Chapter 1

Review of Literature
AGAR, AGAR HYDROLYSIS AND AGARASES

Agar as a complex polysaccharide

Formerly called Agar-Agar, agar is a gel forming extract of agarophytes. The name agar originated in Indonesia and referred to red seaweed of the genus Eucheuma (Chapman & Chapman, 1980) though it actually refers to the phycocolloid of these seaweeds. Indonesians still refer to some seaweeds and to jellies from seaweeds as "agar-agar". Agar is a phycocolloid belonging to a family of polysaccharides obtained from the agarophytes of Rhodophyceae. The term agarophyte refers to the seaweeds used for the manufacture of agar. The most important agarophyte genera are Gelidium and Gracilaria. Agar has also been found in species of Ceramium, Phyllophora, Pterocladia, Ahnfeltia, Campylaephora, Acanthopeltis (Lewis, Stanley & Guist, 1990), Gelidiella and Gracilariopsis (Armisen, 1995).

Properties and structure of agar

Agar is a cell wall polysaccharide. It is soluble in hot water. 1.5% hot agar gives a clear solution after cooling to 30°C – 39°C, it becomes solid and elastic gel which redissolves only above 85°C.

Agar can be separated into two fractions. The first is a neutral polymer agarose composed of repeating units, referred to as agarobiose, of alternating 1,3-linked beta-D-galactopyranose and 1,4-linked 3,6-anhydro-alpha-L-galactopyranose (figure 1.1) (Lewis, Stanley & Guist, 1990).

The second is a variable group known to be highly (Chapman & Chapman, 1980). The fraction has been termed as agaropeptin but due to its variability the name does not adequately define the group (Lewis, Stanley & Guist, 1990). However, it is found that the repeated sequence of neoagarobiose is
Fig. 1: Structure of the galactose disaccharide.

\[ \text{D-galactopyranose} \quad \text{3,6-anhydro-L-galactopyranose} \]
masked in agaropeptin, the charged polysaccharide complex of agar. The complexity of the agar polymers originates from the varying substitution of hydroxyl groups with methoxyl and sulphate groups, mostly as 6-O-hemiesters and, less frequently pyruvic acid bound as 4,6-O-(1-carboxyethylidene)-D-galactose (S. Hirase, 1957; M. Duckworth et al., 1971). The chemical structure of agar in terms of sulphate, methoxyl and pyruvate substituents depends markedly on the algal source (I. Levy and M. Friedlander, 1990). However, the 4-O-sulphate and pyruvate content is normally quite low. Recently, agar, unusually rich in these two substituents, has been described from Gracilaria bursa-pastoris (E. Murano et al., 1995) and Gracilaria dura (E. Murano, 1992).

The agarose fraction is responsible for the gelling properties of agar while the second fraction provides the viscous component. The viscosity varies depending upon species of alga, method of production and sulphate content. If there is a higher proportion of agarose in agar, it forms a stronger gel. It is considered that the agarose molecules are present as double helices and that, in gelling, these aggregate (figure 1.2) to form a matrix of double helical polymer held together by hydrogen bonds (Ying, N. and Yaphe, 1972; Chapman & Chapman, 1980). This double helix structure when disturbed or broken fails to form a brown color with Lugol's iodine (Gram, 1902).

**Agar as a useful bio-product**

There are basically three grades of agar: bacteriological (microbiological), sugar-reactive and food-grade (industrial-grade). Bacteriological grade agar is the most expensive and microbiologically pure (contains no bacteria), representing 4-5% of total agar sales.

Sugar-reactive agar is the form that is used for cake glazes and confectionery for its clarity and microbiological purity. The strength of the gel is weak.
Food-grade is the lowest priced grade of agar being the least refined. It is used in food as a standard thickener.

Natural-grade agar refers to the first extract from seaweed. This extract is usually manufactured in the country (usually a developing country) where the seaweeds, from which it is obtained, are grown. The reason for this is to reduce the bulk that needs to be transported, and thus the transportation costs, to a developed country where it is processed further to obtain a high value gel, such as agarose. *Gelidium* produces the highest grade agar and agarose.

Agarose is a highly refined and specialized product that has played a pivotal role in the biotechnological revolution. This product is manufactured by isolating the less ionic fraction of agar under highly controlled conditions designed to minimize lot to lot variations (Radmer, 1996). With the increase of biotechnological applications, agarose has become the highest-priced seaweed product taking 0.2% of the phycocolloid market (Critchley, 1997).

Agar is primarily used in food, medical/pharmaceutical and cosmetic industries. It is also used in other industries such as its recent application in the packaging industry for the production of biodegradable packaging foam.

**Hydrolysis of agar**

The double helical structure of agar can be disrupted without cleavage of any glycosidic bond; however, it can be disrupted by the destruction of chemical linkages with acids or hydrolyzing enzymes. The chemical cleavage by mild hydrolysis of agar and oligosaccharides with acid have been well characterized (Araki, 1957 and Kwan S. Young *et. al.*, 1978). Mild acid is used for agar hydrolysis, which is then neutralized with alkali.
The biological hydrolysis of agar is done by a class of bacteria called agarolytic bacteria. The biological changes caused by cleavage of agar molecules by these agarolytic bacteria range from the softening of the surface of the gel to the formation of craters and eventual complete liquefaction.

Agarolytic bacteria use agar as their source of carbon and energy. This hydrolysis of agar by agarolytic bacteria may be termed as agarolysis.

