National Library
of Canada

Acquisitions and
Bibliographic Services Branch

395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale
du Canada

Direction des acquisitions et
des services bibliographiques

395, rue Wellington
Ottawa (Ontario)
K1A 0N4

*Your file Votre référence*

*Our file Notre référence*

## NOTICE

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

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

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

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

## AVIS

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

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

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

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

Canada

# Investigation of a ribosomal RNA operon of *Frankia sp.* strain AcN14a

Mireille Prud'homme

Thesis submitted to the
School of Graduate Studies and Research
University of Ottawa
in partial fulfillment of the requirements for the
M.Sc. degree in the

Ottawa-Carleton Institute of Biology

Thèse soumise à
l'École des études supérieures et de la recherche
Université d'Ottawa
en vue de l'obtention de la maîtrise ès sciences à

l'Institut de biologie d'Ottawa-Carleton

ISBN 0-315-95988-6

Canada

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

## ACKNOWLEDGEMENTS

I thank Dr. D.A. Johnson for giving me the opportunity to work with him in his laboratory. His guidance, encouragement and financial support during the course of this project made this work possible. I am also thankful for the friendship and laughs brought by the laboratory members and I would like to thank Dr. L. Bonen, Dr. J. Saddler and Dr. C. Wyndham who served as members of my research committee. In addition, I wish to thank Dr. Hattori and Dr. B. Baum at Agriculture Canada for their help with the sequence analysis.

Lastly, I would like to thank my family for their encouragement, especially my husband François, for the love, support and encouragement they have provided throughout this project.

## ABSTRACT

Members of the genus *Frankia* are gram positive, filamentous actinomycetes that infect roots and induce the formation of nodules on alders as well as on a variety of other non-leguminous (actinorhizal) plants. In root nodules, *Frankia* forms specialized cell clusters that can reduce atmospheric nitrogen to ammonia. This ability of alders to symbiotically fix nitrogen at high rates in soil makes them potentially useful in land reclamation, reforestation and agroforestry projects (Tjepkema *et al.*, 1986). However, in contrast to the nitrogen-fixing eubacteria of the genus *Rhizobium* which form symbiotic association with leguminous plants, little is known about the genetics of the actinomycete *Frankia*.

With the aim of increasing our present knowledge of *Frankia*, we have investigated the structure of the ribosomal RNA genes from *Frankia sp.* strain AcN14a. Preliminary Southern blot analysis of restricted *Frankia* AcN14a genomic DNA indicated the presence of two copies of the 23S rRNA gene. A genomic library was constructed in EMBL3 and two types of clones, each containing a different set of rRNA genes, were identified by restriction analysis. We have indirect evidence suggesting that the three rRNA genes of *Frankia sp.* strain AcN14a are organized into an operon as they are in almost all eubacteria so far studied (Srivastava and Schlessinger, 1990). First, sequence analysis of a part of one of these clones, λFr-8, revealed that the rRNA genes are clustered, with the order 5' 16S-23S-5S 3' as they are in most eubacterial rRNA operons. In addition, putative promoter sequences upstream from the 16S rRNA gene and a putative terminator sequence downstream from the 5S rRNA gene could be identified. Secondly, sequences similar to maturation signals of other eubacterial polycistronic RNAs have been identified in the putative primary transcript of *Frankia* strain AcN14a. Finally, the putative primary rRNA gene transcript of *Frankia* AcN14A could be folded into a secondary structure which shares many features with that proposed for other eubacteria, especially *S. ambofaciens* (Pernodet *et al.*, 1989). As in the *S. ambofaciens* rrnD operon, no tRNA gene sequences were detected in the 16S-

23S rRNA gene spacer region nor were any detected in the region downstream from the 5S rRNA gene.

The rRNA genes of *Frankia* AcN14a have been compared to those of *E. coli* and other eubacterial species. *Frankia* AcN14a rRNA genes show over 98% identity to those of *Frankia* ORS020606 and usually show over 80% identity to other actinomycete rRNA sequences. The mature 16S, 23S and 5S rRNAs are predicted to contain 1509, 3105 and 120 nucleotides, respectively. Thus, *Frankia* 16S rRNA is smaller than that of *E. coli* whereas the 23S rRNA of *Frankia* is longer than its *E. coli* counterpart. In spite of these size differences, all three rRNAs could be folded into secondary structures similar to the *E. coli* models and most tertiary interactions proposed for *E. coli* rRNAs could be formed in *Frankia* AcN14a 16S and 23S rRNAs.

Most of the size variation observed between *Frankia* and *E. coli* rRNAs is accommodated by change in the number of nucleotides present in specific regions of the rRNA molecules. These regions have already been defined as variable by Raué *et al.* (1988). They encompass helices 6, 10-11 and 18 of the 16S rRNA and regions in proximity to helices 13-14, 50-51 as well as helix 41 of the 23S rRNA. Two additional regions of the 16S rRNA for which the nucleotide sequence have been reported to vary among *Frankia* were also analyzed (Nazaret *et al.*, 1991).

While some of these regions, such as helix 11 of the 16S rRNA, are highly variable among *Frankia*, other regions, such as helix 18 of the 16S rRNA, are more or less conserved within the genus but variable among actinomycetes. An unrooted phylogenetic tree has been constructed based upon variable regions of the 16S rRNA of selected actinomycetes and was compared to one based upon complete 16S rRNA sequences. Some regions of both 16S and 23S rRNAs showing promise for the targeting of strain- as well as species-specific oligonucleotide probes, have also been identified.

Finally, the potential usefulness of rDNA RFLP analysis for the categorization of *Frankia* isolates has been investigated. Hybridization of both

16S and 23S rDNA specific probes to total DNA digested with BamH1, revealed sufficient polymorphism to allow classification of twenty-six isolates into six groups (A to F). In addition, eight other isolates exhibited unique patterns.

## RÉSUMÉ

Les membres du genre *Frankia* sont des actinomycètes capables d'induire la formation de nodules sur les racines de certaines plantes non-légumineuses tel les aulnes. Ces nodules sont le site de la fixation de l'azote atmosphérique en ammoniac. Cette habileté de l'aulne à fixer l'azote symbiotiquement en fait un candidat de choix pour les projets de reboisement, de récupération des terres et d'agroforesterie (Tjepkema *et al.*, 1986). Toutefois, contrairement aux membres du genre *Rhizobium* qui forment des associations symbiotiques avec les légumes, nos connaissances sur la génétique de *Frankia* sont très limitées.

Dans le but d'approfondir nos connaissances sur *Frankia*, nous avons étudié la structure des gènes d'ARN ribosomaux (ARNr) de la souche AcN14a. Des données provenant de transferts de type Southern ont suggéré la présence de deux copies du gène d' ARNr 23S chez cette souche. Une banque génomique a été construite dans EMBL3 et deux types de clones, représentant les deux groupes de gènes d'ARNr ont été identifiés par analyse de restriction. La séquence de nucléotides d'une portion du clone génomique λFR-8, contenant une copie des gènes de chaque ARNr, est présentée. Comme chez la majorité des eubactéries, ces gènes sont probablement organisés en opérons (Srivastava and Schlessinger, 1990). Les gènes pour les trois différentes espèces d'ARNr sont liés dans l'ordre 5' 16S-23S-5S 3'. Des promoteurs potentiels ont été repérés dans la région en amont du gène d'ARNr 16S et un terminateur potentiel a été identifié dans la région en aval du gène d'ARNr 5S. De plus, des séquences semblables à des signaux de maturation trouvés chez d'autres ARN polycistroniques eubactériens ont été identifiés dans le transcript d'ARNr primaire potentiel. Enfin, ce transcript d'ARNr primaire potentiel peut former une structure secondaire ayant des caractéristiques en commun avec ceux proposés pour d'autres eubactéries, tout particulièrement celui de *Streptomyces ambofaciens* (Pernodet et al., 1989). Comme chez ce dernier, aucun gène d'ARN de transfert (ARNt) n'est présent entre les gènes d'ARNr ou dans la région en aval du gène d'ARNr 5S.

La séquence de nucléotides des gènes d'ARNr de la souche AcN14a de *Frankia* a été comparée à celles de *Escherichia coli* et d'autres eubactéries. Les gènes d'ARNr des deux souches de *Frankia*, AcN14a et ORS020606, sont identiques à plus de 98% et sont généralement à plus de 80% identiques à ceux d'autres actinomycètes.

Les ARNr 16S, 23S et 5S déduits de leur gène respectif contiennent 1511, 3104 et 120 nucléotides. Ainsi l'ARNr 16S de *Frankia* AcN14a est plus court que celui de *E. coli* mais son ARNr 23S est plus long que celui de *E. coli*. Malgré cette différence, les trois ARNr ont une structure secondaire similaire aux modèles proposés pour les ARNr de *E. coli*. De plus, la plupart des interactions proposées pour les ARNr de *E. coli* sont possibles pour les ARNr 16S et 23S de *Frankia* .

La différence de longueur entre les ARNr de *Frankia* et *E. coli* sont le résultat d'une différence dans le nombre de nucléotides présents dans des régions spécifiques des ARNr. Ces régions correspondent aux hélices 6, 10-11 et 18 de l'ARNr 16S et aux hélices 13-14, 41 et 50-51 de l'ARNr 23S. Deux autres régions de l'ARNr 16S pour lesquelles la séquence de nucléotides diffère parmi des membres du genre *Frankia* ont aussi été analysées (Nazaret et al., 1991). Toutes ces régions ont déjà été décrites comme étant variables chez les ARNr des procaryotes (Raué *et al.*,1988).

Le niveau de conservation de la séquence de nucléotides varie entre les différentes régions. Certaines, comme l'hélice 10 de l'ARNr 16S, semblent très variables chez *Frankia*, d'autres comme l'hélice 18 de l'ARNr 16S sont plus conservées chez *Frankia* mais varient chez les actinomycètes. Ainsi, ces régions pourraient être utiles pour le ciblage de sondes spécifiques à une souche, une espèce ou au genre. Un arbre phylogénétique construit à partir des régions variables de l'ARNr 16S de certains actinomycètes a été comparé à un second basé sur des séquences d'ARNr 16S complets.

Finalement, l'utilité de l'analyse du polymorphisme de longeur des fragments de restriction de l'ADNr 16S et 23S pour la catégorisation d'isolats de *Frankia* a été démontrée. Les patrons d'hybridation de deux sondes spécifiques

à l'ADNr 16S et 23S ont démontré suffisament de polymorphisme permettant la classification de vingt-six isolats en six groupes (A-F). Des patrons d'hybridation uniques ont aussi été identifiés chez huit autres isolats.

## TABLE OF CONTENTS

**LIST OF FIGURES**

## LIST OF TABLES

## LIST OF ABBREVIATIONS

| | |
|---|---|
| A: | adenosine |
| bp: | base pair(s) |
| *B. subtilis*: | *Bacillus subtilis* |
| cDNA: | complementary DNA |
| C: | cytosine |
| Da: | Dalton |
| dATP: | deoxyadenosine triphosphate |
| dCTP: | deoxycytosine triphosphate |
| DNA: | deoxyribonucleic acid |
| *E. coli*: | *Escherichia coli* |
| EDTA: | ethylene diamine tetraacetic acid |
| g: | gram |
| G: | guanosine |
| kb: | kilobase pair(s) |
| L: | litre |
| *M. bovis*: | *Mycobacterium bovis* |
| *M. lepraerium* | *Mycobacterium lepraerium* |
| *M. phlei*: | *Mycobacterium phlei* |
| *M. segmatis*: | *Mycobacterium segmatis* |
| *M. luteus*: | *Micrococcus luteus* |
| °C: | degree Celcius |
| pfu: | plaque forming units |
| PNFI: | Petawawa National Forestry Institute |
| *P. thermophila*: | *Pseudonocardia thermophila* |
| RNA: | ribonucleic acid |
| rRNA: | ribosomal RNA |
| *S. hirsuta*: | *Saccharopolyspora hirsuta* |
| *S. ambofaciens*: | *Streptomyces ambofaciens* |
| *S. coelicolor*: | *Streptomyces coelicolor* |
| *S. griseus*: | *Streptomyces griseus* |
| *S. lividans*: | *Streptomyces lividans* |

| | |
|---|---|
| *S. rimosus*: | *Streptomyces rimosus* |
| SDS: | sodium dodecylsulphate |
| t/yr: | ton per year |
| T: | thymidine |
| tRNA: | transfer RNA |
| U: | uridine |
| UV: | ultraviolet light |
| *Z. mays*: | *Zea mays* |

## CHAPTER ONE: INTRODUCTION

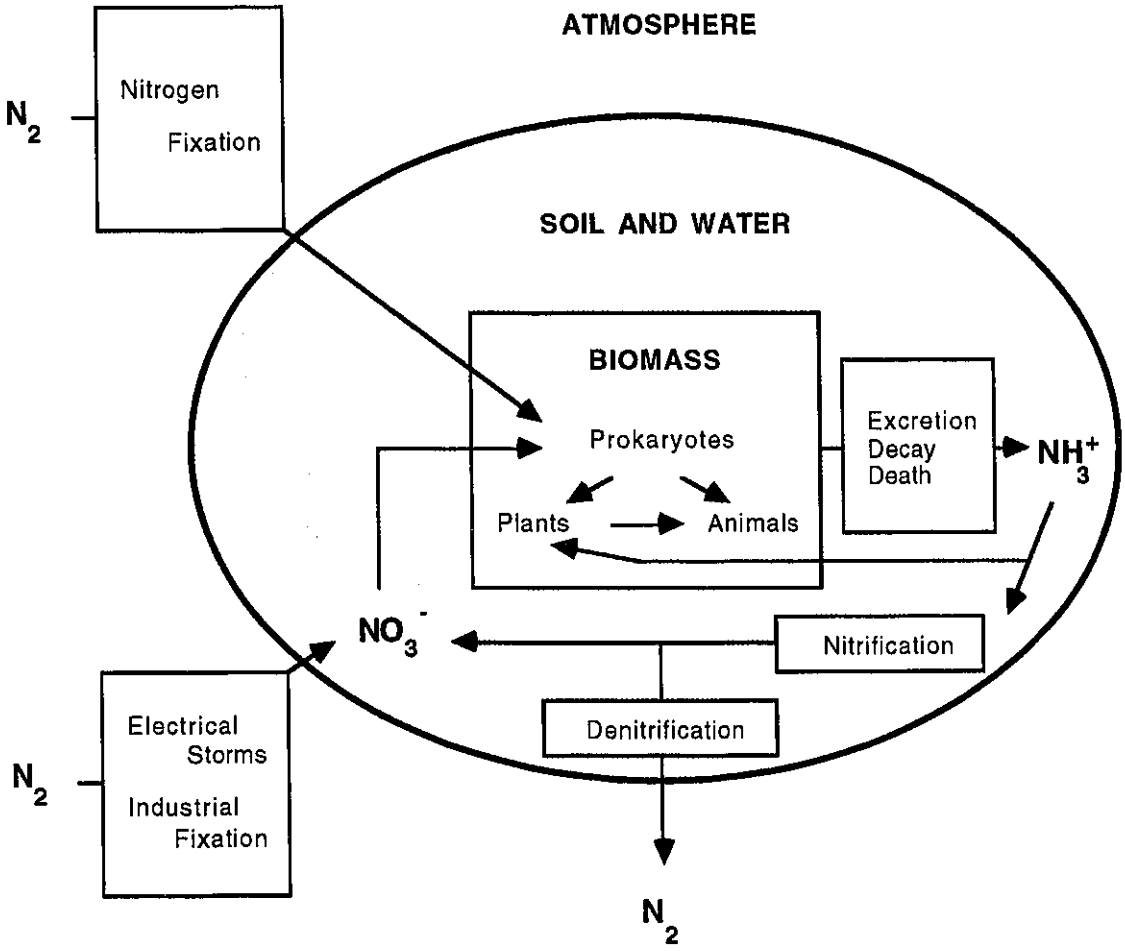
Nitrogen is essential for all biological systems since it is a component of many of the compounds which are necessary for life such as amino acids and DNA. A simplified version of the nitrogen cycle is presented in Figure 1. Most of the nitrogen occurs in the atmosphere in its most inert form ($N_2$), and thus is not readily available to plants. Plants normally acquire nitrogen from the soil in an inorganic form, mainly nitrate and ammonium ($NO_3^-$ and $NH_4^+$). Most of these compounds are produced by microorganisms (e.g. nitrifying bacteria) by the breakdown of more complex molecules in decaying materials. The remainder of the utilizable nitrogen in the soil results from the fixation (reduction) of atmospheric nitrogen to ammonia by various prokaryotic microorganisms, by the application of fertilizers (products of industrial fixation) and lightning. Nitrate is also lost from soils by denitrification. In this process, anaerobic bacteria transform nitrate into nitrogen ($N_2$) and other gases (NO, $N_2O$ and $NO_2$). The nitrogen fixers represent the major contributor to the total annual fixation rate. Their contribution is estimated at 122 x $10^6$ t/yr which represents half the nitrogen fixed in one year (Newton and Burgess, 1981).

Several taxonomically diverse prokaryotic organisms have the capacity to reduce nitrogen into a form utilizable by plants. This ability occurs in a variety of eubacteria and in methanogenic archaeobacteria (e.g. Belay *et al.*, 1984; Murray and Zinder, 1984). As far as is currently known, no eukaryotic organisms fix nitrogen (Sprent and Sprent, 1990).

Although most nitrogen-fixers live as free-living diazotrophs (Sprent and Sprent, 1990), some of them also have the ability of associating with eukaryotic organisms, usually plants. The degree of association between the two organisms varies greatly. In the most complex associations observed with plants, the nitrogen-fixing microorganisms live in the nodules where they gain access to the energy provided by plant photosynthesis. In return, the plant receives a supply of ammonia

**FIGURE 1 THE NITROGEN CYCLE**

A simplified version of the nitrogen cycle. The cycle is a modified version of the one published by Sprent and Sprent (1990).


ATMOSPHERE
Nitrogen Fixation
$N_2$
SOIL AND WATER
BIOMASS
Prokaryotes
Plants
Animals
Excretion
Decay
Death
$NH_3^+$
$NO_3^-$
Nitrification
Denitrification
Electrical Storms
Industrial Fixation
$N_2$
$N_2$

independent of the presence of fixed nitrogen in the soil (reviewed by Schubert, 1986). One of the most well characterized of these symbiotic associations is that between leguminous plants and bacteria of the genera *Rhizobium* (e.g. *R. meliloti* and alfalfa) and *Bradyrhizobium* (e.g. *B. japonicum* and soybean). In addition nitrogen-fixing symbioses have been observed in a variety of non-leguminous (actinorhizal) plants. Recent investigations have been focused on alder (*Alnus sp.*), the actinorhizal genus capable of living in symbiosis with the nitrogen-fixing actinomycete *Frankia*, because of its potential use in the forestry industry.

