other amino acid family contains this feature (125). Unusual, however, is that the tRNA gene is not associated with an isoleucine tRNA as in all other alanine tRNA containing spacers in *E. coli* and *Bacillus subtilis* (64, 76, 152). An alanine tRNA gene with the same anticodon has been identified in one of the 16S-23S intergenic spacers in a related archaeabacterial species, *Methanococcus vanniellii* (49). A different alanine tRNA from *H. cutirubrum* was sequenced by Gu *et al.*, 1983 (41) and was shown to contain a CGC in the anticodon. That tRNA shows 84% homology with the DNA sequenced in this thesis. Gu *et al.* also suggested that archaeabacteria may have unusual decoding properties due to a paucity of U in the first anticodon position of all the archaeabacterial tRNA sequenced. However, the alanine tRNA sequence of *M. vanniellii* and the alanine DNA sequence of *H. cutirubrum* reported in this thesis make this hypothesis uncertain.

A putative tRNA sequence was found located distal to the 5S gene (nucleotides D395-D468) and contained a GCA cysteine anticodon. This tRNA is unusual in that it contains only two base pairs in the dihydrouracil stem with 16 bases in the dihydrouracil loop. A cysteine tRNA with a similar ribonucleotide sequence has recently been characterized from *H. volcanii* but
Figure 23: Cloverleaf structure of putative alanine and cysteine tRNAs.

The nucleotide sequence and universal cloverleaf structure of two putative tRNA sequence are illustrated. The alanine sequence is located in the 16S-23S intergenic space (nucleotides 2152-2229) and the cysteine sequence is located distal to the 5S rRNA (nucleotides D395-D496). Both DNA sequences lack the conserved CCA sequence at the 3' acceptor end of the mature tRNA.
its coding sequence has not been identified (Woese and Gupta, personal communication).

Neither tRNA gene contains the universal CCA trinucleotide found at the 3' end of mature tRNAs above the 7 base pair acceptor stem; this implies that these nucleotides are added by a nucleotidyl transferase maturation system.

**Terminator sequences**

There are two potential sites in the 3' flanking sequence that may function in transcription termination. The first is a short inverted repeat followed by T₁ at position D302-D337, analogous to the *E. coli* rho-independent termination system (47, 103, 110; Figure 24). Although there are 13 base pairs in the stem, the stem is not very stable because only 7 of them are G-C pairings (in contrast to *E. coli* in which almost all are G-C pairings). Partial or complete termination at this site would reduce the level of expression of the cysteine tRNA gene relative to the rRNAs. The second terminator-like signal is a G/C rich region followed by an A/T rich region (D544-D604) located distal to the cysteine tRNA gene. The terminator of the bacteriorhodopsin mRNA consists of a limited region of dyad symmetry overlapping 10 consecutive G/C residues and followed by an A/T rich region; termination occurs in the A/T rich region (21). A strong transcription termination site downstream from the HL20 ribosomal protein gene of *H. cutirubrum* (equivalent to the *E. coli* L12 gene) lacks dyad symmetry but contains a very G/C rich sequence, including 9 consecutive G/C base pairs, followed by an A/T rich region including a T₆ stretch; termination occurs near the beginning of the T₆ sequence (Dennis and Shimmin, unpublished results).
Figure 24: An inverted repeat termination-like signal.

An inverted repeat found at 18 nucleotides beyond the 5S gene is illustrated. This inverted repeat (nucleotides D303-D332) contains a loop of 4 nucleotides and a 13 base pair stem followed by Us. This structure resembles a typical eubacterial rho-independent termination signal (47, 103, 110).
Unique copy of ribosomal RNA genes

Genomic Southern hybridizations have indicated there is only one copy of each of the rRNA gene in *H. cutirubrum*. Even the presumptive promoter sequence is unique to this cluster. In rich medium at 37°C, *H. cutirubrum* grows very slowly with a doubling time of about six hours; therefore, a single copy of the rRNA genes with three tandem repeat promoters is probably more than adequate to produce a sufficient number of ribosomes to satisfy the demand for protein synthesis. Indeed these bacterial cells appear to be rich in ribosomes and the ribosome synthesizes proteins at 10–20% of the rate for the *E. coli* ribosome (Dennis and Chant, unpublished results).

In this investigation, the molecular characterization of the unique ribosomal RNA gene cluster in *H. cutirubrum* was achieved. The complete DNA sequence of the 16S and the 5S rRNAs, partial sequence of the 23S rRNA, and 5', 3' flanking region and the intergenic region of this gene cluster were determined. The arrangement of the rRNA genes are: 5' 16S–23S–5S 3'. From the sequence data, two transfer RNA–like sequences were found—an alanine tRNA (anticodon UGC) located in the intergenic region between 16S and 23S rRNA and a cysteine tRNA (anticodon GCA) located distal to the 5S rRNA gene. Putative sequences which resemble transcript processing, initiation, and termination signals were detected. The 16S and 23S rRNA genes are surrounded by extensive inverted repeat sequences which are presumably used for the processing and maturation of these rRNAs. The 5' flanking sequence of the gene cluster contains three perfect copies of a bipartite direct repeat unit which contains sequences believed to be important components of the *Halobacterium* promoter. The 3' sequence flanking the 5S rRNA gene contains an inverted repeat followed
by T, which resembles the rho-independent termination signal in *E. coli*.
REFERENCES