**Types of agarolysis**

Agar is a polysaccharide, and hence it takes intense brown color when stained with Lugol's iodine (solution of KI + I₂). But agarolytic products fail to stain brown, and give yellow color which was applied by Gran (1902) to show degradation of agar on an agar plate. Agarolytic cleavage can be broadly classified into three main types.

i) Destruction of double helical structure without any glycosidic cleavage

ii) Cleavage of α-linkages of the agar molecules, giving rise to oligosaccharides of agarobiose series with 3, 6-anhydro-L-galactose at the reducing end.

iii) Cleavage of β-linkages giving rise to neoagarobiose series of oligosaccharides with D-galactose at the reducing end.

The hydrolyzing enzymes, which cleave α and β linkages, are called α-agarase and β-agarase, respectively. Such α and β cleavages were first reported by Duckworth and Turvey in 1969.

**Agarolytic bacteria**

Depending upon the extent of degradation of agar, agarolytic bacteria have been divided into two groups.
1) group causing softening of the agar.
2) group causing extensive liquefaction of the agar (Agbo and Moss, 1979).

Agar degraders which only soften the agar have been reported from the genus *Cytophaga* (Agbo and Moss, 1979). β-agarase I of *Pseudomonas atlantica* has also been found to degrade agar mildly (Morrice *et al.*, 1983), whereas β-agarase II of *Pseudomonas atlantica* (Morrice *et al.*, 1983), agarase from *Alteromonas* (Agbo and Moss, 1979), several α-agarases from *Streptomyces coelicolor* (Butner *et al.*, 1987), *Vibrio* strains (Aoki *et al.*, 1990) etc., were found to liquefy the agar extensively.

**Occurrence and distribution of agarolytic bacteria**

Most agarolytic bacteria have been isolated from the marine environment, which require a relatively high concentration of salts for growth. But several non-marine species have also been reported (Araki and Arai, 1954). Hofstein and Malmqvist (1975) isolated agarolytic bacteria from sewage water, which contain large amount of agar discharged from hospitals, research laboratories and pharmaceutical industries. Several strains of agarolytic bacteria belonging to *Cytophaga* and *Alteromonas* have also been reported from low-land River Wey (Agbo and Moss, 1979). These carry organic materials, of which large parts were polymers of plant origin including algae. Hence agarolytic bacteria are found more in association with agarophytes or organic material containing polysaccharides.

**Agarases: A hydrolases of glycosylases family**

Nomenclature committee of International Union of biochemistry and Molecular Biology (NC-IUBMB) has recommended the name 'agarase' for the enzyme which hydrolyzes 1,3-β-D-galactosidic linkages in agarose. The IUBMB code for this enzyme is EC 3.2.1.81. The other name for the same enzyme is
β-agarase and the systematic name given is ‘agarose 3-glycanohydrolase’. The enzyme agarase has been placed in ‘hydrolases’ (EC 3), ‘Glycolsylases’ (EC3.2) and the group ‘Glycosidases’, (EC3.2.1) i.e., enzyme hydrolyzing O- and S- glycosyl compounds. (http://www.chem.qmw.ac.uk/iubmb/enzyme/EC3/2/1/81.html).

Till the preparation of this manuscript, the IUBMB had not given any separate nomenclature for α-agarase.

Types of agarases and their mode of action

Depending on the mode of the action of the enzyme, agarase is precisely classified into α-agarases, which cleave the α-1, 3 linkage; and β-agarases which act on β-1, 4 linkages. (Young, et. al., 1978; Duckworth and Turvey 1969). Many agarolytic bacteria produce α and β-types of agarases. The two types of enzymatic actions can easily be recognized by their formed characteristic products. This in turn can be separated by thin layer chromatography and identified (Duckworth and Yaphe, 1970). As mentioned above, action of α-agarases results in the formation of characteristic oligosaccharides belonging to the agarobiose series, while β-agarase results in the formation of neoagarobiose series of oligosaccharides which have specific ‘R Gal’ value and hence are differentiated easily.

The breakdown of the complex agar polysaccharide to simple sugars is exemplified by a Pseudomonas atlantica agarase system at paradigm, which is 1, 3 enzyme agarase system proposed to operate in the following manner (fig. 1.3). The first enzyme is a β-agarase, designated β-agarase I. β-agarase I is an endoenzyme which acts on the 1, 4 linkages between D-galactose and 3,6-anhydro-L-galactose, resulting in a neoagarotetraose tetramer end product. β-agarase I can also act on neoagarooctaose and neoagarohexaose. The agarase I enzyme is extracellular.
The second enzyme, referred to as either β-agarase II or neoagarotetraose hydrolase, is also an agarase. β-agarase II cleaves the central 1, 4 linkage of the tetramer produced by the β-agarase I to yield a neoagarobiose dimer. Mild acid hydrolysis of agar can result in an agarobiose dimer rather than neoagarobiose. The second agarase enzyme also has residual exoenzymatic activity on the large agar polymer. The β-agarase II enzyme is cell wall associated.

The final enzyme in the agarase system is an α-agarase. The α-agarase enzyme cleaves the α linkage of neoagarobiose to yield D-galactose and 3, 6-anhydro-L-galactose. The cellular location of this neoagarobiose hydrolase is not clear.

**Substrate specificity and chemical analysis of agarase activity**

Duckworth and Tuvey (1969) carried out substrate specificity studies and found that the agarase enzyme is specific for its substrate agarose, although it also uses related substrates like neoagarooligosaccharides, porphyran, carragenan and galactan.