In contrast to *Rhizobium*, very little is known about the genetics of *Frankia*. Until recently, most studies on the subject focused upon the characterization of the nitrogen fixation gene cluster (Normand *et al.*, 1988; Ligton and Nagas, 1987; Nittayajarn *et al.*, 1990) and the gene coding for glutamine synthase II (Rochefort and Benson, 1990). We have characterized, cloned and sequenced the rDNA region from the *Frankia sp.* strain AcN14a which can form a symbiotic association with the speckled alder (*Alnus incana*). The complete nucleotide sequence of the three ribosomal rRNA genes will further increase our present knowledge of the genetics of this important nitrogen-fixing organism.

The present chapter is divided into two sections. The first section consists of an overview of the *Frankia-Alnus* symbiotic association, its potential economical and ecological applications, as well as a description of the problems encountered in their study. In the second section, different characteristics of the rRNA genes and the use of rRNA sequences for phylogenetic as well as ecological studies are reviewed.

## The actinorhizal symbiosis

Members of the genus *Frankia* are gram positive, filamentous actinomycetes that infect roots and induce the formation of nodules in a variety of non-leguminous (actinorhizal) plants. These actinorhizal plants

are primarily trees and shrubs belonging to at least 23 taxonomically diverse genera, and include members of the genus *Alnus* (alder). Some of these genera are listed in Table 1. Not all members of a family can form such associations. More surprisingly, the apparent absence of close phylogenetic relationships among most host-plant groups suggests a polyphyletic origin for actinorhizal symbioses (Mullin *et al.*, 1990).

Alder trees satisfy their total nitrogen requirements through nitrogen fixation and/or soil mineral nitrogen depending upon the availability of soil nitrogen. In a recent study done in a black alder (*Alnus glutinosa*) stand, most of the nitrogen present in leaves (94%) was found to be derived from fixation (Beaupied *et al.*, 1990).

The alders are the most important actinorhizal plants in temperate forests. In addition to their ability to form symbiotic associations, trees of the genus *Alnus* are adapted to a variety of environmental conditions such as water and temperature stresses (Hall and Burgess, 1990). Alders are pioneering trees on soils, especially with low nitrogen content. Due to their low competitive ability towards other tree species, alders are, however, rapidly replaced by other non-actinorhizal trees when the nitrogen content of the soil has been improved. A good example is the colonization by actinorhizal plants of nitrogen-poor glacial deposits in Europe and North America before being displaced by non-fixing plants as soil nitrogen increased (Silvester, 1977). A more recent example is the presence of a dense stand of red alder (*Alnus rubra*) on the Mount Saint Helens mudflow deposit only 4 years after the May 1980 eruption (Heilman, 1989). Alder stands can also exist for many decades without losing their nitrogen-fixing activity in environments where other trees do not grow well. As a consequences, alders are primarily found in harsh environments such as open sites, wetlands, peatlands and in cold areas (reviewed by Tjepkema *et al.*, 1986).

**TABLE 1 LIST OF SELECTED GENERA OF ACTINORHIZAL TREES AND THEIR FAMILIES**

| Families | Genera |
|---|---|
| Betulaceae | Alnus |
| Casuarinaceae | Allocasuarina, Casuarina |
| Coriariaceae | Coriaria |
| Datiscaceae | Datisca |
| Elaeagnaceae | Elaeagnus, Hippophae, Sherpherdia |
| Myricaceae | Comptonia, Myrica |
| Rhamnaceae | Ceanothus, Discaria |
| Rosaceae | Cercocarpus, Dryas, Purshia |
| Ulmaceae | Aphananthe, Celtis |
| Zamiaceae | Macrozamia |

The list is an adapted version from the one published by Brewbaker (1989).

## *Alnus-Frankia* symbiotic association - potential uses

The ability of alders to symbiotically fix nitrogen at high rates in soil makes them potentially useful in land reclamation, reforestation and agroforestry projects. For this reason alders have been one of the genera focused on by the International Energy Agency (IEA) Forestry Energy Agreement (Hall and Burgess, 1990). Improvement of soil nitrogen content is thought to occur via leaf fall, excretion from roots and decay of nodules as well as other parts of the plant; rates varying between several tens to hundreds of kilograms of nitrogen per hectare per year have been reported (reviewed by Sprent and Sprent, 1990).

Excess nitrogen not used by alders is thought to be readily available to associated non-fixing trees (Beaupied *et al.*, 1990); therefore, forest management practices could be changed so that crop rotation might be used. The red alder (*Alnus rubra*) could form part of such a rotation. In the western USA, Douglas-fir (*Pseudotsuga menziesii*) seedlings planted on a former red alder site showed a significant increase in biomass compared to seedlings planted on a former Douglas-fir site (Brozek, 1990). In addition alder grown in mixed stands has been shown to be beneficial to the growth of Douglas-fir (Binkley, 1981) or Black cottonwood (*Populus trichocarpa*)(Pezeshki and Chadwick, 1985). Moreover, the red alder trees in such plantations could be cut for pulpwood (Sprent and Sprent, 1990) or timber (Heilman, 1989). In both North America and Europe alders give high timber yields of 5 or more tons per hectare per year (Sprent and Sprent, 1990).

Species of *Alnus* are also widely used to colonize waste and reclaimed land. Common types of area laid waste by man are those associated with mining activities. Black alder has been successfully used for this purpose in the United States and Europe (e.g. Heilman, 1982).

In summary, alders have already been shown to play an important role in forest management, biomass production, and regeneration of

disturbed lands and in the future, these roles can be expected to become increasingly important.

### Improvement of the actinorhizal symbioses

In the wild, alders are usually nodulated, indicating that the actinomycete *Frankia* is a common soil microorganism. However, large differences in their ability to fix nitrogen have been observed for the different *Frankia* strains, both between strains infecting different plant species (Weber *et al.*, 1987) and between strains infecting the same plant species (Hooker and Wheeler, 1987; Bloom *et al.*, 1989). Moreover, the genetic variation between alders is large (Bousquet *et al.*, 1987). Thus optimal use of the *Alnus-Frankia* symbiosis includes improvement of the symbiosis through selection of superior genotypes of both partners. Such a selection program to find an improved *Frankia* strain involves competition experiments with other *Frankia* strains or natural populations in which the origin of the *Frankia* species within nodules will have to be determined.

### Study of the bacterial symbiont - *Frankia*

The ideal bacterial symbiont would have to be a highly competitive strain able to persist in the soil in order to reinfect the roots during the growth cycle of the trees. The present knowledge of the genetics and the ecology of *Frankia* is far from complete, mostly because of the problems encountered with their isolation, their culture and their identification.

The isolation and culture of *Frankia* are difficult to achieve mostly because of their slow growth rate. Their slow growth rate *in vitro* might be in part due to the uncertainty in the media requirements (reviewed by Tjepkema *et al.*, 1986). Since the first isolation of *Frankia* from the nodules of *Alnus rubra*, a large number of pure cultures of *Frankia* strains isolated from alder root nodules have become available (reviewed by Lalonde *et al.*, 1988). Similarly, numerous strains have been isolated from the different

actinorhizal plant species. Most of these strains have been isolated from nodules which are enriched with this organism (reviewed by Tjepkema *et al.*, 1986); therefore infective strains are preferentially selected although noninfective strains are known to exist (Hahn et al., 1989a, Nazaret *et al.*, 1991). Similarly, the presence of *Frankia* is usually determined by growing host plants in the soil (e.g. Smolander, 1990) and for this reason, the distribution and occurrence of *Frankia* in soil are probably underestimated.

Microorganisms inducing the formation of nitrogen-fixing root nodules on actinorhizal plants have been classified in the genus *Frankia* on the basis of their host specificity, morphology, cell chemistry and physiology (Lechevalier, 1984). With the isolation and the culturing *in vitro* of numerous *Frankia* strains, different groups have tried to develop methods for their classification and identification. Classical actinomycete taxonomic criteria alone have been shown to be of little help for differentiating between members of this genus. There appear to be few morphological differences among isolates other than pigmentation or hyphal size, both of which can be affected by growth conditions. One of the first attempts to classify members of the genus *Frankia* was based on cross-inoculation tests (Baker, 1987). The different *Frankia* strains were categorized into four host specificity groups (HSG): strains infective on *Alnus* and *Myrica* (HSG1), strains infective on *Casuarina* and *Myrica* (HSG2), strains infective on *Elaeagnus* and *Myrica* (HSG3) and strains infective on *Elaeagnus* (HSG4) only. Others have divided the different strains into two major phenotypic groups (A and B), on the basis of a combination of physiological, chemical, and plant infectivity characteristics (reviewed by Tjepkema *et al.*, 1986). Using a different series of criteria (primarily phenotypical characteristics), Lalonde *et al.* (1988) have assigned some *Frankia* strains to two species named *F. alni* and *F. elaeagni*.

For their part, Fernandez *et al.* (1989) have categorized different *Frankia* isolates (from various actinorhizal plants) into nine "genomic species" by the use of DNA-DNA hybridization (reassociation) techniques. *Frankia* isolates from each infectivity group (Baker, 1987) could thus be

categorized into different genomic species. Other methods such as serology, SDS-PAGE of total soluble protein, isozyme patterns and different DNA analyses have also been tested (reviewed by Lalonde *et al.*, 1988). The latter comprise techniques such as restriction pattern analysis of plasmid (Simonet *et al.*, 1985) and genomic DNA (An *et al.*, 1985; Bloom *et al.*, 1989) as well as restriction fragment length polymorphism (RFLP) analysis of the *nif*DH DNA region (Nittayajarn *et al.*, 1990). Most of these methods tend to define fairly large groups but provide insufficient resolution to distinguish between closely related strains. Others such as the analysis of plasmid DNA are of limited use in phylogenetic studies. In addition, most of these methods could not be easily used to identify *Frankia* species in nodules. Application of these methods often requires substantial quantity of material and would thus necessitate reisolation and culture of these slowly growing microorganisms in order to analyze the bacteria or their cellular content.

The difficulties with isolation and culturing of *Frankia* species could be bypassed by the use of oligonucleotide probes for the detection of specific rRNA sequences. The sensitivity of this method has been demonstrated (Hahn *et al.*, 1990) and is attractive for ecological investigations.

Partial 16S rRNA sequences of various *Frankia* strains have now been published (Hahn *et al.*, 1989a, b; Harry *et al.*, 1991). This has permitted the determination of their phylogenetic relationships (Nazaret *et al.*, 1991). Interestingly, a good correlation was obtained between the grouping obtained based on 16S rDNA sequence homology and the grouping proposed by Fernandez *et al.* (1989) based upon DNA-DNA reassociation.

In addition, 16S rDNA/rRNA sequence alignment has permitted the identification of regions where intrageneric variation was observed (Hahn *et al.*, 1990; Harry *et al.*, 1991; Nazaret *et al.*, 1991) and *Frankia*, 16S rRNA specific DNA probes have been shown to be potentially useful for the identification of *Frankia* isolates in inoculation experiments (Hahn *et al.*,

1990). Because of the abundance of *Frankia* rRNA (the target material) in nodules, the necessity for reisolation of the symbiont is eliminated.

At the time this work was initiated, only one *Frankia* 16S rRNA gene, that of the Ag45/Mut15 strain, had been partially sequenced by Hahn *et al.* (1989b) but no information on the *Frankia* 23S or 5S rRNA sequences were yet available. The 23S rRNA is approximately twice the size of the 16S rRNA and consequently can potentially contain numerous additional variable regions. Such variable regions could be used to derived oligonucleotide probes and might allow differentiation of closely related *Frankia* strains.

**Use of rRNA/rDNA for the study of *Frankia* phylogeny**

The analysis of ribosomal RNA (rRNA) as a phylogenetic tool was pioneered by Carl Woese and his colleagues (Woese and Fox, 1977; Fox *et al.*, 1980). Based on oligonucleotide cataloging, the existence of three distinct lineages (eubacterial, archaeobacterial and nuclear) was suggested. Cataloging analyses also supported the endosymbiosis theory for the origin of mitochondria and chloroplasts (e.g, Bonen and Doolittle, 1975; Zablen *et al.*, 1975). Today, partial or complete rRNA sequences are more often derived by direct sequencing of the rRNA (Lane *et al.*, 1985), sequencing of cloned DNA, or by sequencing PCR amplified material (Boettger, 1989).

The use of rRNA sequences for the study of the diversity and the phylogenetical relationships amongst *Frankia* strains has recently been reported by three different research groups (Hahn *et al.*, 1989a; Harry *et al.*, 1991; Nazaret *et al.*, 1991). In addition, the potential value of 16S rRNA specific oligonucleotide probes has been demonstrated in competition experiments between two *Frankia* strains on *Alnus glutinosa* (Hahn *et al.*, 1990).

Although almost any gene may be used as a genetic marker, rRNA genes offer distinct advantages. The characteristics of the rRNAs and their genes will be described in this section.

**rDNA conservation - Organization and primary structure**

Ribosomal RNAs are ancient molecules, functionally constant (homologous) and universally distributed. Their conservation across broad phylogenetic distances make them excellent phylogenetic "chronometers" (Woese, 1987). All these characteristics are most likely a consequence of the central role of rRNAs in protein synthesis.

The rRNAs and their corresponding genes are highly conserved. In eubacteria, there are three different ribosomal RNA genes. The typical transcription products, the mature 23S, 16S and 5S rRNAs have approximate sizes of 3000, 1500 and 120 nucleotides, respectively. The sizes of the two larger rRNAs have been shown to vary between different groups of organisms (reviewed by Noller, 1984). Such variation is often due to the presence of large deletions or expansion segments in relation to the *E. coli* rRNAs (reviewed by Brimacombe *et al.*, 1989) which are commonly used as reference sequences.

The nucleotide sequence of the two larger rRNAs contains several highly conserved as well as less conserved regions regions useful for phylogenetic distance measurements. The 5S rRNA has also been used for phylogenetic studies, but its small size limits the information obtainable from this molecule (Olsen *et al.*, 1986). The more highly conserved regions are the basis for phylogenetic analysis (Woese and Fox, 1977) and for the design of universal oligonucleotide probes and primers used for the identification and amplification, respectively, of rDNA from distantly related species. Depending on the degree of variability, the less conserved regions, in principle, allow the allocation of more closely related organisms (i.e., genus, species, and strains) by sequence analysis (Woese, 1987), oligonucleotide probing (e.g. Stackebrandt and Charfreitag, 1990), or

procedures employing polymerase chain (PCR) amplification (e.g. Nazaret *et al.*, 1991).

The organization of the rRNA genes is also conserved. In most eubacterial genomes, rRNA transcription units are arranged in operons with the order 5' 16S-23S-5S 3'. Between the 16S and 23S rRNA genes and downstream of the 5S gene are regions known as the intergenic spacer and distal spacer, respectively. These regions are also contained in the primary transcript. These two relatively short regions have been shown to contain transfer RNA (tRNA) genes in some but not all rRNA operons (reviewed by Srivastava and Schlessinger, 1990). In the intergenic spacer, an arrangement with an alanine tRNA (Bacot and Reeves, 1991), a glutamate tRNA, or tandem isoleucine-alanine tRNAs may be found. Some distal spacers have been shown to encode tandem aspartate-tryptophan tRNAs or an individual tRNA for either aspartate, isoleucine (Bacot and Reeves, 1991), threonine or tryptophan.

Variants of this rRNA gene organization have been identified (reviewed by Srivastava and Schlessinger, 1990). One of the *E.coli* ribosomal RNA operons contains two consecutive 5S rRNA genes (Duester and Holmes, 1980). In *Mycoplasma hyopneumoniae*, the 5S rRNA gene was found to be approximately 4 kb downstream of the 3' end of the 16S-23S rRNA gene unit (Taschke *et al.*, 1986). Interestingly, *Mycoplasma gallisepticum* contains one typical rRNA operon and another copy of the 16S rRNA gene separated from a still-linked 23S-5S rRNA gene unit (Chen and Finch, 1989). Similarly, a great distance separating the 16S rRNA gene and 23S-5S rRNA genes was observed in the two rRNA operons of *Thermus thermophilus* (Hartmann *et al.*, 1987). In this organism, the 16S rRNA gene has been found to be independently transcribed (Hartmann *et al*, 1989). Finally, all the rRNA genes of *Leptospira interrogans*, and probably of *Leptospira bifiexa*, have been found to be unlinked (Fukunaga and Mifuchi, 1989).