Thus apart from digesting agar, agarose and neoagarooligosaccharides, agarase also acts on other substrates such as galactomannan, starch, pectin, carboxymethylcellulose, etc. Sepharose 2B, 4B and 6B, washed free of the bacteriostatic agents were found to be completely degraded by agarolytic bacteria but no degradation occurs when sepharose is cross-linked (Hofsten and Malmqvist, 1975). The highly substituted agaroid, prophyran was also degraded by β-agarase I and β-agarase II of Pseudomonas atlantica and β-agarase of Cytophage species in endo-fashion. However, in case of alkali treated porphyran the activity was found to be half of that found in agarose because of the presence of β-O-methyl and sulphated groups on some of the
Fig. 1.3 Sites of action of β-agarase I and β-agarase II.
D-galactose units. These units resulted in hindrance in hydrolysis of $\beta$ 1, 4-linkages.

Similar studies by substituting sulphated esterified groups in agarose have reported reduction in agarolytic activities (Duckworth et. al., 1969).

Because of degradation of varied polysaccharides, certain agarases have been successfully used for protoplast isolation from algae (Chen et. al., 1994) and oligosaccharides preparation from agar and related polysaccharides.

**Assay of agarase activity and definition of unit**

There are two commonly used procedures for assays of agarase activity:

i. **Viscometric method**

This method was first used by Turvey and Christison (1967) for measurement of agarase activity. It is based on the decrease in viscosity of a substrate solution by action of enzyme, which is in turn measured by Ostwald viscometer. In this case the unit of activity is defined as the amount of enzyme that would have decreased the specific viscosity of the digest in 100 mins, under the specified conditions.

ii. **Reducing power method**

Agarase enzyme acts on its substrate i.e. agarose or similar polysaccharides and results in formation of reducing sugar, most commonly galactose, which can be estimated by the method described by Somogyi (1952) and Nelson, (1944). Miller (1959) used dinitrosalicylic acid (DNSA) to determine reducing sugar spectrophotometrically.
In this case, one unit of enzyme activity is defined as that which liberates one μ-mole of D-galactose per mg protein per min at pH 7.0.

**Nature of biochemical reaction**

Most of the agarases require pH in neutral or slightly acidic range for optimum activity. Agarase from *Pseudomonas atlantica* requires neutral pH whereas *Vibrio* species show an optimum activity at pH 5.5. But agarase of *Pseudomonas* sp. PT - 5 has its maximum activity at pH 8.5. The buffer system found suitable includes mainly Tris-C1, 10-50 mM, pH 6.8 - 8.0 (Sugano et. al., 1993; Oscar Leon et. al., 1992). Optimum temperature for the agarase activity was found to range from 30°C to 45°C. Agarase from *Alteromonas* sp. strain C-1 requires 30°C for its optimum activity and β-agarase from *Vibrio* sp. AP-2 was fully active till 45°C (Aoki et. al., 1990). As gelling strongly retards attack on the agar by agarase, an elevated temperature is desirable for the initial stages of a digest. But at higher temperature, enzymes lose their activity due to thermal denaturation (Morrice et. al., 1983).

**Genetic composition and molecular properties**

Agarase has been found to be chromosomally encoded. *Aga A* gene from *Pseudomonas atlantica* coding for 57.48 kD β-Agarase (Belas, 1989), *dag A* gene from *Streptomyces coelicolor* for 32 kDa agarase (Buttner et. al., 1987) and *aga A* gene of *Vibrio* sp. Strain JT0107 coding agarase of 107 kD (Sugano et. al., 1993) have been studied extensively to detect the homology using the perfect match analysis and the functional level analysis. A common characteristic of these three proteins was that they had a highly hydrophobic N-terminal region that seemed to function as a signal peptide for secretion of the protein. The first 20 residues in *Vibrio* agarase were found to be the signal peptide. Although these agarases have different molecular weights yet perfect
match analysis and functional level match analysis indicated the presence of the two homologous regions which might be related to agarase activity (Sugano et. al., 1993). Molecular weight of agarase from Vibrio sp. JT0107 was 107kD (Sugano et. al., 1993). The molecular weights of other agarases were 32 kD from Streptomyces coelicolor (Buttner et. al., 1987), 32 kD β-agarase I from Pseudomonas atlantica (Morrice et. al., 1983), 20 kD from Pseudomonas sp. PT-5 (Yamaura et. al., 1991) etc.

Applications of agarase

In the modern industrialized world, agarase enzyme seems to have a very promising potential and a wide range of applications. Agarase is generally used as an anti-convectional agent in electrophoresis of DNA. As agarose is a substrate for agarase enzyme, the enzyme is used for elution of DNA from agarose gel after electrophoresis. This technique has been utilized very successfully by many scientists to recover high molecular weight DNA fragments from agarose gels after electrophoresis. For example, 1.6 megabase pairs Yeast artificial chromosome had been purified after pulsed-field gel electrophoresis by Maurice et. al., (1994). This method has also been utilized by Kai et. al., (1990) for human genomic library preparation. Many other scientists who are working on human genome project are utilizing agarase for recovery of large DNA fragment from agarose gels.

It is well known that seaweeds are very valuable marine resources for various important substances like unsaturated fatty acids, vitamins, caratenoids, betaine, etc. One of the main obstacles in making use of these valuable resources is the degradation of the polysaccharide in the cell wall, since these agarophytes are known to contain large quantity of agar and related oligosaccharides like porphyran and carrageenan. Hence agarases with broad substrate specificity can be used for enzymatic destruction of cell wall and for obtaining the valuable substances from it. Other chemical degradation
methods are harsh and thus cannot be utilized for complete recovery of these resources.