Although several different types of organization have been observed, the linkage of the rRNA genes (in the order 16S, 23S, and 5S rRNA genes),

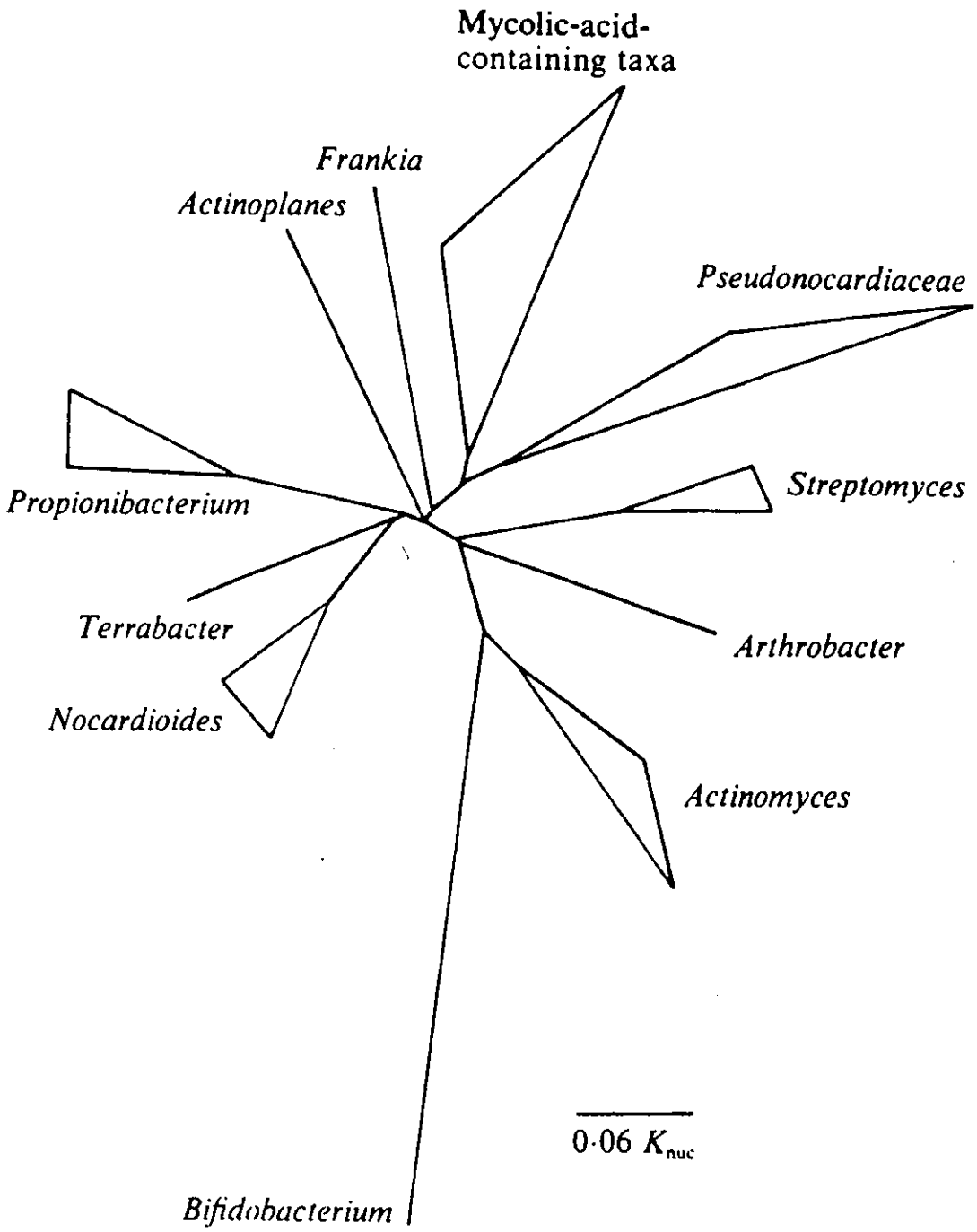
appears to be ancestral (Pace and Burgin, 1989) since this organization is also found in archaeobacteria (Garrett *et al.*, 1991). In eukaryotic organisms, the 5S gene is not linked to the other rRNA genes but a search in the primitive eukaryotic organisms needs to be done. One could suggest that linkage of the three different rRNA genes guarantees coordinate synthesis of the rRNA molecules required for ribosome assembly. However, the fact that rRNA genes are physically and transcriptionally linked in most but not all eubacteria, suggests that this organization is not essential to cell function.

We would expect the order and the organization of the rRNA genes to be conserved in *Frankia* AcN14a since a typical rDNA organization has been reported for all the actinomycetes so far examined (e.g. Suzuki *et al.*, 1987; Pernodet *et al.*, 1989). *Frankia* AcN14a rDNA region will be compared to that of *Frankia sp.* strain ORS020606 (Normand, unpublished) and *Streptomyces ambofaciens* (Pernodet *et al.*, 1989), the only two actinomycetes to have had their rRNA operons sequenced. The relative taxonomic position of *Frankia* to other actinomycete groups, such as *Streptomycetes* and those containing mycolic acid is still uncertain. The most recent unrooted phylogenetic tree that has been proposed is presented in Figure 2. In this tree, *Frankia* appears to be more closely related to the mycolic-acid containing eubacteria, such as *Mycobacteria*, *Corynebacteria* and *Nocardia* than to *Streptomyces* (Stackebrandt and Charfreitag, 1990).

The number of rRNA genes in bacterial species varies greatly (reviewed by Srivastava and Schlessinger, 1990). It has been shown that the *Escherichia coli* chromosome carries seven sets of ribosomal genes and that ten are carried in *Bacillus subtilis*. In contrast, only one or two sets of rRNA genes have been reported for *Mycoplasma* (Amikam *et al.*, 1982), *Chlamydia* (Engel and Ganem, 1987), *Thermus* (Ulbrich *et al.*, 1984), and *Mycobacterium* species (Suzuki *et al.*, 1987).

**FIGURE 2 PHYLOGENETIC TREE BASED UPON SELECTED NUCLEOTIDES OF THE 16S RIBOSOMAL RNAS**

Unrooted phylogenetic tree showing the intergeneric relationships of *Frankia* and various actinomycete genera, based on comparison of selected nucleotides of the 16S rRNA (from Stackebrandt and Charfreitag, 1990).


Mycolic-acid-containing taxa
Frankia
Actinoplanes
Pseudonocardiaceae
Propionibacterium
Streptomyces
Terrabacter
Arthrobacter
Nocardioides
Actinomyces
Bifidobacterium
0·06 $K_{nuc}$

The reasons for the multiplicity of ribosomal RNA operons in *E. coli* and *B. subtilis* are not clear. Amikam *et al.* (1982) have speculated that for those bacteria capable of growing in versatile habitats, a high number of rRNA gene copies arranged into operons would allow faster RNA synthesis and would thus enable their fast growth even under adverse conditions. In contrast, microorganisms with a slow growth rate and/or adapted to a constant environment would require fewer sets of ribosomal RNA genes. Like most soil eubacteria, members of the genus *Frankia* grow slowly and consequently, a low number of rRNA gene operons are expected to be contained in their genomes. Determination of the number of copies could be achieved by RFLP analysis.

All typical rRNA operons studied so far have been shown to be transcribed into long precursor molecules called the 30S precursor-rRNAs (30SprRNA) (reviewed by Srivastava and Schlessinger, 1989; 1990). In addition to the 16S, 23S and 5S rRNAs, the precursor molecule includes a leader sequence, two spacers and a trailer sequence. Mature rRNA species are the end result of a series of cleavages which are collectively known as rRNA processing. The details of rRNA processing are relatively well understood for *E. coli* and are thought to be similar in other eubacteria including *Frankia*.

The various enzymatic steps necessary for the production of the mature rRNA molecules fall into two categories: the primary cleavages that separate the long precursor molecule into its different gene products and the secondary cleavages that lead to the formation of the mature 5' and 3' termini of the rRNA molecules. In general, enzymes that affect the primary cleavages can cut any naked RNA substrates, whereas the terminal maturases involved in secondary cleavages require ribonucleoprotein (pre-rRNA + associated ribosomal proteins) substrates. RNase III is responsible for generating precursors for the three rRNA species. It produces the 16S precursor rRNA (p16SrRNA) that contains 115 additional nucleotides at its 5' end and an additional 33 nucleotides at its 3' terminus. The same enzyme also produces a 23S rRNA precursor (p23SrRNA) containing 7 and 8 additional

nucleotides at the 5' and 3' termini, respectively. Consequently, the downstream 5S rRNA precursor (p5SrRNA) is produced with 85 nucleotides at the 5' end and additional 3' nucleotides extending to the terminator site. The p16SrRNA and p23S rRNA species are further cleaved at both termini, but the putative enzymes for those reactions, RNase M16 and RNase M23 have not been characterized (Pace 1984; King *et al*, 1986). *E.coli* p5SrRNA is rapidly cleaved by RNase E producing a species with only 3 extra nucleotides at each termini. In *Bacillus subtilis*, the extra nucleotides at both ends of the p5SrRNA species are removed by RNase M5 (reviewed by Pace and Pace, 1990). For those operons that contain tRNAs, RNase P and RNase D are responsible for maturation of the 5' and 3' ends of these species, respectively (reviewed by King *et al.*, 1986).

**Conservation of the secondary and tertiary structures - consequence of the functional role of rRNAs in translation**

The *E. coli* 16S and 23S rRNAs have been divided into structural domains containing several hundred bases (Noller, 1984). Three major domains and a small 3' domain have been distinguished in *E. coli* 16S rRNA while six domains have been defined in the 23S rRNA (reviewed by Raué *et al.*, 1988). These domains can be individually folded through base pairing (secondary interactions) and are joined to one another by so-called tertiary interactions. The information for generating such secondary and tertiary structures is contained in the nucleotide sequence (primary structure). Secondary structure is defined by pairings of complementary sequences at least two base pairs in length. The so-called Watson-Crick (or canonical) interactions are the main determinants of the rRNA secondary structure. However, other base pairing arrangements (non-canonical), such as the commonly found G-U interaction, are also possible (Woese *et al.*, 1983). Through these different types of interactions, the rRNA can tentatively be folded into a series of hairpins and loops but with such large molecules, many alternative conformations are possible.

The most prevalent approach to analyzing secondary structure is to compare the sequences of an rRNA in related organisms (James *et al.*, 1989). This is because rRNA secondary structure has remained relatively conserved across the living kingdoms (Gutell and Fox, 1988; Woese *et al.*, 1983). Thus, even though the nucleotide sequences of rRNAs have diverged through evolution the structural organization of the molecule has been maintained through compensatory base changes. The covariation of paired residues in putative helical regions offers support for the proposed secondary structure. On the other hand, if equivalent complementarity is not present in homologous sequences from different organisms, the structure is unlikely to exist *in vivo* (reviewed by Woese *et al.*, 1983). The molecules compared must be sufficiently different to provide sequence variation with which to test pairing possibilities, yet, the molecules must be similar enough that homologous residues can be aligned with confidence. Sequence similarity of 60-80% is thought optimal for secondary structure analysis (James *et al.*, 1989). Biochemical and genetical approaches have also been used to test the validity of the proposed secondary structures (Moazed *et al.*, 1986; Triman *et al.*, 1989).

From such comparisons, secondary structure models have been constructed for both 16S and 23S rRNAs (Neefs *et al.*, 1991; Gutell and Fox, 1988). Comparative sequence analysis has also been successfully used to elucidate the secondary structure in many other, although smaller, RNA molecules such as tRNA (Sprinzl *et al.*, 1989), 5S rRNA (Wolters and Erdmann, 1988), small nuclear RNA (Siliciano *et al.*, 1987), and the RNA component of RNase P (James *et al.*, 1989).

In addition, the secondary structure of the region flanking the future 16S and 23S mature rRNAs on the precursor molecule of eubacteria appears to be conserved (King *et al.*, 1986). Sequences flanking the 16S and the 23S rRNA can potentially form two large stable RNA stems. Such structures have been proposed for *E. coli* (Bram *et al.*, 1980; Young and Steitz, 1978) as well as for many other eubacteria such as *B. subtilis* (Ogasawara *et al.*, 1983), *Mycoplasma* (Rasmussen *et al.*, 1987), *S. ambofaciens*

(Pernodet *et al.*, 1989), and are expected to be present in *Frankia*. These long stem structures have been shown to contain RNase III cleavage sites (reviewed by Srivastava and Schlessinger, 1990). The evolutionary conservation of such putative stems implies that they have a function in the pre-rRNA (King *et al.*, 1986).

Ribosomal RNA constitutes the major part of the mass of the ribosome. Ribosomal RNAs have been shown to interact with mRNA and tRNA molecules as well as with elongation factors and antibiotics, and thus participate in most if not all steps of protein synthesis. This finding is consistent with the idea that the translation mechanism evolved from a preexisting RNA-based system (Noller *et al.*, 1989). For this reason, an important part of the understanding of the translation process, depends on the elucidation of the specific three-dimensional folding of the rRNA molecules.

The three-dimensional folding of the 16S and 23S ribosomal RNA molecules using the phylogenetic comparison analysis alone, has proven to be a difficult task (reviewed by Noller *et al.*, 1989). This is thought to be mainly due to the high degree of conservation of the bases involved in such interactions, and consequently the covariations required for their detection are difficult to find. In addition, each of these tertiary interactions probably involves one to three consecutive base pairs, thus comparably fewer than what is generally observed in the secondary structure. Moreover, these interactions are thought to involve different combinations of base pairs (canonical and non-canonical base pairings)(Gutell and Woese, 1990). In spite of these difficulties, comparison of the rRNA sequences has permitted the identification of some tertiary interactions in both 16S and 23S rRNAs (Gutell and Woese, 1990). Different three-dimensional models for the 16S *E. coli* rRNA in the 30S ribosomal subunit have been proposed. These models are based on the rRNA secondary structure deduced from comparative analysis, the positions of the ribosomal proteins obtained from neutron diffraction and chemical probing data on protein-rRNA interactions

(Brimacombe *et al.*, 1988; Stern *et al.*, 1988, 1989). To date, no such model has been constructed for the 23S rRNA molecule.

Antibiotics have played an important role in the study of ribosome structure and function as they were found to block the translation process at specific steps. Moreover, antibiotics have been found to protect (bind) bases in the highly conserved regions of the 16S and 23S rRNAs implicated in functions related to the known mode of action of the different antibiotics (reviewed by Noller *et al.*, 1989). For example, nucleotide mutations in the highly conserved 3' minor domain of 16S rRNA, which is thought to play an important role in translational initiation and decoding, have conferred resistance to various antibiotics. The role of individual nucleotides in conferring antibiotic resistance on the ribosome has been studied by chemical probing, which identifies sites of direct interaction between the antibiotic and the rRNA, and by site-directed mutagenesis followed by functional analysis of the mutant rRNA *in vivo* or *in vitro* (reviewed by Raué *et al.*, 1989).

Only one study on the intrinsic antibiotic resistance of various *Frankia* strains has been reported (Normand and Lalonde, 1986). Comparison of the primary and secondary structures of the two large rRNAs of *Frankia* to those of other eubacteria known to contain modifications conferring resistance to different antibiotics might give indications about *Frankia* AcN14a intrinsic antibiotic resistance.

## Research objectives

The *Alnus- Frankia* symbiotic association has potential value in the forestry and land reclamation and merits further investigation. Since the long term goal of research on *Frankia* symbiosis is likely to be the identification of the most efficient association, a better understanding of the symbiotic process, competitiveness between *Frankia* species and ecological events will be needed. For such studies, methods for the

preliminary grouping as well as for the identification of the different *Frankia* isolates (or strains) would be useful.

This project has as its primary objective the cloning and characterization the rDNA region of the *Frankia* strain AcN14a. Sequencing data for the ribosomal RNA genes and their comparison to other *Frankia* rRNA complete and partial sequences or to other related bacterial rRNA sequences already published will allow the identification of highly as well as less conserved regions. Such information could then be used for phylogenetic as well as ecological studies.

Depending on the region to be used genus, species and even strain (or isolate) specific probes could be designed. Probes specific to AcN14a could be used as markers during competition experiments. Similarly, oligonucleotides specific to the different *Frankia* strains (or isolates) could be designed from highly variable rRNA regions or from intergenic regions and used for their detection by PCR amplification.

In addition, comparative analysis of specific variable regions of the rDNA of different *Frankia* isolates or strains could be informative about their phylogenetic relationships and would permit their grouping. The sequence information required could be obtained by direct sequencing or after cloning of the PCR amplified material. The latter option has recently been successfully used (Nazaret *et al*, 1991).

Categorization would certainly facilitate the screening of the numerous *Frankia* isolates and strains available. Only one representative isolate per group would have to be tested. Such grouping of the isolates could be first obtained by RFLP analysis of the rDNA region. Thus the cloning of the rDNA region of one strain will provide DNA fragments that could be used as probes. One can assume that related isolates sharing similar hybridization patterns may also share similar competitiveness and symbiotic activity characteristics. A more elaborate analysis could even allow identification of the individual isolates or strains. In addition, such an

analysis could be informative on the number of rRNA operons present in these organisms.

The sequencing of the rDNA region from *Frankia* AcN14a will have additional value since it will contribute to the rRNA/rDNA sequence collection already available. To date, partial sequences of the 16S rRNA/rDNA of numerous strains and isolates have been determined (Hahn *et al.*, 1989a,b; Harry *et al.*, 1991; Nazaret *et al.*, 1991) but the sequence of an rRNA gene operon, thus the complete sequence of all three rRNA genes, their intergenic and flanking regions has been reported only once (Normand, unpublished). This will provide additional data for studies of the primary, secondary as well as tertiary structures.

Superimposition of *Frankia* AcN14a rRNA sequence derived from the gene sequence on the secondary structure model proposed for the prokaryotic rRNAs will provide an additional sequence to test the proposed model. If done with sequences of closely related bacterial species, such comparisons could elucidate the secondary structure of variable regions as well as regions characteristic of a given group of bacteria (in this case, actinomycetes). One can also test the validity of the secondary structures proposed for the sequences flanking the mature rRNAs in the long precursor RNA molecule and locate putative processing sites. In addition, *Frankia* AcN14a rRNA sequences might be useful in the identification of possible intrinsic antibiotic resistance and for the verification of the proposed tertiary interactions.

# CHAPTER 2: MATERIALS AND METHODS

## Bacterial strains, plasmids, and phage

### Growth and maintenance of *Frankia*

*Frankia sp.* strain AcN14a was obtained from M. Lalonde, Université Laval, Québec. QmodB culture medium (Normand *et al.*, 1983) was used for growth of *Frankia*. The medium was prepared, autoclaved and incubated at 28°C for at least a week to ensure sterility as this actinomycete grows slowly. From an original inoculum, 20-30 ml of cell suspension were transfered into one liter of fresh medium. The cultures were left standing with occasional swirling to provide aeration until sufficient cell density was obtained for DNA or RNA extractions. Similarly, a collection of *Frankia* isolates obtained from D. Lachance (Petawawa National Forestry Institute) were cultured in 50 mL aliquots of QmodB medium. These cultures were grown at 28°C for 10 months before DNA extraction.

### *E. coli* strains and vectors

*Escherichia coli* strains DH5α (BRL) and JM101 (Messing *et al.*, 1981) were used as hosts for the growth and purification of plasmids, the latter was also used as host for M13 vectors. *E. coli* strain LE392 ($F^-$, *hsd* R514 ($r_K^-$,$m_K^-$), *sup* E44, *sup* F58, *lac* Y1 or Δ(*lac* IZY)6, *gal* K2, *gal* T22, *met* B1,*trp* R55, $\lambda^-$) (Promega), a permissive host strain (Maniatis *et al.*, 1982) was used as host for EMBL3 recombinants. Plasmid cloning vectors included pGEM4Z (Promega) and pTZ18R (Pharmacia). M13mp18 and M13mp19 (Yannish-Perron *et al.*, 1985) phages were used as sequencing vectors.

## Preparation, isolation and purification of DNA

### Genomic DNA from *Frankia*

DNA was isolated from *Frankia* cultures by a modification of a procedure designed for *Streptomyces* (Hunter, 1985). Basically 0.5-1.5 g of

cells resuspended in 10 mL of TE buffer* with lysozyme at 1mg/mL were incubated for 30 minutes at 37°C. 0.1 volume of 20% SDS was added and the solution was incubated for 15 minutes at 65°C. Meanwhile, disruption of the cell clumps was achieved by slowly pipetting the solution up and down in a 10 mL glass pipette. Following the addition of NaCl to 0.6 M, the sample was extracted with 10 mL of Tris-saturated phenol and two times with chloroform. The DNA was precipitated by the addition of one volume of isopropanol, usually DNA could usually be spooled out after an incubation of 10 minutes at room temperature or was pelleted by centrifugation. The pellet was resuspended in 1 mL of TE buffer* and extracted with phenol and two times with chloroform in order to further purify the DNA. The DNA was precipitated by adding 0.5 volume of 7.5 M ammonium acetate and two volumes of 95% ethanol.

**Plasmids and replicative form DNA of M13 phage**

Small scale preparations of plasmid DNA and M13 replicative form DNA were prepared by a modified alkaline lysis method of Birnboim and Doly, 1979 (Morelle, 1989). Large scale preparations were also obtained following the same procedure simply by scaling up the quantities of the different reagents. The resulting DNA was used for restriction analyses and sequencing.

**Single-stranded DNA from M13**

Single-stranded templates for DNA sequencing were prepared and purified, according to the M13 cloning and sequencing handbook from Amersham. The phage DNA was then used as single-stranded template for the preparation of deletions and for sequencing.

To estimate the size of the inserts, 20 µL of supernatants obtained after centrifugation of the cell cultures were incubated for 10 minutes at 65°C with 1 µL 2% SDS, 3 µL loading buffer (50% glycerol, 0.2M EDTA pH 8.0, bromophenol blue) and run on a 0.5X TBE gel. The single-stranded M13 vector with no insert and a clone containing an insert of known size were run as markers. Similarly, the orientation of the inserts was

determined by a complementation tests and a more precise insert size was obtained by digestion of the complementation test samples with S1 nuclease (Baird and Johnson, 1991).

**Lambda DNA**

Lambda phage was grown, titred and purified as described in the Promega handbook. Briefly the small scale lysate was obtained by resuspending the putative positive plaque in 5 mL of L Broth* with 10 mM magnesium sulfate in a 15 mL glass tube. After the addition of 100 µL of an overnight culture of LE392, the culture were shaken vigorously at 37°C for 4-6 hours. When lysis of the culture was observed, the cell debris was pelleted by centrifugation (15 minutes, 8000 rpm, 4°C) and the supernatant was collected in a sterile tube. Titres were low, approximately 5 X $10^5$ pfu/mL. DNA extraction of small scale lysate was done according to a standard protocol (Lévesque, 1991).

**Specific restriction fragments**

DNA digestions were electrophoresed on 0.8-1.0% agarose gels, with lambda DNA digested with HindIII and pBr322 DNA digested with HaeIII as molecular weight standards. Restriction fragments to be ligated into a vector were excised from stained gels and the DNA fragment was isolated by using a Geneclean kit as recommended by the manufacturer (Bio 101). Any DNA restriction fragment to be used as a probe was electrophoresed through a second gel before isolation of the DNA fragment.

**PCR amplification of DNA fragments**

Double-stranded amplification was performed on part of the 16S rRNA genes following the procedure described by Innis and Gelfand (1990). Four 17-base oligonucleotides (F16A1 and F16A2, F16C1 and F16C2) were used. The nucleotide sequences of these primers are presented in Table 2. The PCR reaction was carried out in a final volume of 50 µL containing 1 µL of template DNA, 200 µM of each deoxynucleoside triphosphate, 0.02 µM of each primer, and 2.5 U of Taq I DNA polymerase (Bio/Can) in PCR buffer. A control reaction without any DNA template

was always included. The amplification reactions were carried for a total of 37 cycles in a DNA thermo cycler (Perkin Elmer Cetus). The DNA was denatured at 94°C for 1 minute, annealed at 57°C for 1 minute, and the extension was done at 72°C for 2 minutes. The denaturation step of the first cycle was carried out for 5 minutes and the extension step of the last two cycles for 2 minutes. The samples were then stored at 4°C. The amplification products were analyzed by electrophoresis on a 2.5% agarose gel. The DNA products were extracted sequentially with phenol, phenol-chloroform and chloroform. Precipitation was done by the addition of ammonium acetate and ethanol. DNA was resuspended in 25 µL of 0.1X TE buffer and 5 µL was used for their cloning in M13mp18 and M13mp19 vectors.

**Vectors**

Plasmids and M13 phage replicative form were digested with the appropriate restriction enzymes as per the manufacturer's recommendation, and treated with calf intestinal alkaline phosphatase (Boeringer Mannheim). The reaction was then heated at 65°C for 10 minutes and electrophoresed on a 1.0% agarose gel. The band corresponding to the linear form of the DNA molecule was excised and purified with Geneclean.

**Isolation of RNA**

Total RNA from *Frankia* AcN14a culture was isolated following the hot phenol method. Cultures were allowed to settle and most of the supernatant decanted. Cells were harvested in a 40 mL polypropylene centrifuge tube by centrifugation at 8000 rpm for 12 minutes at 4°C in an HB4 rotor. Cells were resuspended in 0.1 volume of ice cold lysis buffer (0.02M sodium acetate, 0.5% SDS, 0.001M EDTA, pH 5.5) and extracted two times at 65°C with 1 volume of phenol saturated with lysis buffer. For each extraction, the mixture was shaken for 10 minutes and the phases were separated by centrifugation at 3500 rpm at room temperature for 10 minutes. The aqueous phase was transfered to a new sterile centrifuge tube, potassium chloride was added to 0.1 M potassium chloride and the RNA was precipitated by the addition of 2 volumes of 95%

ethanol. The solution was kept at -20°C overnight. The precipitate was collected by centrifugation for 30 minutes, 8000 rpm at 4°C, washed once with 80% ethanol and dried at room temperature for 30 minutes or until almost dry. The precipitate was then resuspended in sterile water and the preparation was analyzed by formaldehyde gel electrophoresis.

**Construction and screening of *Frankia* genomic library**

**Fractionation of *Frankia* DNA restricted with MboI**

*Frankia* DNA (50 µg) was partially digested with MboI (half with 0.25U/µg and the other half with 0.12U/µg) for one hour at 37°C and the enzyme was heat inactivated by addition of EDTA to 10mM. The fragments were separated by electrophoresis in a 0.75% LMT (Low Melting Temperature) agarose gel to allow for separation of the large fragments. Lambda DNA digested with HindIII or BglII were used as molecular weight markers. The gels were run overnight at 4°C in TAE buffer*. Fragments between 10-13 kb were identified after staining the gel with ethidium bromide and visualization under UV light. The fragments were isolated from the agarose gel with Geneclean. To prevent shearing of the long DNA fragments, the isolation was performed without vortexing.

**Ligation of MboI fragments to EMBL3 BamH1 arms, *in vitro* packaging and titration**

The insert DNA (approximately 10 µg) and the lambda arms (0.5 µg) were then coprecipitated by the addition of 0.5 sample volume of 7.5 M ammonium acetate and 2 volumes of 95% ethanol. The DNA mixture was resuspended in 4.4 µL of TE buffer*. Ligation buffer* and T4 DNA ligase were added to 4 µl of the DNA mixture and, following ligation overnight at 4°C, the mixture was packaged *in vitro* using a commercial packaging system (Promega) according to the manufacturer's protocol. After titration of the library, different amounts were plated on 310 $cm^2$ Petri dishes in order to obtain 0.5, 1.0, and 1,6 X $10^4$ plaques/plate. The library was plated using *E.coli* LE392 strain as host. The plates were incubated at 37°C overnight. At this time plaques did not come into contact with one

another. Plates were stored at 4°C. The plate with approximately 4 X $10^4$ plaques was used for screening.

To amplify the library, all the plates were overlaid with approximately 6 mL/ 4 X $10^4$ plaques of SM buffer* and shaken for one hour at room temperature. The top layer of each plates were then scraped off, pooled in a 250 mL Erlenmeyer with an additional 50 mL of SM buffer* and mixed at room temperature with a magnetic stirrer. After 15 minutes chloroform was added to 5%. The mixture was stirred for 15 minutes, then was transfered to two sterile 30 mL Corex centrifuging tubes. After centrifugation (10 minutes, 8000 rpm), the amplified library was transfered to a 250 mL bottle and stored at 4°C over 0.3% chloroform.

**Screening of genomic library**

The phages were transfered onto a Biotrans membrane and hybridized as described in the Southern transfer and hybridization sections. The membrane was probed with a DNA fragment containing 23S rDNA sequence. Following hybridization, the membrane was washed, before overnight exposure.

**Plaque purification**

Positive plaques (regions of ~1 $mm^2$) were picked from the plate, stored in 3 mL SM buffer* with few drops of chloroform and dilutions were titred. The plates with sufficient but manageable number of plaques (50-200) were lifted onto Biotrans membranes and hybridized as previously described. Additional rounds of purification were made until all the plaques on the plate hybridized to the probe.

**Subcloning into plasmids and M13 vectors**

**Ligations into double-stranded vectors**

Restriction fragments of interest were subcloned into one or more of the following vectors: M13mp18/mp19 RF DNA, pGEM4Z, pTZ18R.

Ligation mixtures containing vector and DNA fragment to a concentration of 2-5 ng/μL were incubated overnight at 4°C. The T4 DNA ligase was inactivated by heating the reaction mixture at 65°C for 15 minutes. A control of vector without any DNA fragment was also prepared to verify the dephosphorylation of the vector.

### Transformations

10-20 ng of each ligation was introduced into competent *E. coli* JM101 cells by transformation (Hanahan, 1985). The transformation efficiency was tested with 10 pg of uncut M13mp19 and was usually over $10^7$ plaques/μg.

## Labeling of probes

### Preparation of radioactively labeled fragments

Isolated restriction fragments were labeled by the random primer method (Feinberg and Vogelstein, 1983) using kits purchased from Amersham and NEN. The radioactive label was [α-$^{32}$P]-dCTP (3000 Ci/mmol). Unincorporated nucleotides were removed by passing the labeled fragments through Sephadex G-50 by the spun-column procedure as indicated in Maniatis *et al.* (1982).

## Agarose gel electrophoresis

### Gel electrophoresis of DNA

DNA restriction fragments were separated by electrophoresis through horizontal agarose slab gels. The concentration of agarose in the gels varied from 0.8 to 1.2% depending on the size range of the fragments to be resolved. The electrophoresis buffer contained 40 mM Tris-acetate, 2mM EDTA pH 8.0 (Maniatis *et al.*, 1982). HindIII-digested lambda or Taq-digested pBR322 DNAs were generally used as molecular size markers.

M13mp19 phage supernatants were electrophoresed on a 0.8-1.0% agarose TBE gel to verify approximate insert size or orientation and hybridization to the proper probes.

**Gel electrophoresis of RNA**

RNA was separated by electrophoresis through 1.5% agarose gels containing formaldehyde (Lehrach *et al.*, 1977). RNA samples were prepared as described in Maniatis *et al.* (1982). The gels contained 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA, 2.2 M formaldehyde. The electrophoresis buffer contained 20 mM MOPS (pH 7.0), 5 mM sodium acetate, 1 mM EDTA and 1.1 M formaldehyde. Following electrophoresis, the gel was washed with 4 to 5 changes of water for two hours and two changes of 0.1 M ammonium acetate for one hour. The gel was stained with 0.5 ug/mL of ethidium bromide in 0.1 M ammonium acetate for 45 minutes. The gel was destained for 2 hours with 2 to 3 changes of 0.1 M ammonium acetate and then photographed. *E. coli* 23S, 16S and 5S rRNAs or *E. coli* total RNA were used as size standards.

**Hybridization to DNA transfers**

**DNA transfer to membrane**

Briefly, the DNA (in an agarose gel, or dotted onto a Biotrans membrane) was denatured in 1.5 M sodium chloride / 0.5 M sodium hydroxide and neutralized in 3 M sodium acetate as described in the ICN protocol handbook. When DNA was dotted onto a membrane (for control dots or plaque lifts), the membrane was placed DNA side up on blotting paper soaked in denaturation and neutralizing solutions. The membranes were then left to air dry for 5 minutes. For agarose gels, DNA was allowed to transfer with 20X SSC* as the transfer buffer (Southern, 1975). Following transfer or denaturation, the DNA was crosslinked to the membrane by UV illumination (Black Ray lamp, Ultraviolet Products, Inc., San Gabriel, CA) for 5 minutes.

For genomic transfers, about 0.5-1.0 µg of genomic DNA per sample were restricted with various restriction enzymes according to the manufacturer's instructions and electrophoresed on a 0.8-1.0% agarose gel in TAE buffer* (Maniatis *et al.*, 1982). After staining with ethidium bromide and visualization by UV light, the gel was pretreated with 0.2 N

hydrochloric acid for 5 minutes and processed as described in previous section.

Dot blots controls were added to some hybridizations in order to determine the sensitivity of the probe after the appropriate exposure. Various dilutions of plasmids containing the control DNAs (100 - 0.001 ng/µl in 10X SSC) were dotted on a Biotrans membrane, then air dried and prepared as Southern transfers.

**Hybridization**

The following hybridization protocol is outlined in the ICN Biotrans membranes handbook (ICN Biomedicals, Inc.). All the membranes were prehybridized with 4 mL/100 $cm^2$ membrane, and hybridized with 2 mL/100 $cm^2$ membrane of hybridization solution (5X Denhardt's solution*, 5X SSC, 50 mM sodium phosphate, pH 6.5, 0.1% SDS, 250 µg/mL single-stranded Herring sperm DNA and 50% (v/v) deionized formamide). Prehybridization was done for a minimum of one hour and the hybridization overnight, both at 42°C with gentle shaking. Blots were then dipped in 2X SSC, O.1% SDS, then washed four times in the same buffer with vigorous agitation for 5 minutes at room temperature and finally washed in 0.1X SSC, 0.1% SDS at 250 mL/100 $cm^2$ twice for 15 minutes at 50°C. For some Southern hybridization experiments, the last washes were done at room temperature, 37°C, 42°C, 50°C or 65°C. The membranes were then exposed to Kodak XAR-5 film, with Quanta III intensifying screens.

**Rehybridization**

The probe was removed by washing the membranes in 50% deionized formamide, 10mM sodium phosphate pH 7, solution at 100 mL/100 $cm^2$, followed by 15 minutes at room temperature in 2X SSC, 0.1% SDS at 250 mL/100 $cm^2$. Care was taken not to let the membranes dry between rounds of hybridization, washing, and stripping. Stripped membranes were exposed on Kodak XAR-5 film for at least 24 hours to ensure that the probe had been removed properly.

**DNA sequencing**

Restriction fragments for sequencing were subcloned into either M13 mp18/mp19 vectors (Yannish-Perron et al., 1985) or pGEM4Z (Promega) plasmid vectors. Fragments for subcloning were determined by restriction analysis of a genomic clone from *Frankia* AcN14a: λFR-8. In some cases, restriction of the clone did not yield useful restriction sites for subcloning, in others, the restriction fragments were larger than 0.5 kb. In these circumstances, the sequence was determined either by the used of synthesized 17-nucleotide oligomers or by the preparation of deletion clones.

DNA was sequenced using the dideoxy chain termination method (Sanger *et al.*, 1977) with [α-$^{35}$S]-dATP as the radioactive label. Sequencing of M13 recombinants phage DNA and double-stranded plasmids was as described in the different kits that have been used (Pharmacia, Promega and USB). Separation of the strands was done by electrophoresis through 7% polyacrylamide gels* and visualized by autoradiography. The synthesized primers and their respective composition are listed in Table 2. In some case, manganese reaction buffer was used to determine the sequence close to the 3' end of the primer. In circumstances where there appeared to be some difficulty in determining sequence due to possible secondary structure, dITP was used instead of dGTP.

**Buffers and other solutions**

Buffers and solutions indicated here are from Maniatis *et al.*, 1982 unless otherwise stated.

A. 7% acrylamide gel for sequencing: 33.6 g urea, 37 mL water, 14 mL 40% bis/acrylamide (20:1), 4 mL 10X TBE, 56 µL TEMED, 560 µL ammonium persulfate (100 mg/mL).

B. 50% Denhardt's solution: 1% (w/v) Ficoll (400 000 MW, 1% (w/v) polyvinylpyrolidone (360 000 MW), 1% (w/v) bovine serum albumin, (Denhardt, 1966)

C. L broth: (per liter) 10 g tryptone, 5 g yeast extract, 10 g NaCl, 0.72 g Tris base, 10 mL 1M $MgSO_4$, 5 mL 0.5 M $CaCl_2$, pH 7.25, 9 g agar.

D. 10X ligase buffer: 400 mM Tris-HCl, pH 7.5, 100 mM $MgCl_2$ , 100 mM DTT, 10 mM ATP, 500 mg/mL bovine serum albumin (Promega).

E. SM buffer: (per liter) 5.8 g NaCl, 2 g $MgSO_4$ .$7H_2O$, 50 mL 1M Tris-HCl, pH 7.5, 5 mL 2% gelatin.

F. 20X SSC: 3 M NaCl, 0.3 M sodium citrate, pH 7.0.

G. 10X TAE: (per liter) 96.8 g Tris base, 32.8 g sodium acetate, 7.4 g disodium EDTA, pH 8.2 with the addition of glacial acetic acid.

H. 10X TBE: (per liter) 108 g Tris base, 55 g boric acid, 8.5 g disodium EDTA, pH 8.3.

I. TE buffer: 10mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0.

**Nucleotide sequence analyses**

**Primary and secondary structure analyses**

The sequence data generated using the different primers were aligned with the aid of the the PCGENE programme (Intelligenetics, Inc). Analysis and comparison of *Frankia* AcN14a to other eubacterial sequences were done using IBI Pustell Sequence Analysis Package version 2.04 (IBI, New Haven, Connecticut). Secondary structures were designed with the help of the RNAFOLD programme of the PCGENE package. A programme called DNASIS of an updated version of the IBI Pustel package was used to search for tRNA sequences.

**Phenetic and phylogenetic analyses**

DNA sequences were aligned using CLUSTAL V (Higgins, D.G. *et al.*, 1991, submitted to CABIOS) modified to run on the VAX/VMS (Version 5). Phenetic analysis was conducted using Jukes-Cantor's model (Jukes and Cantor, 1969) that assumes the independent changes at all sites with equal probability. The matrix of distances was also subjected to a cluster analysis using the neighbor-joining method (Saitou and Nei, 1987). These methods do pair-wise comparisons of the sequences, calculate the sequence similarity allowing the construction of the tree that fits the distances calculated. Some preliminary data are presented in Appendixes 2 and 3.