Agarophytes contain large amounts of agar as one of the main components in their cell wall, and hence agarase enzyme can be utilized for degradation of cell wall and obtaining protoplast from agarophytes which has potential in molecular cloning. Japanese scientists have already obtained a patent for using agarase from *Vibrio* sp. to obtain protoplast from algae (Japanese Patent et. al., 1993). Agarases were used by Magnama et. al., (1993) to release protoplast from the red algae *Gelidium robstrum*. However, the most impressive example of this series is the utilization of protoplast obtained from *Gelidium* in aquaculture as conchospore.

In addition to the above, agarase enzyme has other diverse applications. For example, the Japanese have patented a method of softening agar culture medium for plant propagation using agarase enzyme. Further, in 1990 Japanese patented a method for preparation of agar oligosaccharides using agarase enzyme, which has got commercial importance. Agarase enzyme has also been utilized for preparation of low calorie food and beverages. A strong promotor of agarase gene is also utilized in creating new vectors for the production of extracellular proteins (Malek et. al., 1990; Buttner et. al., 1986).

**IDENTIFICATION OF UNKNOWN BACTERIA**

Biodiversity, a concept used to describe the range of living organisms in a given area, takes into consideration the variety of life forms, the genes they contain, and the ecosystems they form. Life forms within an ecosystem vary in their size and shape from the simplest unicellular prokaryote, the bacteria, to the more complex multicellular eukaryotic organisms, such as elephants or bottlenose dolphins. Each organism plays an important role and contributes to
ecosystem stability. Each type of environment, aquatic or terrestrial will have its own characteristic type of microorganisms or microbes.

During the last decades bacterial taxonomy has undergone remarkable changes. New categories of information of potential taxonomic value have become available (e.g., chemotaxonomy, DNA base composition, DNA-DNA hybridization, etc) which distinguish between organisms and reveal similarities not detected before. In addition, phylogeny has attracted increasing attention and phylogenetic relationships are now an important basis in the classification of bacteria. The integrated use of phenotypic and genotypic characteristics will have great influence not only on classification but also on bacterial nomenclature, and will continue to do so.

To identify bacteria, biochemical testing gives important clue for the taxonomical position of the organism. The types of biochemical reactions each organism undergoes act as a "fingerprint" for its identification. This is based on the following:

- Each different species of bacterium has DNA with a unique series of nucleotide bases.
- Since DNA codes for protein synthesis, then different species of bacteria synthesize different protein enzymes.
- Enzymes catalyze all the various chemical reactions of which the organism is capable. This in turn means that different species of bacteria must carry out different and unique sets of biochemical reactions.

**Genetic analysis**

The ultimate indicator of taxonomic classification is at the level of DNA, the genetic material. Thus if the DNA from an unknown organism is compared to the DNA of a known organism, the degree of relatedness can be determined. Some genetic identification is accomplished through the use of DNA probes.
A labelled piece of DNA representing a unique genetic sequence is incubated with DNA from the unknown organism. Detection of the labelled probe indicates the presence of the DNA sequence in the unknown. Phylogenetic analysis of bacteria is typically performed by comparing the sequence of a portion of the 16S ribosomal RNA gene. Polymerase chain reaction is used to amplify this gene and the amplified region is then isolated and sequenced. Comparisons can be made using databases that are available on-line. Recently gyrB sequences have also provided a useful tool to discriminate bacterial isolates at strains level.

**Phylogenetic analysis using 16S rDNA sequencing**

The development of molecular methods during the 1980’s and 1990’s led to the adoption of ribosomal RNA molecules as good indicators of phylogenetic relationships (Woese, 1987). The 5S, 23S and the more widely used 16S rDNA gene sequences were proposed to reflect best phylogenetic relationships among bacteria because they are stable, highly conserved and present in all bacteria (Woese, 1987). The 16S rRNA gene is approximately 1500 bp long and it is usually present as several copies in a bacterial genome (Woese, 1987). Genus or species specific primers or probes can be designed to amplify the 16S rDNA.

In recent years, the extensive use of sequence-based analyses of ribosomal genes has led to the establishment and rapid expansion of large international sequence databases such as EMBL (European Molecular Biological Laboratory) and GenBank. It should be noted that this rapid development has led to a situation where 16S rDNA sequencing presently dominates, or has even replaced the use of many classical taxonomic methods. However, restrictions apply to this method (Vandamme et. al., 1996).
It has become clear that phylogenetic analyses cannot replace polyphasic taxonomic studies. It was noted in 1992 that 16S rDNA sequence identity is not sufficient to guarantee species identity (Fox et al., 1992). Two clearly separate *Bacillus* spp. were shown to be distinct species by means of DNA-DNA hybridization experiments and phenotypic properties, yet they shared 99.5% 16S rDNA sequence similarity (Fox et al., 1992). A literature survey was performed to compare the results from a number of 16S rDNA sequence analyses and DNA-DNA hybridization experiments (Stackebrandt and Goebel, 1994). The authors considered that, if the 16S rDNA sequence homology of a given group of bacteria was less than 97% to existing species, they were not likely to demonstrate more than 60% relatedness in DNA-DNA hybridization experiments. However, if the sequence homology exceeded 97%, DNA-DNA hybridization experiments were necessary to define species boundaries as it was noted that above this value of sequence similarity, the DNA-DNA hybridization value may be either low or as high as 100% (Stackebrandt and Goebel, 1994). It was also concluded that 16S rDNA analysis is a valuable addition to polyphasic taxonomy and could be used as a screening method to determine if more time-consuming DNA-DNA hybridization experiments should be performed (Stackebrandt and Goebel, 1994, Vandamme et al., 1996).