**TABLE 2 LIST OF SYNTHESIZED SEQUENCING AND PCR PRIMERS**

| Primer | Sequence (5' - 3') | Position in rRNA gene operon | |
|---|---|---|---|
| | | *Frankia* AcN14a [1] | *E. coli* [2] |
| F1601 | TTACAACCCGAAGGCCGT | 1014-997 | 1014-947 |
| F1602 | GCTCCTTAGAAAGGAGG | 2116-2000 | 3068-3052 |
| F2301 | GTGTAGGGACTTGCTTC | 4449-4465 | n.a. |
| F2302 | GACTCGGCGAAATTGCA | 4738-4754 | 5503-5519 |
| F2303 | TCAGCCTGGTTGGCAAT | 5029-5045 | 5781-5797 |
| F2304 | GCGAGCTGGGTTTAGAA | 5307-5323 | 6073-6089 |
| F2305 | TAATAGGCCGAGGGCTT | 5611-5627 | 6380-6396 |
| F2306 | GACTGTGTGGGAGAGTA | 5795-5811 | 6583-6599 |
| F2307 | GAAGCAAGTCCCTACAC | 4465-4449 | n.a. |
| F2308 | TCCCCTACCCATACACA | 3836-3820 | 4693-4677 |
| F2309 | TGTGTATGGGTAGGGGA | 3820-3836 | 4677-4693 |
| F2310 | GAGTAGTGGTGAGCGGG | 2772-2788 | 3739-3755 |
| F2311 | GTGATGCTGGTTCTCC | 3418-3434 | 4298-4314 |
| F16A1 | GCGGCGTGCTTAACACA | 637-653 | 1556-1572 |
| F16A2 | ATTCCCCACTGCTGCCT | 948-932 | 1882-1866 |
| F16C1 | GGGGCATGATGACTTGA | 1773-1757 | 2711-2727 |
| F16C2 | TGGGGAGTACGGCCGCA | 1449-1465 | 2401-2417 |

1 Positions in the *Frankia* AcN14a rRNA gene operon presented in Figure 10 of this thesis.

2 Positions in the *E. coli* sequence published by Brosius *et al.*, 1982.

n.a. indicates that the corresponding region is not present in the rRNA gene operon of *E. coli*.

# CHAPTER 3: RESULTS

## 1A. CLONING AND CHARACTERIZATION OF *FRANKIA* ACN14A rDNA

### Analysis of *Frankia* AcN14a genomic DNA

Like most actinomycetes, members of the genus *Frankia* have a genome with a high proportion of guanosine and cytosine residues (G+C content). Their G+C contents have been reported to range between 66% and 75% (An *et al.*, 1983; Fernandez *et al.*, 1989).

For DNA with such a high G+C content, restriction endonucleases that have a high % G+C within their recognition site are expected to cut more frequently than enzymes that have a low % G+C within their recognition site. Indeed, this correlation is observed when *Frankia* AcN14a DNA is digested with selected restriction endonucleases (data not shown). Restriction endonucleases such as BglII, EcoR1 and HindIII which recognize 6 base pair sequences with a relatively low G+C content, cut *Frankia* DNA infrequently and, consequently the average DNA fragment size is high. In contrast, digestion with restriction endonucleases such as SmaI and SstII which recognize 6 base pair sequences containing exclusively guanosine and cytosine residues generated, on average, fragments of lower molecular sizes. The frequency of cutting also depends on the length of the site recognized by the restriction endonuclease. Consequently, even smaller DNA fragments were generated after digestion of *Frankia* AcN14a DNA with the endonuclease HaeIII which recognizes the 4 base pair sequence GGCC (data not shown).

### Characterization of the 23S rDNA probe

The probe used for the preliminary experiments and for the screening of the *Frankia* genomic library was derived from pnod4 (Gleeson, 1992) which consists of a 1,718 bp fragment cloned into pGEM4Z. The 1.5 kb fragment used as a probe results from the digestion of pnod4 with BamH1 and EcoR1. This probe comprises the 3' portion of *Frankia* AcN14a 23S ribosomal RNA gene

(Figure 6A), and shares 68% and 77% homology with the corresponding sequence of *E. coli* and *B. subtilis* respectively.

Since the ribosomal RNA gene sequences are known to be highly conserved throughout evolution, the specificity of the fragment to be used as probe was evaluated. In a Southern hybridization experiment, replicates of *E. coli* and *Frankia* AcN14a genomic DNAs were digested with BamH1 hybridized to the 1.5 kb fragment and subjected to wash conditions of increasing stringency (room temperature (RT), 37°C, 42°C, 50°C or 65°C). After an exposure of 12 hours, no hybridization signals were observed for *E.coli* DNA even for the membrane washed at RT (data not shown). In all cases, two strong bands (1.9 and 2.8 kb) were detected in *Frankia* AcN14a DNA suggesting the presence of two copies of the 23S rRNA gene. Bands in the *E. coli* lanes were observed after a longer exposure of 48 hours for all membranes except that washed at 65°C. The bands had sizes larger than those detected in *Frankia* AcN14a. Thus, the 1.5 kb fragment does indeed hybridize specifically to *Frankia* AcN14a 23S rRNA gene sequences and could be used to screened a *Frankia* AcN14a genomic library. We will refer to this fragment as the 23S rDNA specific probe.

**Southern analysis of *Frankia* AcN14a rDNA**

The number of rRNA operons in *Frankia* AcN14a was estimated by Southern hybridization. Figure 3 represents the autoradiogram of *Frankia* genomic DNA digested with various restriction endonucleases and hybridized to the 23S rDNA specific probe. A list of the sizes of the fragments hybridizing to the 23S rRNA gene specific probe as well as to other probes is presented in Appendix 1. The position of the 23S rDNA specific fragment and those of other probes derived from a genomic clone are given in Figure 6A. In most cases, two or more bands hybridized to the 23S rDNA specific probe, suggesting the presence of two copies of the 23S rRNA gene in this *Frankia* strain.

Hybridization of the same membrane to probes derived from other regions of a *Frankia* AcN14a genomic clone containing a 23S rRNA gene further supports the presence of two copies of the 23S rRNA gene. The regions flanking the 23S rRNA gene are probably also present in two copies in the

genome since in most cases, at least two bands are detected by hybridization. Characterization of the different DNA fragments used as probes revealed the presence of the different rRNA genes (see Figure 6A). In addition, the fact that probes containing different rRNA genes hybridized to a doublet corresponding to two 5.1 kb fragment of the PstI digest of genomic DNA, suggests clustering of the 16S, 23S and 5S rRNA genes. As in most eubacterial genomes, the rRNA genes are probably organized into operons. The two rRNA gene clusters are probably not in close proximity since two large fragments (15.0 and 21.0 kb) of the BglII digest hybridized to all probes.

Polymorphism in the *Frankia* AcN14a rDNA could be detected with most of the restriction endonucleases. Digestion with enzymes, such as BamH1, BstEII as well as BglII revealed a simple band pattern and for this reason these enzymes could be useful in order to differentiate between the two rRNA operons when they have been cloned (Figure 3).

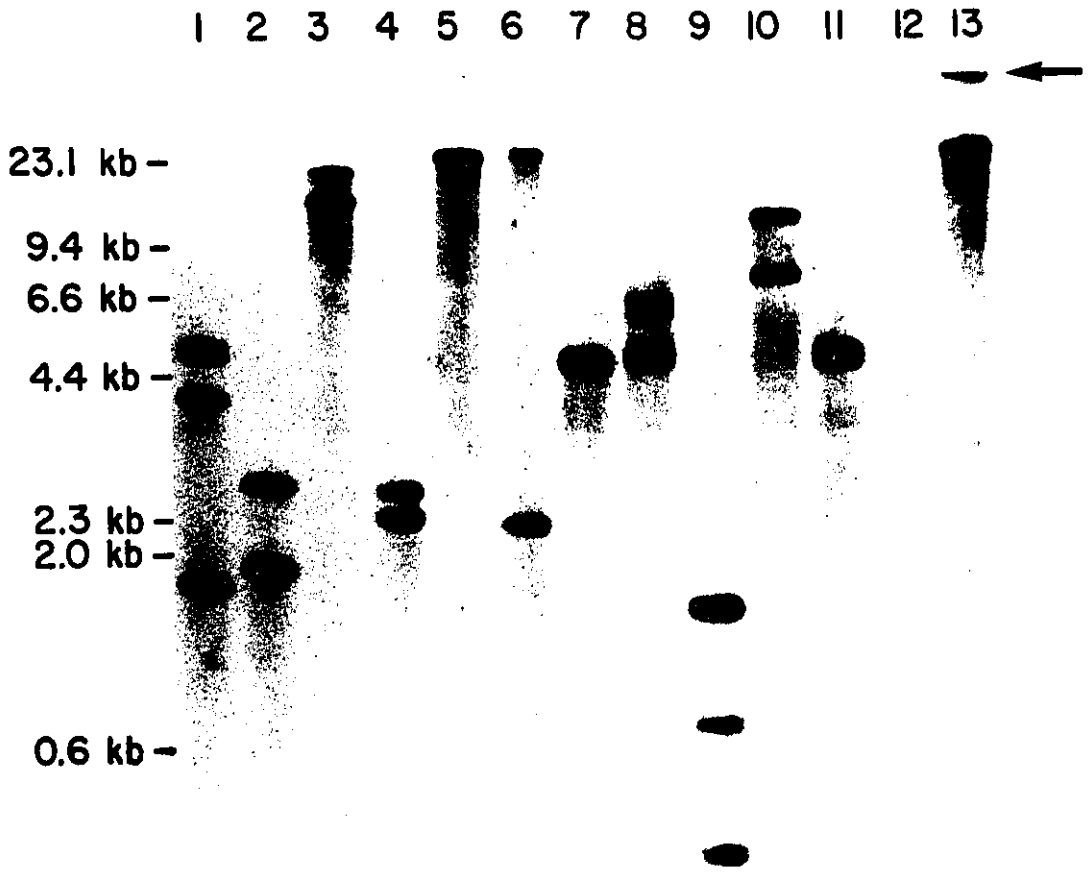
Two copies of the rRNA genes seem to be the rule in members of the genus *Frankia*. The *Frankia* strain ArI3 and thirty-eight other *Frankia* isolates have been investigated and all except one were found to contain two copies of the rRNA genes. Three copies are present in the latter strain. These results will be further discussed in Part 2 of the Results Section. As could be expected from its slow growth rate, only two copies of each rRNA gene are present in *Frankia* AcN14a. They are probably arranged into operons since this organization is typical of most prokaryotes.

The number as well as the organization the rRNA genes appear to be conserved in members of the genus *Frankia*. Of the various strains examined only one was found to contain an additional operon. Similarly, variation in the copy number of the rRNA gene cluster has been reported among the actinomycetes. *M. phlei* and *M. segmatis* possess two sets of rRNA genes whereas only one rRNA set is thought to be present in *M. bovis* BCG and *M. lepraerium* (Yamada *et al.*, 1986; Suzuki *et al.*, 1987). *S. coelicolor* (Baylis and Bibb, 1988a), *S. griseus* (Kim *et al.*, 1991), *S. lividans* (Suzuki *et al.*, 1988b), and

**FIGURE 3 SOUTHERN BLOT ANALYSIS OF *FRANKIA* ACN14A GENOMIC DNA**

Southern hybridization of *Frankia* AcN14a genomic digests to the 23S rDNA specific probe (pnod4), corresponding to the 3' portion of *Frankia* AcN14a 23S rRNA gene.

**Lane 1** digest with Acc1 (GT $^{A}/_{C}$ $^{T}/_{G}$ AG); **lane 2** digest with BamH1 (GGATCC); **lane 3** digest with BglII (AGATCC); **lane 4** digest with BstEII (GGTNACC); **lane 5** digest with EcoR1 (GAATTC); **lane 6** digest with HindIII (AAGCTT); **lane 7** digest with Pst1 (CTGCAG); **lane 8** digest with Sal1 (GTCGAC); **lane 9** digest with Sma1 (CCCGGG); **lane 10** digest with SstI (GAGCTC); **lane 11** digest with SstII (CCGCGG); **lane 12** HindIII-digested λ DNA and Taq-digested pBR322 DNA as molecular weight markers; **lane 13** uncut DNA. The guanosine and cytosine residues within the recognition sites are in bold.


1 2 3 4 5 6 7 8 9 10 11 12 13
23.1 kb
9.4 kb
6.6 kb
4.4 kb
2.3 kb
2.0 kb
0.6 kb

S. *rimosus* (Plohl and Gamulin, 1991) contain six rRNA operons whereas *S. ambofaciens* contains only four (Pernodet.*et al.*, 1989). Thus the number of rRNA operons has been reported to vary greatly among actinomycete species of different genera as well as for species of the same genus.

Nucleotide methylation has been reported to occur in eubacterial DNA (Raleigh, 1987). In order to investigate methylation of the rDNA region of *Frankia* AcN14a, Southern hybridization of genomic DNA digested with three pairs of isoschizomers exhibiting different sensitivity to methylation (BamH1/Bst1, HpaII/Msp1 and Mbo1/Sau3A) was done. A mixture of two DNA fragments (1.9 kb and 8.0 kb BamH1 fragments) derived from one *Frankia* AcN14a genomic clone, comprising one entire rRNA operon and approximately 4 kb of the sequence upstream from the 16S rRNA gene, was used as a probe (see Figure 6A). The first pair of restriction endonucleases allows detection of 5-methylcytosine residues in the sequence GGAT$^{m5}$CC since methylation of this cytosine residue inhibits restriction by BamH1 but not by Bst1. The second pair of restriction endonucleases detects methylation of the middle cytosine residue of the recognition site CCGG (C$^{m5}$CGG) since cutting by HpaII but not by Msp1 is inhibited by methylation of this residue. The last pair of enzymes allows detection of 5-methylcytosine residues as well as the 6-methyladenine residues in the sequences GAT$^{m5}$C and G$^{m6}$ATC. Mbo1 cuts the sequence GATC even if the cytosine residue is methylated whereas methylation of the adenosine residue inhibits it. Sau3A will cut if the adenosine residue is methylated but will not cut if the cytosine residue is methylated. As judged by the absence of discrepancies between the hybridization band patterns of the various genomic digests (data not shown), we can conclude that these two types of methylated residues are probably not present in rRNA operons or in the region upstream from them. It therefore seems likely that extensive methylation of these residues is not found in *Frankia* AcN14a.

**Construction of *Frankia* AcN14a genomic library**

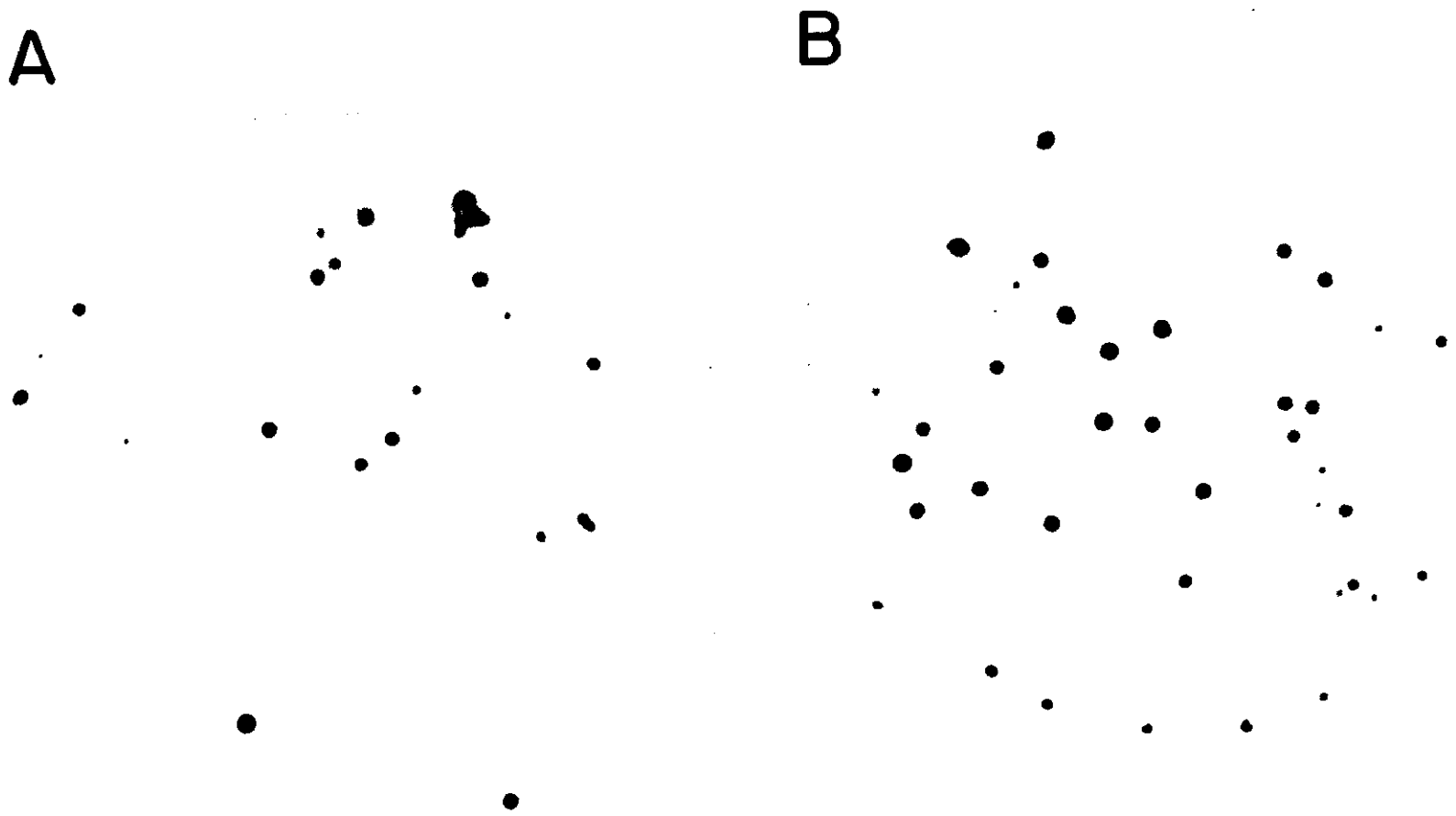
The minimum number of clones required to construct a representative library was estimated with the assumption that the genome size of *Frankia*

AcN14a was close to that of *E. coli* (4.3 x $10^6$ bp). According to the calculations described in Maniatis *et al.*, 1982, 1.6 x $10^2$ clones of 12 kb in length are necessary in order to have a 99.9% probability of obtaining a single copy gene. Obviously, this number would vary depending on an accurate genome size and the gene copy number. The mean genome size of two *Frankia* strains, EuI1A and ArI4, has been estimated to be 7.1 x $10^9$ Da, which corresponds to 1.1 x $10^7$ bp (An *et al.*,1985), thus representing a genome size about 2.5X that of *E. coli* and similar to that of *Streptomyces* (Hunter, 1985). In addition, two rRNA operons are present in *Frankia* AcN14a. Therefore a representative library would need to contain at least 2.1 x $10^3$ clones.

Randomly fragmented DNA is necessary to obtain a representative library with the minimum number of clones. For this purpose, restriction endonucleases such as Mbo1 and Sau3A are often used (Kaiser and Murray, 1985). These enzymes are isoschizomers, recognize the sequence, GATC, and are expected to restrict, on average, every 256 ($4^4$) base pairs in a random DNA sequence of 50% G+C content. Mbo1 will cut if the cytosine residue is methylated but will not cut if the adenosine residue is methylated. Even though no such methylation has been detected in *Frankia* AcN14a (see previous section), the genomic library was constructed using partial digestion with Mbo1. The appropriate enzyme/DNA ratio was determined by digesting one µg aliquots of *Frankia* AcN14a DNA with dilutions of Mbo1 and the ratio was determined empirically to be 0.25 U/ µg of Mbo1. As suggested by Kaiser and Murray (1985), the digest was scaled up, half was digested with 0.25 U/ µg and the other with 0.125 U/ µg of Mbo1. After combining the two digestion reactions, the fragments generated were electrophoretically separated in a 0.75% LMT (low melting temperature) agarose gel within a standard agarose frame to provide support. In addition, the gel was run at 4°C to prevent collapse of the gel. After visualization of the ethidium bromide stained DNA under the UV light, the region comprising fragments in the 10-13 kb range was excised and the DNA was isolated.

**FIGURE 4 SCREENING OF A *FRANKIA* AcN14A GENOMIC LIBRARY**

**A**. Autoradiogram of the membrane representing approximately 4 x $10^4$ clones from the *Frankia* AcN14a library after hybridization to the 23S rDNA specific probe (pnod4), corresponding to the 3' portion of *Frankia* AcN14a 23S rRNA gene. **B**. Autoradiograms of membranes for the second round of purification.


A
B

The *Frankia* genomic library was constructed using the EMBL3 vector BamH1 arms system (Promega). DNA partially digested with BamH1, Mbo1 and Sau3A can be ligated into the dephosphorylated arms. Religation of these arms with the central stuffer fragments has been prevented by secondary digestion with EcoR1 followed by the removal of the central fragment. The background of non-recombinants is therefore dramatically decreased and the number of plaques to be screened is lowered. DNA fragments ranging in size from 9-23 kb can be cloned in this vector. Because of the size of the DNA fragments isolated (10-13 kb), ligation of two fragments that are not contiguous cannot be completely eliminated.

The *Frankia* AcN14a genomic library was found to contain 4.7 x $10^4$ recombinants which represents 11 *Frankia* genome equivalents, assuming that the average insert size is 12 kb. The genomic library was amplified and part of the amplified library, representing 4 x $10^4$ recombinants, was screened for the presence of clones containing 23S rDNA sequences. Figure 4 represents autoradiograms of the initial screening and further plaque purifications using the same DNA fragment as probe. Twenty three of the 4 x $10^4$ plaques screened (0.06%) were positive. Eight plaques hybridizing to the probe and a control plaque that did not were purified. Six of the eight plaques that initially gave a positive signal still hybridized strongly to the probe after two rounds of purification.

**Characterization of the genomic clones**

DNA was prepared from small scale lysates and analyzed by Southern hybridization. The six λ DNA clones were digested with BamH1 or BstEII and were categorized into groups according to the sizes of the fragments that hybridized to the 23S rDNA specific probe (Table 3). Three distinct band patterns were observed, of which only two had been predicted by Southern hybridization analysis of *Frankia* genomic DNA. Comparison of the band sizes obtained for each group of genomic clones to those obtained by Southern hybridization of genomic DNA suggests that clones of groups #1 and #2 contain the two different copies of the 23S rRNA gene. These two groups of clones show bands with sizes predicted from previous experiments. A third

**TABLE 3 CHARACTERIZATION OF *FRANKIA* ACN14A GENOMIC CLONES**

| Groups | Fragments hybridizing to the 23S rDNA specific probe (kb) | | Genomic clones |
|---|---|---|---|
| | BamHI | BstEII | |
| 1 | 1.9 | 2.7 | λFr-3<br>λFr-8 |
| 2 | 2.8 | 2.5 | λFr-1<br>λFr-4<br>λFr-5 |
| 3 | 2.0 | 4.0 | λFr-7 |

Categorization of the genomic clones according to the band pattern observed after Southern hybridization of the individual clone DNAs digested with BamHI or BstEII. The 23S rDNA specific probe was used for this experiment.

band pattern corresponding to a 2.0 kb BamH1 fragment and a 4.0 kb BstEII fragment, is observed in one of the genomic clone (λFr-7). Unlike the two other band patterns, no bands with such sizes were detected in genomic DNA. This unexpected pattern is probably observed because one end of the λFr-7 insert generated by partial digestion with Mbo1, is found within both the 2.8 kb BamH1 fragment and the 2.5 kb BstEII fragment. A BamH1 fragment of intermediate size could then be the result of creation of a BamH1 site after ligation of the λ arms to the insert. The large BstEII fragment would thus be obtained after cutting of the clone at a site within the λ arm and at one of the sites present in the genomic fragment. Thus, even though no further analyses were done on λFr-7, it most likely contains the same 23S rDNA gene copy as the clones of group #2.

Large scale preparations of DNA were performed on representatives of each group (λFr-1, λFr-7, and λFr-8) and were analyzed by restriction with BamH1. The yield was found to vary greatly for the three lysate preparations. Subsequent work was limited to analysis of the λFr-8 clone since it was thought to contain a complete rRNA operon and large quantities were obtained.

## CHARACTERIZATION OF λFR-8, A GENOMIC CLONE CONTAINING ONE OF THE TWO RIBOSOMAL RNA OPERONS

### Cloning and subcloning

All the DNA fragments generated after digestion of the genomic clone λFr-8 with BamH1 have been subcloned and analyzed. Their approximate sizes are 0.3, 0.85, 0.9, 1.9, and 8.0 kb. The insert sizes determined by restriction analysis of the clones are generally in agreement with the approximation obtained by Southern hybridization analysis of *Frankia* AcN14a genomic DNA. However, the size of the largest DNA fragment which had an apparent size of 9.4 kb on Southern blots, was reduced to 8.0 kb after cloning and restriction analysis. For further analysis, other DNA fragments have subsequently been cloned.

In general, DNA fragments were subcloned in both orientations into M13 vectors and pTZ18 or pGEM4Z in order to facilitate sequencing and further mapping, respectively. Selected clones and their respective positions in λFr-8 are shown in Figure 6.

The vector M13mp19 was chosen to facilitate the construction of deletions by the Dale procedure (Dale *et al.*, 1985). M13 phage vectors have routinely been used for the preparation of single-stranded templates for the sequencing of relatively short DNA fragments. In order to facilitate data accumulation, the clones were named according to the scheme outlined here. Subclones were named according to the vector used for cloning. The abbreviations mp18, mp19 (or simply 18, 19) pG, pT were used for the vectors M13mp18, M13mp19, pGEM4Z and pTZ18, respectively. This is generally followed by the approximate size in kilobase pairs and the digestion from which the fragment originated (B, BamH1; E, EcoR1; H, HindIII; P, Pst1, S, Sal1; Sac, Sac1). In addition fragments cloned in both orientations could be differentiated by a bracketed number at the end of the name. Since the DNA templates used to obtained deletions were always in M13mp19, deletion clones were named according to the restriction enzyme(s) used for cloning, followed by a number representing the orientation of the clone (when the cloning was not directed) and the number of the respective deletion clone (e.g. PST(1)#3).

Verification of the plasmid clone identity was first achieved by measurement of the insert size by agarose gel electrophoresis of the DNA digested with the enzyme(s) used for the cloning. In addition, restriction analysis and/or hybridization experiments were done to further confirm the identity of the clones. The identity and the relative orientation of the M13 clones were determined by complementation tests (C-tests) and, in some cases, confirmed by Southern transfer followed by hybridization to the proper probe.

Sequence information was obtained from both strands in order to minimize the occurrence of sequencing errors. Deletion clones of the larger fragments were obtained by following the Dale procedure (Dale *et al.*, 1985).

**FIGURE 5 SOUTHERN HYBRIDIZATION ANALYSIS OF λFr-8**

Southern hybridization of the 23S rDNA specific probe (pnod4) to λFr-8 genomic clone DNA partially digested with BamH1. The probe corresponds to the 3' portion of Frankia AcN14a 23S rRNA gene. Aliquots of λFr-8 DNA (approximately 0.8 µg) were digested with decreasing amounts of BamH1.

**Lane 1** pnod4 digested with BamH1; **lane 2** pGEM4Z digested with BamH1; **lanes 3-9** λFr-8 DNA partially digested with BamH1. **lane 3** 5 units; **lane 4** 2.5 units; **lane 5** 1.25 unit; **lane 6** 0.75 unit; **lane 7** 0.325 unit; **lane 8** 0.15 unit; **lane 9** 0.07 unit.

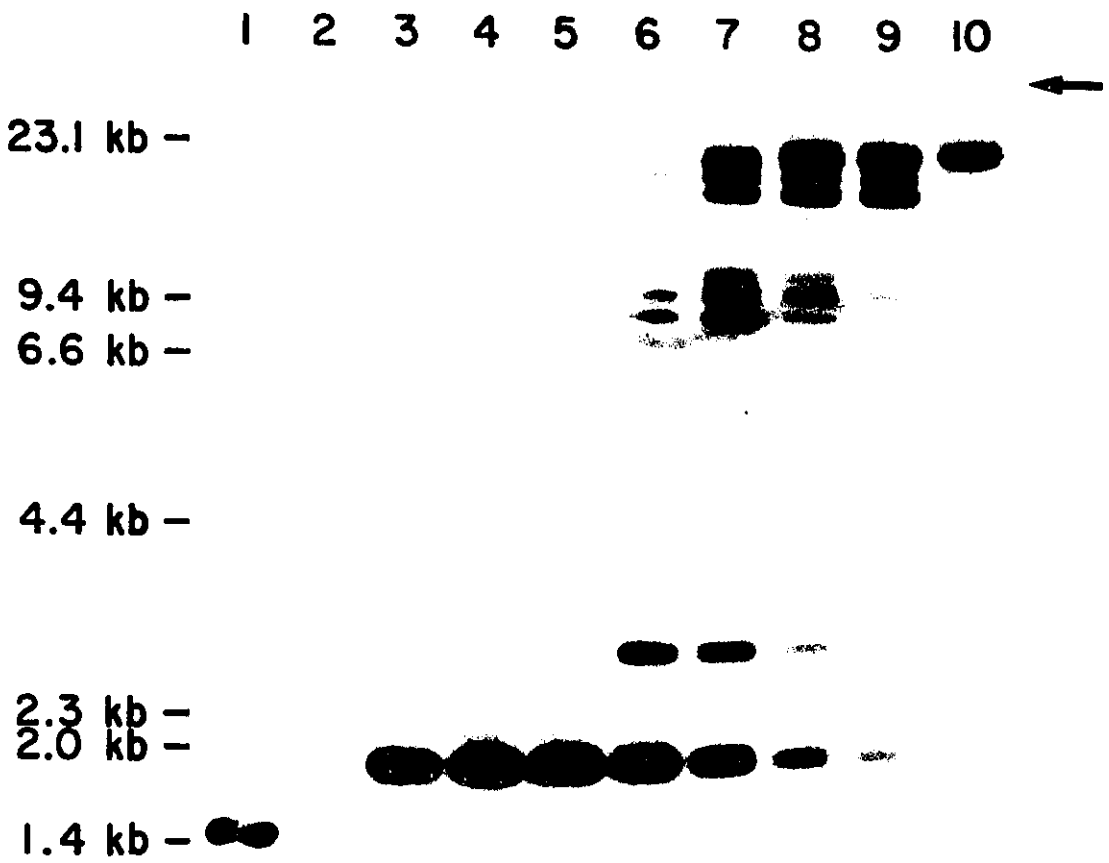
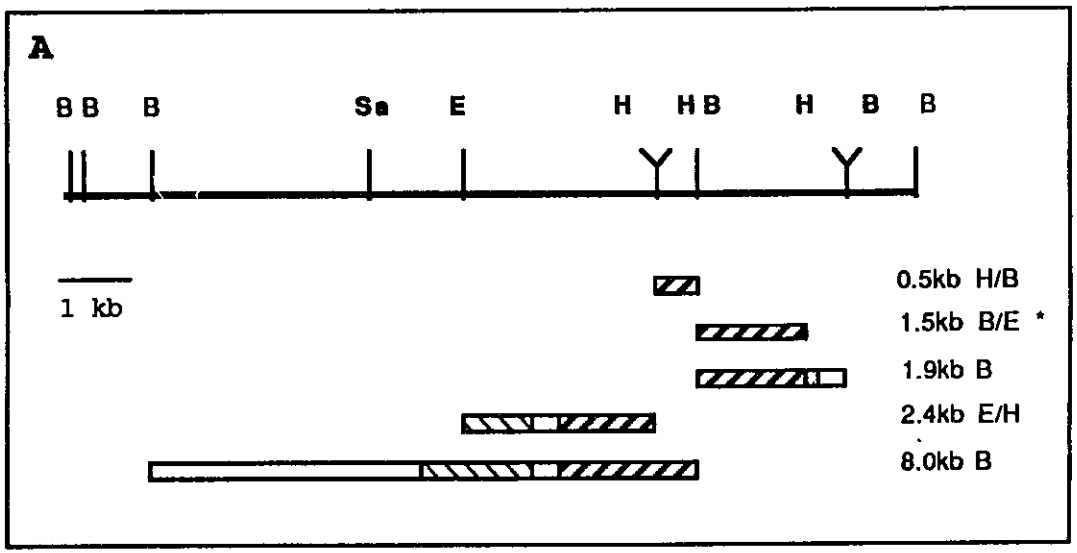

1 2 3 4 5 6 7 8 9 10
23.1 kb –
9.4 kb –
6.6 kb –
4.4 kb –
2.3 kb –
2.0 kb –
1.4 kb –

**FIGURE 6 PHYSICAL AND GENETIC MAP OF λFr-8.**

**A**. Simplified physical map of λFr-8 and relative positions of the various DNA fragments used as probes in Southern hybridization experiments. **B** Detailed physical and genetic map of the portion of the λFr-8 clone that comprises one of the two rRNA operons. The restriction site symbols written in bold are shown in both figures. The DNA fragment indicated with an asterisk has been derived from pnod4; the EcoR1 site was located within the vector polylinker region.




**LEGEND:**

putative promoter

putative terminator

16S rRNA gene

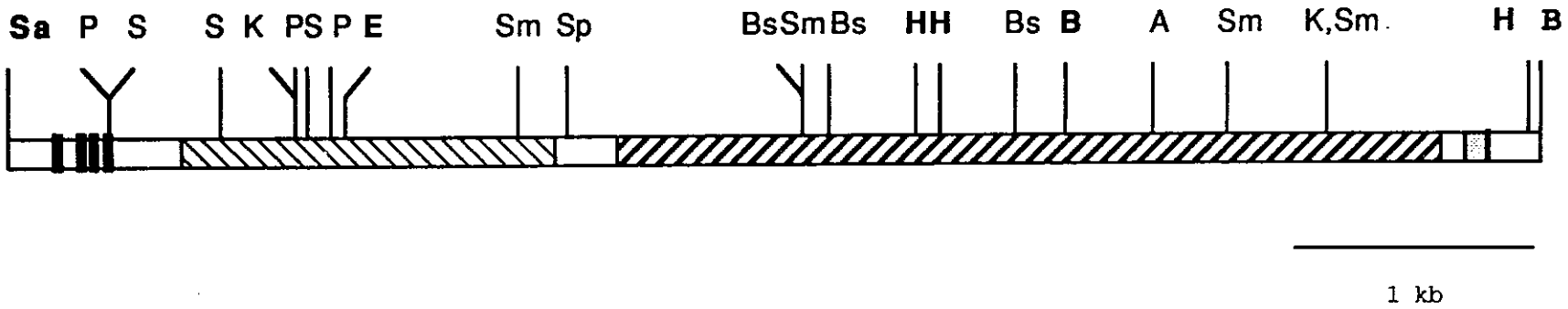
23S rRNA gene

5S rRNA gene

intergenic or intragenic regions

A: Acc1 K: Kpn1
B: BamH1 P: Pst1
Bs: BstEII S: Sac1
E: EcoR1 Sa: Sal1
H: HindIII Sm: Sma1
Sp: Sph1

**B**



Sequencing of the complementary strand was achieved either by the same method, by the use of synthesized primers or by further subcloning of the DNA fragments.

**Genetic and physical map of λFr-8**

Figure 5 represents a Southern blot of λFr-8 partially digested with BamH1. This partial digestion allowed us to order the BamH1 restriction fragments and to draw a preliminary restriction map. A more detailed map has been established with data obtained from additional restriction analyses, Southern hybridization analyses and sequencing (Figure 6).

Preliminary positioning of the rRNA genes was determined by sequencing the extremities of the λFr-8 DNA fragments generated by digestion with BamH1 followed by a sequence comparison search. The 16S and the 5S rRNA genes were located on the 8.0 kb and the 1.9 kb fragments respectively while the 23S rRNA gene is split between these two fragments (Figure 6B).

## 1B.COMPARATIVE ANALYSIS OF *FRANKIA* AcN14A RIBOSOMAL RNA OPERON

**General organization of the rRNA genes in *Frankia***

Approximately half (6.0 kb) of the λFr-8 clone has been sequenced. The high conservation of the rRNA gene sequences allowed the identification of sequences corresponding to the mature rRNAs by comparison with *E. coli*, *B. subtilis* and *S. ambofaciens* sequences. Sizes of 1509 and 3105 bases are predicted for *Frankia* AcN14a 16S and 23S rRNAs, respectively. Staining of *Frankia* AcN14a total RNA electrophoretically separated on a gel confirmed the size difference between *E. coli* and *Frankia* AcN14a rRNAs (Figure 8); bands corresponding to rRNAs of 3.1 and 1.5 kb in length are observed.

The nucleotide sequence obtained from λFr-8 is presented in Figure 9. As in most bacterial genomes, the order of the rRNA genes in *Frankia* AcN14a

**FIGURE 8 ANALYSIS OF *FRANKIA* ACN14A 16S AND 23S RIBOSOMAL RNAS**

*Frankia* AcN14a total RNA was electrophoretically separated on a denaturing formamide-agarose gel and stained. Prominent bands corresponding to the 16S and 23S rRNAs are observed. The sizes of *E. coli* rRNAs are included as references.