A more complex situation is evident when molecular sequencing data is used as a sole criterion to define species boundaries. This occurs when DNA from unculturable bacteria is subjected to direct PCR amplification with primers designed to amplify the 16S rDNA region and the product is cloned and sequenced. As a consequence, descriptions of several new, less than adequately described, bacterial species and genera have been published (Murray and Schleifer, 1994). A category "*Candidatus*" has been proposed for the descriptions based on sequences of unculturable organisms to solve this problem (Murray and Schleifer, 1994).
Polyphasic taxonomy

In simple terms, taxonomy is an attempt to organize the constantly expanding information base concerning natural diversity into a meaningful heirarchical system. Classically, it has been divided into three parts; 1) classification of organisms into groups on the basis of similarity, 2) nomenclature of these groups according to international rules and 3) identification of unknown organisms (Krieg and Holt, 1984, Vandamme et. al., 1996). However, studies using just one or few methods to classify bacteria often give confusing results. An approach which utilizes a large amount of data from phenotypic and genotypic analyses is called polyphasic taxonomy. In this regard, a taxon (whether it be a species, genus, family, superfamily, subdivision, division or domain of any group of living organisms) is defined with all necessary information from classical phenotypic analysis to highly sophisticated nucleic acid sequence data (Vandamme et. al., 1996). Colwell first used the term "polyphasic taxonomy" in 1970 (Colwell, 1970). It should be remembered that this approach contains no strict or defined guidelines describing its application, but involves a proper definition of the problem to be resolved and the subsequent application of the most suitable methods for this purpose (Vandamme et. al., 1996). As a consequence, the end result should serve all practical users of taxonomy; clinicians, veterinarians, epidemiologists, and other scientists to the best possible extent (Vandamme et. al., 1996).

The basic unit of taxonomy is the species, since it is the lowest category that cannot be further subdivided. It must be noted that bacterial species are not like animal species, which may be defined by specific criteria, namely the possibility of two individuals of opposite gender to produce offspring with the capacity to reproduce. The definition of bacterial species requires considerable flexibility. At present, consensus guidelines define a bacterial species as a group of strains (including a type strain) that is characterized by a certain degree of phenotypic consistency, a significant degree of DNA-DNA
homology (preferably over 70%) and over 97% of 16S rDNA homology (Wayne et al., 1987; Stackebrandt and Goebel, 1994; Vandamme et al., 1996). In addition to these requirements, it is important to note that at least two phenotypical differences should exist to other defined species (Vandamme et al., 1996) Thus, these guidelines give precedence of the phenotype over the genotype (Wayne et al., 1987, Vandamme et al., 1996).

Much confusion has been caused by introducing more taxonomic groups. A category of subspecies or biotype has been referred to organisms that differ phenotypically, but exhibit hybridization value of approximately 40-60%. The opposite also occurs. If two genetically distinct groups can be identified, yet not differentiated from each other by clear phenotypic means, they should not be considered as separate species, but be designated as genospecies, genomic species, genomovars or genomic groups (Wayne et al., 1987, Ursing et al., 1995). Many of these other categories have caused more problems than they have solved (Vandamme et al., 1996).

Two groups of strains can be considered as two separate species if they exhibit clear phenotypic differences and share insignificant degree of binding in DNA-DNA hybridisation experiments (less than 30%) or show less than 97% similarity in their 16S rDNA sequences (Stackebrandt and Goebel, 1994; Vandamme et al., 1996). If two such groups of strains can be differentiated phenotypically, but the 16S rDNA sequence similarity is higher than 97%, DNA-DNA hybridisation experiments should be performed to delineate species boundaries (Stackebrandt and Goebel, 1994). In many cases these rules cannot be applied, but then the polyphasic taxonomic approach is the method of choice (Vandamme et al., 1996).
**GyrB** : A novel phylogenetic marker

For the use of phylogenetic analysis, it is necessary to select target genes. Such genes should be universal among bacterial species, should not show any adaptive mutation to specific growth conditions, and should not be transferred horizontally. Many genes could be discarded from candidates according to these criteria: *fliC*, *ospC* may probably not be useful because as they are not universal among bacterial species and acquire adaptive mutations in some bacteria such as pathogens which should cope with host defense mechanisms including host immune systems. Catabolic genes may also not be appropriate because many of them are transferred from one host to the other horizontally. However, these genes are generally more specific for the identification of strains.

DNA topoisomerases are essential for DNA replication, transcription, recombination and repair and they control the level of supercoiling by cleaving and resealing the phosphodiester bond of DNA. They are classified into type I (EC 5.99.1.2) and type II (EC 5.99.1.3) according to their enzymatic properties. The bacterial DNA gyrase is a type II topoisomerase that is capable of introducing negative supercoiling into a relaxed closed circular DNA molecule.

The *gyrB* gene appears to meet all the requirements of a phylogenetically useful protein-coding gene. It does not appear to be frequently horizontally transmitted and can be found in most, if not all, bacterial species. Speeds of amino acid substitution are different in different proteins or in different sites in the same protein. In this regard, *gyrB* is also suitable. The rate of evolution is not only faster than ribosomal genes, but also appears fast relative to other protein coding genes (Yamamoto, S. and Harayama, S., 1996). This makes it especially useful for strain discrimination and identification.
Its main advantage is the availability of universal primers which have been shown to successfully amplify DNA across a wide range of bacterial classes (Yamamoto, S. and Harayama, S., 1995). Furthermore, the resulting amplified fragment, around 1.2 to 1.4 kb, is not only a useful length for phylogenetic studies but also contains both variable and conserved regions (Kasai, H. et. al., 2000).

A validity of gyrB sequences as taxonomic markers was evaluated mainly from two points of view: the rate of their base substitutions and the consistency between the results of gyrB-based analysis, and those of the DNA-DNA hybridization analysis. Protein-coding genes are thought to evolve faster than rRNA genes because synonymous substitutions mainly at the third positions of codons in the protein-coding genes are permitted without causing any changes in the amino acid sequences of their gene products. In fact, the average base-substitution rate of 16S rRNA genes was 1% per 50 million years, while that of gyrB at the synonymous sites was estimated to be 0.7 - 0.8% per one million years. Therefore, some species with completely identical 16S rDNA sequences can be differentiated by using their gyrB gene sequences. Four species belonging to *Mycobacterium tuberculosis* complex, *Mycobacterium kansasii* and *Mycobacterium gastri* (Kasai, H. et. al., 2000) are such examples.

DNA-DNA hybridization is the recommended standard for the delineation of a bacterial species (Wayne et. al., 1987), but it has a drawback that this method is not effective in the estimation of genetic distances between distantly related species. As a complement to DNA-DNA hybridization, sequence analysis of 16S rRNA or its gene (16S rDNA) is frequently used in taxonomy (Moore et. al., 1996). However, the degree of resolution obtained with 16S rRNA sequence analysis is not sufficiently discriminatory to permit resolution of intrageneric relationships, because of the extremely slow rate of evolution of 16S rRNA. Due to the gap between the valid genetic ranges of the two
methods, a detailed intrageneric structure of many bacterial genera remains to be resolved. In order to evaluate the usefulness of \textit{gyrB} as a phylogenetic marker, to fill the resolution gap between the 16S rRNA sequence analysis and DNA-DNA hybridization study, groupings based on the analysis of the \textit{gyrB} sequences and DNA-DNA hybridization studies have been compared. Scientists have worked extensively on strains of genus \textit{Acinetobacter} of \textit{Proteobacteria} gamma subclass (Yamamoto, S. and Harayama, S., 1998), strains of genus \textit{Comamonas} of \textit{Proteobacteria} beta subclass, strains of genus \textit{Caulobacter} of \textit{Proteobacteria} alpha subclass, strains of \textit{Cytophaga-Fravobacterium-Bacteroides} complex (Suzuki et. al., 1999), strains of genus \textit{Streptococcus} of low G+C Gram-positive bacteria, strains of genera \textit{Micromonospora} (Kasai, H. et. al., 2000) and \textit{Mycobacterium} in \textit{Actinobacteria}, etc. All results showed that the grouping by the phylogenetic analysis based on the \textit{gyrB} sequences was congruent with the grouping based on DNA-DNA hybridization. These results suggest that \textit{gyrB} is the better phylogenetic marker for the classification of bacterial strains at the species level. The \textit{gyrB}-based phylogenetic classification was applied to the analyses on the intergeneric and intrageneric relationships on genus \textit{Pseudomonas} (Yamamoto, S. and Harayama, S., 1998; Yamamoto et. al., 1999). In these papers, they analyzed the phylogenetic structures based not only on \textit{gyrB} but also on \textit{rpoD}, which is the gene for RNA polymerase sigma subunit, as molecular phylogeny deduced from a single locus may be unreliable because of the stochastic nature of base substitutions or due to rare horizontal gene transfer events. According to their data, the basic topologies of these two NJ trees were similar to each other but slightly different in detail. Theoretically, the influence of stochastic drift on the rate of evolution cannot be eliminated from the molecular phylogeny. Hence, these minor discrepancies may have their origin in such drift. If this were the case, the use of longer sequences, e.g. the combined \textit{gyrB} and \textit{rpoD} sequences, in the analysis would give more accurate estimate of the phylogeny. Sawada et. al., (1999) showed that the \textit{gyrB} and \textit{rpoD} genes had not
experienced any intergroup horizontal gene transfer within *P. syringae* but have been stable and evolved along with the *P. syringae* genome. They argue the possibility of lateral transfer pathogenicity-related gene, *argK* or *argK-toxR* gene cluster based on the phylogenetic studies based on these genes including *gyrB* and *rpoD*. Their data suggests that *gyrB* and *rpoD* can be used as indicators of genome evolution.

**Application of *gyrB* for identification and detection of bacteria**

Because of the high ratios of evolution in the *gyrB* nucleotide sequences, it was very easy to design PCR primers or probes having specificity for clusters of species, subspecies, or higher. On the other hand, the design of such primers or probes based on the variable regions of 16S rRNA presents a problem, because the sequence similarity among the variable regions does not always guarantee a close phylogenetic relationship (Yamamoto, S. and Harayama, S., 1998). In conclusion, classification, identification and detection systems based on the *gyrB* can be very useful for molecular diagnostics in microbial ecology and other fields of bacteriology. Several recent reports showed more closely related bacteria can be differentiated by *gyrB*-based methods (Kasai, H. et. al., 2000). The *gyrB*-based detection system has also been developed by Venkateswaran et. al., (1998), Yamada et. al., (1999), Vuddhakul et. al., (2000). For microbial ecological analyses, Watanabe et. al., (1998) developed the *gyrB*-based competitive PCR methods. Watanabe et. al., (1999) also show applicability of the *gyrB* to resolve flocculating and non-flocculating bacteria in activated sludge system.