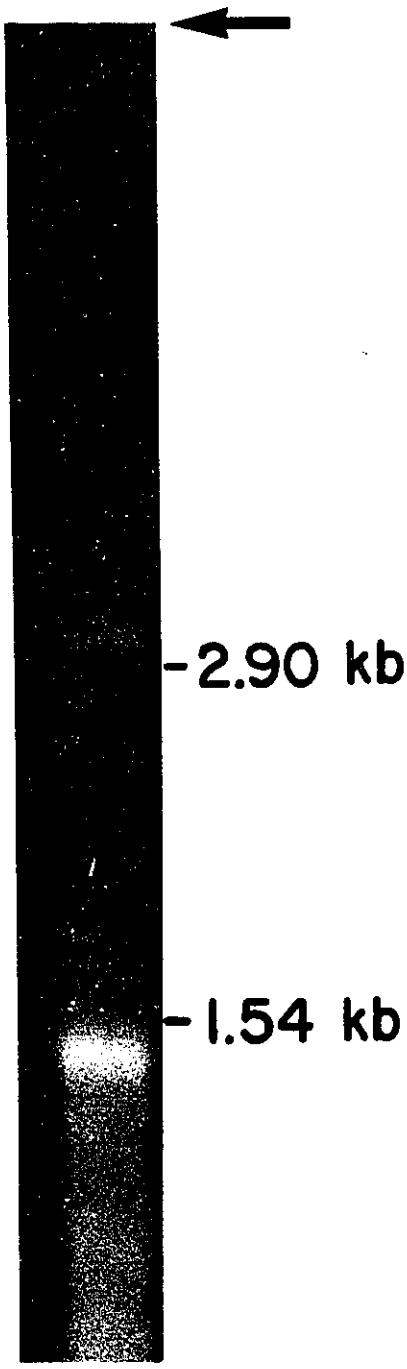
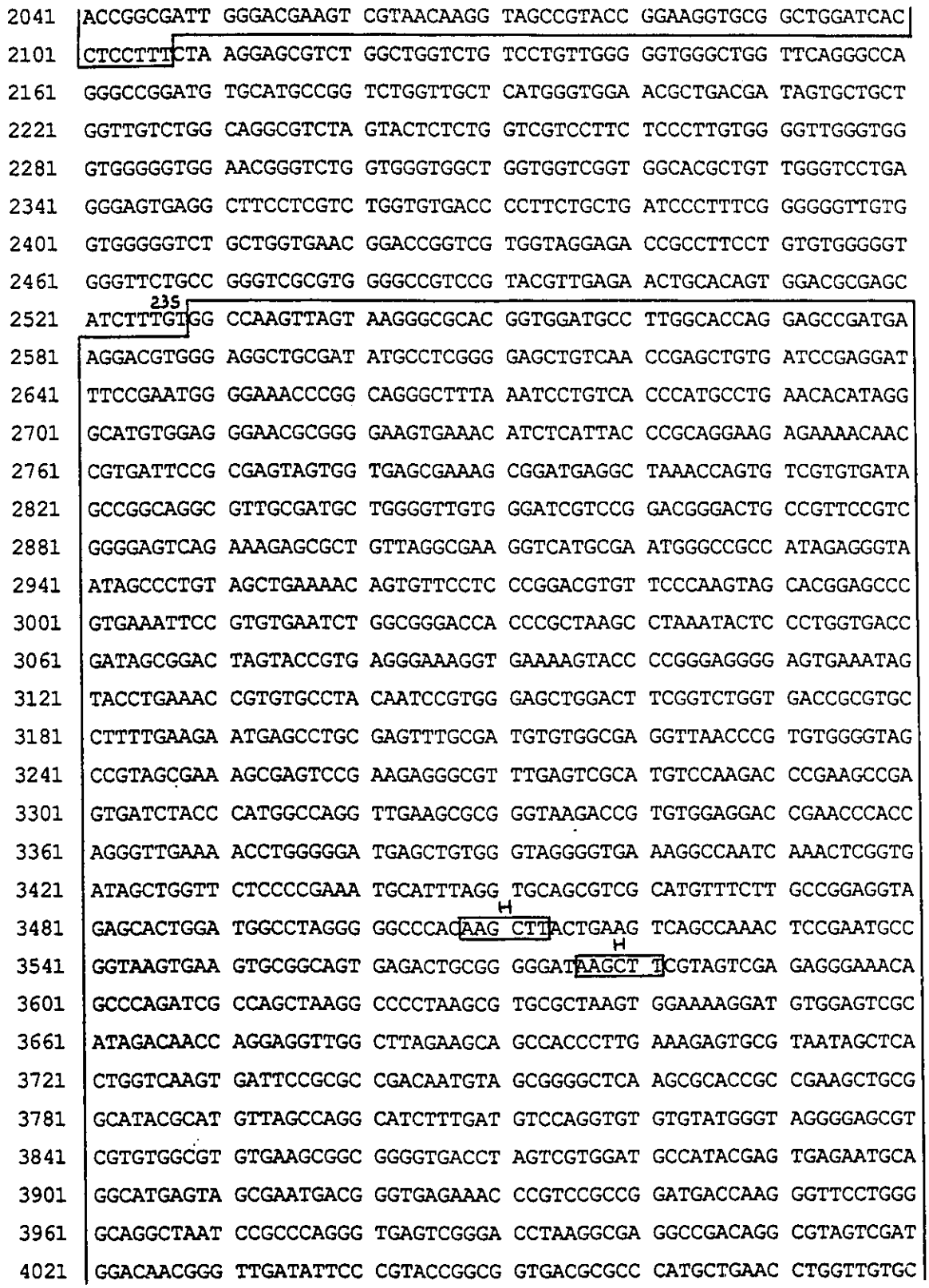
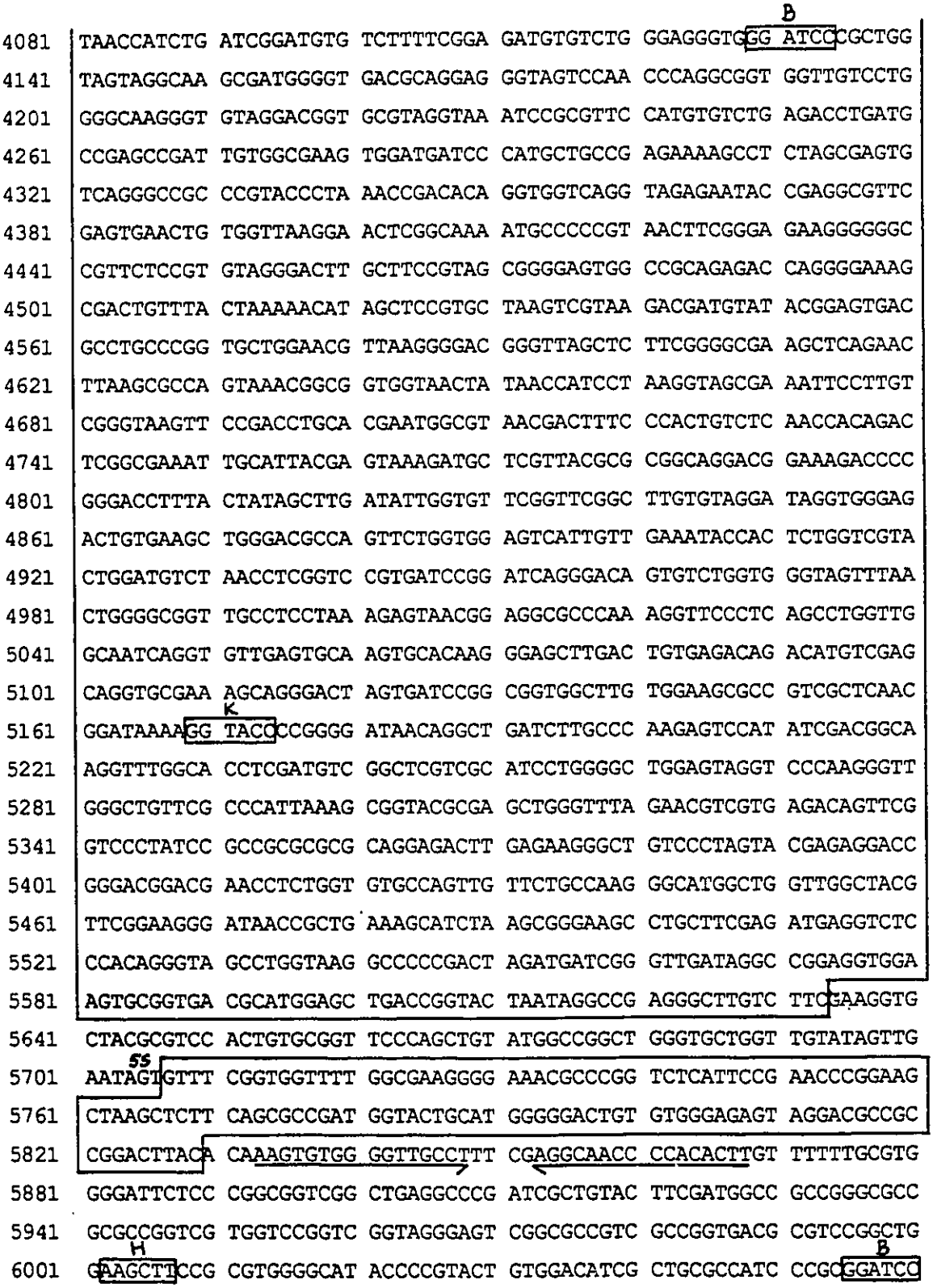

2.90 kb
1.54 kb

**FIGURE 9 NUCLEOTIDE SEQUENCE OF ONE OF THE TWO *FRANKIA* ACN14A RIBOSOMAL DNA REGIONS.**

Nucleotide sequence of *Frankia* AcN14a rDNA contained in λFr-8. The putative promoters are underlined, the regions corresponding to the putative mature 16S, 23S and 5S rRNAs are boxed and the putative terminator is indicated by two arrows. Sites for the restriction endonucleases, BamH1 (B), EcoR1 (E), HindIII (H) and Kpn1 (K) are boxed.


1 GGTCGCCGCA GGTGACTGGG CCCCGCGGAC AGATGCGGAC GGACGGTACG CGGCCTGCTC
61 GTGGGCAACT TGGCCCAGAC CGCCAAGCCT CCCATATCAG AACATCACAA CGAACCTCTC
121 CGCGAGTGGT GTGACATAGG TCACACATCT GGGCGGGGGA TTGATTTGAC GCCCCCCGTC
181 ATCCACCCGT AATGTTCTCT CTGTCAGCGC AGCAGGCCAG ACGCCGAACA GGGCGGCTCC
241 ACTCGCTGCA GAGCCCCTCC CGGAGGGAGC TCGGGGAGAG CAGGTGCGAG GCGGCCCCTT
301 CTTGGGCGTA CTGACGAGCA CCGCTCCATA AAACTTCGTA AGTGGCTAGG AATCCGCACG
361 AGTACGACAT CTCCCGATGT CATCTCGTTC GCGCCTGCTC CTTGAGAACT CAACAGCGTG
421 CCGATTGTCA GTGCCAATTG TTTTACCCCG TTCATGGCCT GGCTGGTTGT TTTCTGCGCA
481 TCGTTTGTGG TGTGTGGGGG GTGGCTGGTT GTGGTTGTGA GGGATTCCTT TGGACACGGC
541 ATCACGGTTG TGGTGTCGAT GTCGGGGTCT GCCGGTTTGT CCGGCACTCT CGAAATCATT
601 GATGGAGAGT TTGATCCTGG CTCAGGACGA ACGCTGGCGG CGTGCTTAAC ACATGCAAGT
661 CGAGCGAGGG GCTTCGGCTC TAGCGGCGAA CGGGTGAGTA ACACGTGGGC AACCTGCCCC
721 GAGCTCTGGA ATAACCTCGG GAAACCGGGG CTAATGCCGG ATATGACGCT ACCGGGCATC
781 TGGTGGTGTG GAAAGATTTA TCGGCTCGGG ATGGGCCCGC GGCCTATCAG CTTGTTGGTG
841 GGGTGATGGC CTACCAAGGC GACGACGGGT AGCCGGCCTG AGAGGGCGAC CGGCCACACT
901 GGGACTGAGA CACGGCCCAG ACTCCTACGG GAGGCAGCAG TGGGGAATAT TGCGCAATGG
961 GCGAAAGCCT GACGCAGCGA CGCCGCGTGG GGGATGACGG CCTTCGGGTT GTAAACCTCT
1021 TTCAGCAGGG ACGAAGCGCA AGTGACGGTA CGTGCAGAAG AAGCACCGGC CAACTACGTG
1081 CCAGCAGCCG CGGTAATACG TAGGGTGCAA GCGTTGTCCG GAATTATTGG GCGTAAAGAG
1141 CTCGTAGGCG GCCTGTCGCG TCGGCTGTGA AAACCCGGGG CTCAACCCCG GGCCTGCAGT
1201 CGATACGGGC AGGCTAGAGT CCGGCAGGGG AGACTGGAAT TCCTGGTGTA GCGGTGAAAT
1261 GCGCAGATAT CAGGAGGAAC ACCGGTGGCG AAGGCGGGTC TCTGGGCCGG AACTGACGCT
1321 AAGGAGCGAA AGCGTGGGGA GCGAACAGGA TTAGATACCC TGGTAGTCCA CGCCGTAAAC
1381 GTTGGGCGCT AGGTGTGGGG GACCTTCCAC GGCCTCCGTG CCGCAGCTAA CGCATTAAGC
1441 GCCCCGCCTG GGGAGTACGG CCGCAAGGCT AAAACTCAAA GGAATTGACG GGGGCCCGCA
1501 CAAGCGGCGG AGCATGTGGC TTAATTCGAT GCAACGCGAA GAACCTTACC AGGGCTTGAC
1561 ATGCAGGGAA ATCTCGTAGA GATACGGGGT CCGTAAGGGT CCTGCACAGG TGGTGCATGG
1621 CTGTCGTCAG CTCGTGTCGT GAGATGTTGG GTTAAGTCCC GCAACGAGCG CAACCCTCGT
1681 CCTATGTTGC CAGCGAGTTA TGTCGGGGAC TCATAGGAGA CTGCCGGGGT CAACTCGGAG
1741 GAAGGTGGGG ATGACGTCAA GTCATCATGC CCCTTACGTC CTGGGCTGCA CACATGCTAC
1801 AATGGCCGGT ACAAAGGGCT GCGATACCGT GAGGTGGAGC GAATCCCAAA AAGCCGGTCT
1861 CAGTTCGGAT CGGGGTCTGC AACTCGACCC CGTGAAGTCG GAGTCGCTAG TAATCGCAGA
1921 TCAGCAATGC TGCGGTGAAT ACGTTCCCGG GCCTTGTACA CACCGCCCGT CACGTCACGA
1981 AAGTCGGTAA CACCCGAAGC CGGTGGCCTA ACCCTTGTGG GGGGAGCCGT CGAAGGTGGG
16S
K
E


2041 ACCGGCGATT GGGACGAAGT CGTAACAAGG TAGCCGTACC GGAAGGTGCG GCTGGATCAC
2101 CTCCTTTCTA AGGAGCGTCT GGCTGGTCTG TCCTGTTGGG GGTGGGCTGG TTCAGGGCCA
2161 GGGCCGGATG TGCATGCCGG TCTGGTTGCT CATGGGTGGA ACGCTGACGA TAGTGCTGCT
2221 GGTTGTCTGG CAGGCGTCTA GTACTCTCTG GTCGTCCTTC TCCCTTGTGG GGTTGGGTGG
2281 GTGGGGGTGG AACGGGTCTG GTGGGTGGCT GGTGGTCGGT GGCACGCTGT TGGGTCCTGA
2341 GGGAGTGAGG CTTCCTCGTC TGGTGTGACC CCTTCTGCTG ATCCCTTTCG GGGGGTTGTG
2401 GTGGGGGTCT GCTGGTGAAC GGACCGGTCG TGGTAGGAGA CCGCCTTCCT GTGTGGGGGT
2461 GGGTTCTGCC GGGTCGCGTG GGGCCGTCCG TACGTTGAGA ACTGCACAGT GGACGCGAGC
2521 ATCTTTGTGG CCAAGTTAGT AAGGGCGCAC GGTGGATGCC TTGGCACCAG GAGCCGATGA
2581 AGGACGTGGG AGGCTGCGAT ATGCCTCGGG GAGCTGTCAA CCGAGCTGTG ATCCGAGGAT
2641 TTCCGAATGG GGAAACCCGG CAGGGCTTTA AATCCTGTCA CCCATGCCTG AACACATAGG
2701 GCATGTGGAG GGAACGCGGG GAAGTGAAAC ATCTCATTAC CCGCAGGAAG AGAAAACAAC
2761 CGTGATTCCG CGAGTAGTGG TGAGCGAAAG CGGATGAGGC TAAACCAGTG TCGTGTGATA
2821 GCCGGCAGGC GTTGCGATGC TGGGGTTGTG GGATCGTCCG GACGGGACTG CCGTTCCGTC
2881 GGGGAGTCAG AAAGAGCGCT GTTAGGCGAA GGTCATGCGA ATGGGCCGCC ATAGAGGGTA
2941 ATAGCCCTGT AGCTGAAAAC AGTGTTCCTC CCGGACGTGT TCCCAAGTAG CACGGAGCCC
3001 GTGAAATTCC GTGTGAATCT GGCGGGACCA CCCGCTAAGC CTAAATACTC CCTGGTGACC
3061 GATAGCGGAC TAGTACCGTG AGGGAAAGGT GAAAAGTACC CCGGGAGGGG AGTGAAATAG
3121 TACCTGAAAC CGTGTGCCTA CAATCCGTGG GAGCTGGACT TCGGTCTGGT GACCGCGTGC
3181 CTTTTGAAGA ATGAGCCTGC GAGTTTGCGA TGTGTGGCGA GGTTAACCCG TGTGGGGTAG
3241 CCGTAGCGAA AGCGAGTCCG AAGAGGGCGT TTGAGTCGCA TGTCCAAGAC CCGAAGCCGA
3301 GTGATCTACC CATGGCCAGG TTGAAGCGCG GGTAAGACCG TGTGGAGGAC CGAACCCACC
3361 AGGGTTGAAA ACCTGGGGGA TGAGCTGTGG GTAGGGGTGA AAGGCCAATC AAACTCGGTG
3421 ATAGCTGGTT CTCCCCGAAA TGCATTTAGG TGCAGCGTCG CATGTTTCTT GCCGGAGGTA
3481 GAGCACTGGA TGGCCTAGGG GGCCCACAAG CTTACTGAAG TCAGCCAAAC TCCGAATGCC
3541 GGTAAGTGAA GTGCGGCAGT GAGACTGCGG GGGATAAGCT TCGTAGTCGA GAGGGAAACA
3601 GCCCAGATCG CCAGCTAAGG CCCCTAAGCG TGCGCTAAGT GGAAAAGGAT GTGGAGTCGC
3661 ATAGACAACC AGGAGGTTGG CTTAGAAGCA GCCACCCTTG AAAGAGTGCG TAATAGCTCA
3721 CTGGTCAAGT GATTCCGCGC CGACAATGTA GCGGGGCTCA AGCGCACCGC CGAAGCTGCG
3781 GCATACGCAT GTTAGCCAGG CATCTTTGAT GTCCAGGTGT GTGTATGGGT AGGGGAGCGT
3841 CGTGTGGCGT GTGAAGCGGC GGGGTGACCT AGTCGTGGAT GCCATACGAG TGAGAATGCA
3901 GGCATGAGTA GCGAATGACG GGTGAGAAAC CCGTCCGCCG GATGACCAAG GGTTCCTGGG
3961 GCAGGCTAAT CCGCCCAGGG TGAGTCGGGA CCTAAGGCGA GGCCGACAGG CGTAGTCGAT
4021 GGACAACGGG TTGATATTCC CGTACCGGCG GTGACGCGCC CATGCTGAAC CTGGTTGTGC