**Ribosomal Intergenic Spacer (RIS)**

The development of molecular genetics methodology during the 1980’s led to the use of ribosomal RNA molecule as a good indicator of phylogenetic relationships (Woese, 1987). The genes for prokaryotic rRNA (5S, 16S and
23S clusters) appear in several different copies at different loci within the genome having different flanking restriction sites locations.

Despite of the usefulness of 16S rDNA in characterizing microbial communities, 16S rDNA sequences are sometimes not divergent enough to distinguish species of the same genus. Bacterial strains with distinct physiology have been reported to have identical 16S rRNA gene. Unlike the conserved 16S rDNA, 16S-23S rDNA ribosomal intergenic spacer (RIS) are variable. The RIS appears to be genetically stable over many generations of a particular bacterial strain (Zhongtang Yu and William W. Mohn, 2001). Thus, the variability in the spacer region between 16S and 23S rRNA genes provides yet another tool to distinguish closely related organisms, which, otherwise, are difficult to distinguish using 16S rDNA sequencing.

In this method different strains were differentiated by PCR amplification of the inter spacer region between 16S and 23S rRNA genes. Significant sequence or length heterogeneity within 16S-23S spacer region of bacteria has been used to discriminate between different species of bacteria and to detect heterogeneity within species, whereas closely related microorganisms often have very similar spacer regions.

Chromosomal DNA from the bacterial isolates is used as a template in PCR with oligonucleotide primers complementary to highly conserved sequences flanking the spacer regions between 16S and 23S rRNA genes. Spacer region amplification of the rDNA (PCR-Rribotyping) provides simple, rapid, reproducible and accurate differentiation and identification of microbial strains.
Restriction fragment length polymorphism (RFLP) using 23S+16S rRNA probe or ribotyping

First, whole-genomic DNA is isolated from bacterial cells, which is then digested with restriction enzymes that cut the DNA into fragments approximately 1000 to 20,000 base pairs (bp) in length. These are separated electrophoretically according to their size in an agarose gel to give a complex genomic "fingerprint". These patterns can be analysed and stored as computer files and compared with other profiles. The advantages of this method is its wide applicability and universal use, whilst the disadvantages include the need for pure, good quality DNA for analysis and the exceptionally complex DNA fragment patterns that are difficult to interpret (Sambrook et al., 1989). As the number of bands generated by restriction enzymes are so numerous, reducing the number of component bands can facilitate interpretation of the data. One such approach is to transfer DNA from agarose gels to a membrane (Southern blotting) and to subsequently apply a labelled probe which hybridises only to a subset of DNA fragments that can be specifically recognised by the probe. Variations in position and number of these fragments are called as restriction fragment length polymorphism, RFLP.

The most widely used probe involves the use of ribosomal DNA, usually 16S, 23S or 16+23S rRNA genes, that are present in several copies in most bacterial species. This method is called ribotyping (Grimont and Grimont, 1991). It is a widely adopted typing method, because the same universal probe (usually derived from E. coli rDNA) can be used for different species and most species harbour multiple copies of the rRNA operon. Thus banding patterns composed of several bands are obtained, depending on the cleavage sites of the restriction enzyme within the probe. The discriminatory power is good with bacterial species carrying multiple copies of rRNA operons. The main benefits of ribotyping are the relatively simple banding
patterns, good reproducibility and possibility for automatisation. The disadvantages include the laborious procedure, taking several days using standard methods, and low discriminatory potential when applied to some bacterial species. Nonetheless, it is a widely accepted taxonomic tool as a species identification method with some difficult groups of bacteria due to its benefits (Swaminathan et. al., 1993).

It was demonstrated that the ubiquitous and polymorphic ribosomal DNA loci are highly conserved in the eubacterial kingdom (Pace et. al., 1986). This finding led to the use of E.coli rRNA as a universal probe for studying the restriction patterns for taxonomic purposes (Grimont and Grimont, 1986). Later Stull and co-workers (1988) independently developed this concept of universal probe, demonstrating its utility for epidemiological studies and coined the term ribotyping. The three species of rRNA (23S, 16S, and 5S) are present in equimolar amounts in 70S ribosome particles. They are co-transcribed as a large 30S precursor RNA, which undergoes maturation process to yield the three different mature species of rRNA. The corresponding genes are thus organized into a polycistronic transcription unit, the \textbf{rrn operon}. Starting from the promoter, the individual genes are arranged in the following order: 5'-16S-23S-5S-3'. A typical \textit{E.coli} rrn operon spans 6,000 to 7,000 bp of DNA. While most genes in bacteria are present in one copy, rrn operons are unique in the property that they are present in multiple copies. Depending on the species, the number of copies vary from 2 to 11 per bacterial cell.
The *E. coli* chromosome contains seven rrn operons located in a region that spans half of the genome. Rare exceptions have been reported for some strains of *Mycoplasma* and *Mycobacteria* spp. and for *Borrelia burgdorferi* strains, which carry only one rrn operon (Postic et al., 1990). The number of rrn operons is an important consideration in the context of ribotyping because the higher the number, the higher the potential discriminative power of the technique.

Between the genes of 16S and 23S rRNAs is a spacer region that contains genes encoding several tRNAs and several direct repeat sequences. These regions are heterogeneous. In addition some rrn operons also have tRNA genes in their 3'-flanking regions (Fellner and Ebel, 1970). Depending on the operon, these genes vary in nature and in number. The rrn nucleotide sequences encoding for 16S and 23S rRNAs appear to have changed little.
during evolution. A high degree of homology for these sequences exists among various bacteria, despite the genetic diversity of the remainder of the chromosome. Grimont and Grimont. (1986) in their broad-ranging study used $P^{32}$ labeled *E. coli* 16S + 23S rRNA to probe rrn loci in 40 different species of gram-positive and gram-negative bacteria. They found that the probe reacted with the DNA of species that were phylogenetically remote from *E. coli* and that a wide variety of restriction fragment patterns was observed among the strains studied. The explanation is that the highly conserved coding sequences in each rrn operon are surrounded by the heterogeneous spacer and flanking sequences which vary from operon to operon in a given species and from one species to another. The peculiar structure of rrn operons allows both a broad-spectrum hybridization of *E. coli* RNA probe and a high level of polymorphism. RFLP patterns actually reveal two levels of genetic diversity. Some of the observed fragments are species specific while others within a given species are strain specific. Accordingly, ribotyping can be considered as a tool for taxonomic as well as epidemiological studies. Several studies have shown that these polymorphic markers are stable *in vitro* and *in vivo* during an outbreak, and the value of probing rrn loci with *E. coli* rRNA has now been clearly demonstrated for the epidemiological typing of any species of bacteria (Grimont and Grimont, 1986; Irino et. al., 1988; Stull et. al., 1988).

Most often ribotyping uses labeled probes containing *E. coli* 23S and 16S rRNA sequences (Grimont and Grimont, 1986; Stull et. al., 1988) but, the nature of the probe and the mode of labeling may vary. Initially Grimont and Grimont used commercially available *E. coli* 16S and 23S rRNA (Grimont and Grimont., 1986), a material known to be heavily contaminated with 5S rRNA. Similarly Stull et. al., (1988) used total rRNA. Now the probe most often used is DNA. It can be synthesized directly into cDNA from rRNA by reverse transcriptase. A recombinant plasmid in which the rrn DNA has been cloned or a synthetic oligonucleotide constructed from the 16S or 16S+23S rRNA
gene sequence of *E. coli* rRNA can also be used as probe. De Byser et al., (1989) used *Bacillus subtilis* rRNA to probe *Staphylococcus* DNA. Some workers have used homologous probes derived from the specific rRNA sequences of the organism under study, as has been done in studies with *Mycoplasma species*. (Yoge et al., 1988), *Pseudomonas cepacia* and *Haemophilus influenzae* (Stull et al., 1988), *Legionella sp.* (Saunders et al., 1988), *Compylobacter sp* (Patton et al., 1991) and *Corynebacterium jeikeium* (Pitcher et al., 1990). Since the hybridization depends on the region of DNA that is probed the results from separate studies are quite often not directly comparable.

Initially the probes used were labeled with $^{32}\text{P}$. For RNA probes and oligonucleotide end labeling is used, and nick translation or random oligopriming is used for cloned probes (Sambrook et al., 1989). One advantage of isotopic probes is that relatively few steps are required for hybridization and subsequent washing. But the effective non-isotopic cold-labeling systems that have now been developed offer many advantages over the radioactive systems including safety, easy disposal, and stability of the probes, which could be stable for several months. Moreover, the results are as good as or even superior to those obtained with isotopic labelling. Ribotyping may be an extremely powerful epidemiological tool when the classical phenotyping methods fail to discriminate between strains. The use of rRNA gene restriction pattern was first proposed by Grimont and Grimont (1986) as potential taxonomic tools. Adopting the same approach, many workers have successfully used the ribotyping method for epidemiological studies of nosocomial infections.

aeruginosa infections in burn unit with four restriction enzymes (BamHI, CiaI, EcoRI and PstI). The 55 unrelated isolates could be classified into 33 ribotypes. He found that the typeability and reproducibility of ribotyping are very high.

Garaizer et al. (1991) using forty-five isolates of Enterobacter cloacae recovered from 36 patients in nine hospitals, compared ribotyping with O-serotyping, phage susceptibility and biotype pattern. By combining the results of the three phenotyping approaches, 29 different strains were identified. With three restriction enzymes, EcoRI, BamHI and HindIII, 30 ribotypes were identified. EcoRI patterns alone were sufficient to distinguish between the large majority of strains, illustrating the importance of the choice of restriction enzymes. EcoRI generated largest number of fragments over the widest size range. The author concluded that ribotyping is a highly discriminatory and reproducible method for typing E. cloacae.

Epidemic methicillin-resistant isolates from Australia and from the United Kingdom showed a high degree of similarity. All the strains could be typed including the non-phage-typeable strains. Ribotyping was not affected by changes in plasmid, transposon, enterotoxin A or phage content. It was concluded that ribotyping may provide new clues to the clonal evolution of the S. aureus population. The capacity to differentiate Salmonella serotypes enteritidis strains by ribotyping with a mixture of PstI and SphI, PCR ribotyping and RAPD typing was evaluated on a series of 65 strains associated with human infection and 11 reference strains. All the methods typed all the strains, however, only ribotyping showed good reproducibility and sensitivity. Twenty-two ribotypes were obtained, whereas by PCR ribotyping a single profile was found and by RAPD typing one, two or three RAPD types were identified with the primers MK22, OPB6 and OPB17, respectively. The banding patterns defining the ribotypes were interpreted more easily and the patterns could be compared more accurately than
banding patterns defining Random amplification of polymorphic DNA (RAPD) types. This work indicated that ribotyping was the most useful genetic procedure to differentiate *enteritidis* strains (Landeras and Mendoza, 1998).