B
4081 TAACCATCTG ATCGGATGTG TCTTTTCGGA GATGTGTCTG GGAGGGTGGG ATCCCGCTGG
4141 TAGTAGGCAA GCGATGGGGT GACGCAGGAG GGTAGTCCAA CCCAGGCGGT GGTTGTCCTG
4201 GGGCAAGGGT GTAGGACGGT GCGTAGGTAA ATCCGCGTTC CATGTGTCTG AGACCTGATG
4261 CCGAGCCGAT TGTGGCGAAG TGGATGATCC CATGCTGCCG AGAAAAGCCT CTAGCGAGTG
4321 TCAGGGCCGC CCGTACCCTA AACCGACACA GGTGGTCAGG TAGAGAATAC CGAGGCGTTC
4381 GAGTGAACTG TGGTTAAGGA ACTCGGCAAA ATGCCCCCGT AACTTCGGGA GAAGGGGGGC
4441 CGTTCTCCGT GTAGGGACTT GCTTCCGTAG CGGGGAGTGG CCGCAGAGAC CAGGGGAAAG
4501 CGACTGTTTA CTAAAAACAT AGCTCCGTGC TAAGTCGTAA GACGATGTAT ACGGAGTGAC
4561 GCCTGCCCGG TGCTGGAACG TTAAGGGGAC GGGTTAGCTC TTCGGGGCGA AGCTCAGAAC
4621 TTAAGCGCCA GTAAACGGCG GTGGTAACTA TAACCATCCT AAGGTAGCGA AATTCCTTGT
4681 CGGGTAAGTT CCGACCTGCA CGAATGGCGT AACGACTTTC CCACTGTCTC AACCACAGAC
4741 TCGGCGAAAT TGCATTACGA GTAAAGATGC TCGTTACGCG CGGCAGGACG GAAAGACCCC
4801 GGGACCTTTA CTATAGCTTG ATATTGGTGT TCGGTTCGGC TTGTGTAGGA TAGGTGGGAG
4861 ACTGTGAAGC TGGGACGCCA GTTCTGGTGG AGTCATTGTT GAAATACCAC TCTGGTCGTA
4921 CTGGATGTCT AACCTCGGTC CGTGATCCGG ATCAGGGACA GTGTCTGGTG GGTAGTTTAA
4981 CTGGGGCGGT TGCCTCCTAA AGAGTAACGG AGGCGCCCAA AGGTTCCCTC AGCCTGGTTG
5041 GCAATCAGGT GTTGAGTGCA AGTGCACAAG GGAGCTTGAC TGTGAGACAG ACATGTCGAG
5101 CAGGTGCGAA AGCAGGGACT AGTGATCCGG CGGTGGCTTG TGGAAGCGCC GTCGCTCAAC
K
5161 GGATAAAAGG TACCCCGGGG ATAACAGGCT GATCTTGCCC AAGAGTCCAT ATCGACGGCA
5221 AGGTTTGGCA CCTCGATGTC GGCTCGTCGC ATCCTGGGGC TGGAGTAGGT CCCAAGGGTT
5281 GGGCTGTTCG CCCATTAAAG CGGTACGCGA GCTGGGTTTA GAACGTCGTG AGACAGTTCG
5341 GTCCCTATCC GCCGCGCGCG CAGGAGACTT GAGAAGGGCT GTCCCTAGTA CGAGAGGACC
5401 GGGACGGACG AACCTCTGGT GTGCCAGTTG TTCTGCCAAG GGCATGGCTG GTTGGCTACG
5461 TTCGGAAGGG ATAACCGCTG AAAGCATCTA AGCGGGAAGC CTGCTTCGAG ATGAGGTCTC
5521 CCACAGGGTA GCCTGGTAAG GCCCCCGACT AGATGATCGG GTTGATAGGC CGGAGGTGGA
5581 AGTGCGGTGA CGCATGGAGC TGACCGGTAC TAATAGGCCG AGGGCTTGTC TTCGAAGGTG
5641 CTACGCGTCC ACTGTGCGGT TCCCAGCTGT ATGGCCGGCT GGGTGCTGGT TGTATAGTTG
SS
5701 AATAGTGTTT CGGTGGTTTT GGCGAAGGGG AAACGCCCGG TCTCATTCCG AACCCGGAAG
5761 CTAAGCTCTT CAGCGCCGAT GGTACTGCAT GGGGGACTGT GTGGGAGAGT AGGACGCCGC
5821 CGGACTTACA CAAAGTGTGG GGTTGCCTTT CGAGGCAACC CCACACTTGT TTTTTGCGTG
5881 GGGATTCTCC CGGCGGTCGG CTGAGGCCCG ATCGCTGTAC TTCGATGGCC GCCGGGCGCC
5941 GCGCCGGTCG TGGTCCGGTC GGTAGGGAGT CGGCGCCGTC GCCGGTGACG CGTCCGGCTG
H
6001 GAAGCTTCCG CGTGGGGCAT ACCCCGTACT GTGGACATCG CTGCGCCATC CCGCGGATCC
B

is 5' 16S-23S-5S 3'. In addition the genes are closely linked suggesting their organization into an operon. The 16S-23S rRNA spacer region is 419 nucleotides in length and 5S rRNA gene is 73 nucleotides downstream from the 23S rRNA gene. A similar organization has been reported for the rRNA genes of the *Frankia* strain ORS020606; the 16S-23S and 23S-5S rRNA spacer regions have 410 and 68 nucleotides in length, respectively.

Analysis of the sequences between the *Frankia* AcN14a 16S and 23S rRNA genes, and downstream of the 5S rRNA gene failed to identify any tRNA sequences. The absence of tRNA genes has also been reported for the rRNA operon of *Streptomyces* (Baylis and Bibb, 1988a; Suzuki *et al*, 1988b; Pernodet *et al.*, 1989) and *Mycobacterium* (Suzuki et al., 1988a), two other groups of actinomycetes. The lack of tRNA genes in these two regions has been suggested to be typical for filamentous bacteria (van Wezel *et al.*, 1991).

**Analysis of the 16S rRNA gene - primary and secondary structures**

The presumptive 5' and 3' ends of the region coding for the 16S rRNA were assigned to nucleotide positions 600 and 2109, respectively (Figure 9), by comparison with the 16S rRNA of other eubacterial sequences (Brosius *et al.*, 1981; Pernodet *et al.*, 1989; Baylis and Bibb, 1987). The *Frankia* AcN14a 16S rRNA is 1509 bp in length, thus almost identical to that of *Frankia* strain ORS020606 (1510 bp) and comparable to those of *Leptospira interrogans* and *Thermus thermophilus* (Neefs *et al.*, 1991). *Frankia* AcN14a 16S rRNA is, however, relatively smaller than the sizes for sequenced eubacterial species. *E.coli* and *B. subtilis* 16S rRNAs are respectively, 1542 and 1550 nucleotide long whereas the length of those of more closely related eubacteria such as *Mycobacterium* and *Streptomyces* ranges from 1528 to 1548 nucleotides. As it has been reported for other eubacteria with a G+C rich genome, the *Frankia* AcN14a rRNA gene base composition is much lower and does not reflect that of the genome (to be discussed later).

Three deletions and one insertion in the first domain of the molecule account for most of the size difference observed between *Frankia* AcN14a and

*E. coli* 16S rRNAs. Similarly, the size variation observed between other eubacteria and chloroplast 16S rRNAs also results from changes in the number of nucleotides present in this same region (Raué *et al.*, 1988). The variation observed in this region as well as others will be discussed in further detail in the next section.

The nucleotide sequence of *Frankia* AcN14a 16S rRNA gene was compared to that of selected eubacterial 16S rRNA sequences. This is possible because of the presence of strongly conserved regions of primary structure through out the molecule (Raué *et al.*, 1988). *Frankia* AcN14a 16S rRNA sequence is 99% identical to that reported for the strain ORS020606 (GeneBank) and shows more homology with *M. bovis* and *S. ambofaciens*, two other actinomycete species, than to *B. subtilis* or *E. coli*. Homology values of 91%, 92% and 84% are obtained with *M. bovis*, *S. ambofaciens* and *B. subtilis*, respectively. By comparison, the homology value with *E. coli* is 78%. The similar homology values obtained between *Frankia* AcN14a and both *M. bovis* and *S. ambofaciens* 16S rRNA sequences does not mean that there is insufficient variability between these sequences to establish phylogenetic relationships. In fact, some regions of *Frankia* AcN14a 16S rRNA resemble more those of *S. ambofaciens* whereas other regions are more similar to those of *M. bovis*. This provides an explanation for the different relative distances calculated from partial 16S rRNA sequences of these three bacterial groups by different researchers (e.g. Hahn *et al.*, 1989b; Stackebrandt *et al.*, 1990). Homology values generally depend upon the length and the number of the sequences to be compared as well as the method used for the calculations.

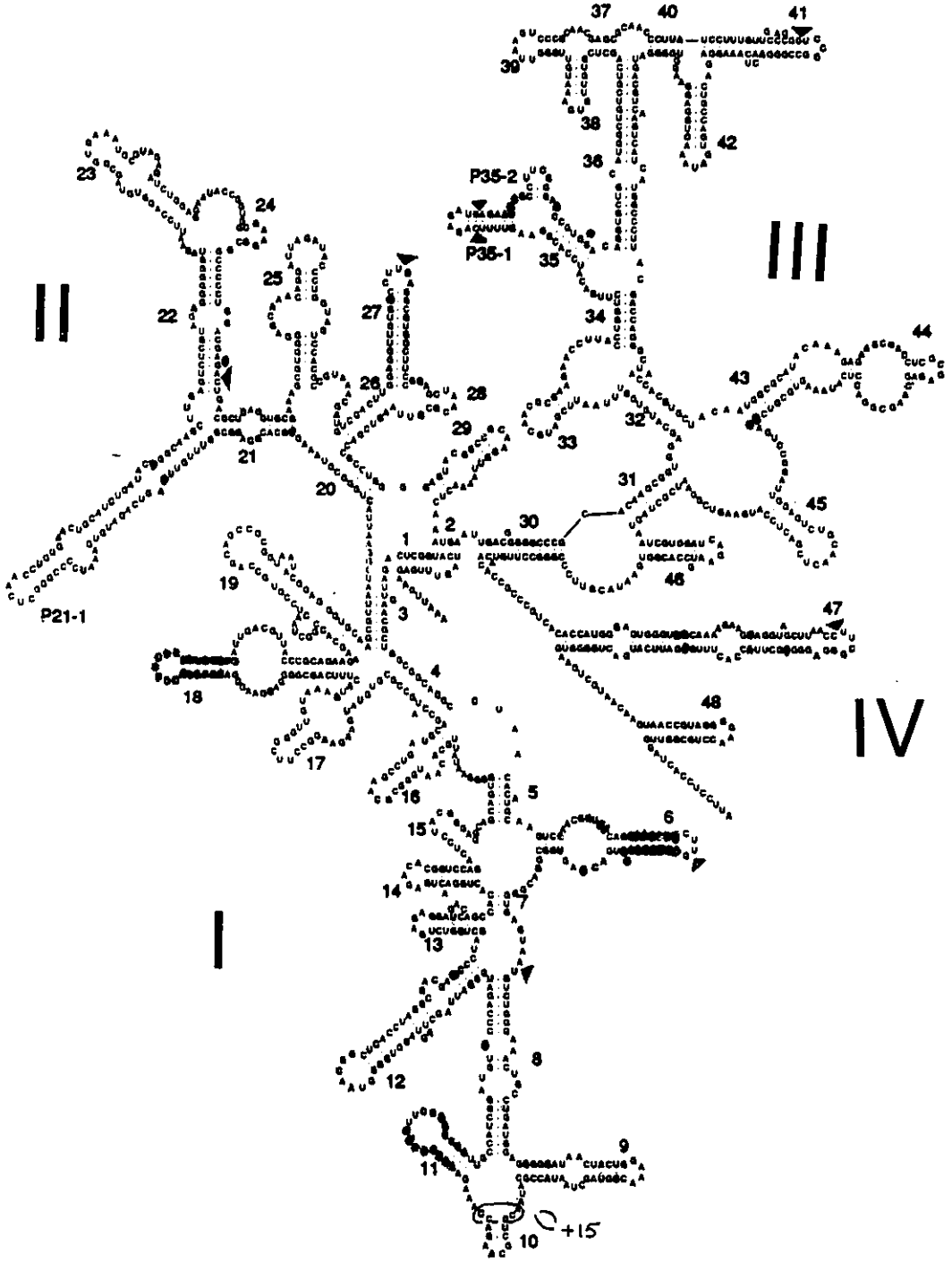
*Frankia* AcN14a 16S rRNA can be folded into a secondary structure model similar to that proposed for *E. coli* (reviewed by Raué *et al.*, 1988). The variation observed between the 16S rRNAs of these two organisms is shown in Figure 10. All helix and nucleotide numbering corresponds to the *Frankia* AcN14a sequence (Figure 10) unless stated otherwise. Most nucleotide substitutions observed between *Frankia* AcN14a and *E. coli* 16S rRNAs are located in base-paired regions and are compensatory or neutral in nature. A compensatory base change refers to the mutation of a canonical base pair to

**FIGURE 10 PRIMARY AND SECONDARY STRUCTURE VARIATION BETWEEN *FRANKIA* ACN14A AND *E. COLI* 16S RIBOSOMAL RNAS**

*Frankia* AcN14a 16S rRNA sequence deduced from the corresponding gene sequence is compared to that of *E. coli* (Neefs et al., 1991). The positions conserved between the *Frankia* AcN14a and *E. coli* 16S rRNAs are indicated by the corresponding nucleotides and the positions that vary between these two organisms are circled. Roman numbers indicate domains. Helices are numbered in the order of occurrence from 5' to 3'. Helices bearing a single number are common to the prokaryotic and eukaryotic models. Helices bearing a composite number preceded by P are prokaryote-specific. The locations and size of the large insertions present in *Frankia* AcN14a sequence are indicated by thick lines. The smaller insertions, the deletions and the different substitutions are indicated as follow:

Color code

| | |
|---|---|
| yellow circle: | compensatory substitution |
| orange circle: | neutral substitution |
| blue circle: | noncompensatory substitution |
| green circle: | deletion |
| black triangle: | small insertion (1 to 4 nucleotides) |


I
II
III
IV
P21-1
P35-1
P35-2
+15

another canonical base pair (e.g. G-C → A-U) whereas a neutral base change refers to the mutation of a canonical base pair to a noncanonical base pair (e.g. G-C → G-U or U-G). The *Frankia* AcN14a secondary structure model differs from the latter in three regions, which encompass helices 6, 10-11 and 18. In these regions, the *Frankia* AcN14a sequence can be folded into secondary structures similar to the those proposed for the corresponding sequences of *S. ambofaciens* (Pernodet *et al.*, 1989).

As stated earlier, *Frankia* AcN14a 16S rRNA differs from that of *E. coli* by three deletions and one addition. Superimposition of the 16S rRNA sequences of *Frankia* AcN14a, *S. ambofaciens* and *E. coli* has allowed identification of two additional regions, located in domain III, in which differences in nucleotide composition and minor modifications in secondary structures are observed. All these changes are located in previously defined variable areas of the 16S rRNA molecule (Neefs *et al.*, 1991) and consequently are unlikely to interfere with ribosome function. Half the nucleotide substitutions observed between the *Frankia* strains AcN14a and ORS020606 are in fact located within these variable regions. In addition, for three of these five regions, the nucleotide sequences of four additional *Frankia* strains have been published (Hahn *et al.*, 1989a; Harry *et al.*, 1991) and will thus be included in the comparative analysis.

The first region consists of helix 6 of which the nucleotide composition and the length vary (Figure 11). In *Frankia* AcN14a 16S rRNA, this helical structure comprises twenty-nine nucleotides and corresponds to nucleotides 61 to 89. Two neutral substitutions (positions 69 and 72) and one additional nucleotide are present in this region of *Frankia* strain ORS020606 16S rRNA. The position of the additional cytosine residue is indicated by a triangle in Figure 11. The corresponding regions of *B. subtilis* (Green *et al.*, 1985), *E. coli* (Brosius *et al.*, 1981), *M. bovis* (Suzuki *et al.*, 1988) and *S. ambofaciens* (Pernodet *et al.*, 1989) are longer that of *Frankia*. Between six and nine additional nucleotides are present in the corresponding helices of the three other Gram positive eubacteria whereas seventeen are present in *E. coli*. Part of the middle loop and the terminal base-paired structure proposed for *E. coli* are absent in *Frankia* AcN14a. Shortened versions of the terminal stem structure have also

been proposed for the corresponding regions of *B. subtilis*, *M. bovis* and *S. ambofaciens*. In *M. bovis*, there exists the possibility for the formation of two additional base-pairs in the middle loop as indicated in Figure 11. Other portions of helix 6 are more conserved. The nucleotide composition of the proximal stem is almost identical in these five organisms. In addition, the nucleotide composition of the terminal loop as well as the last nucleotide pair of the terminal stem are conserved among the three actinomycetes.

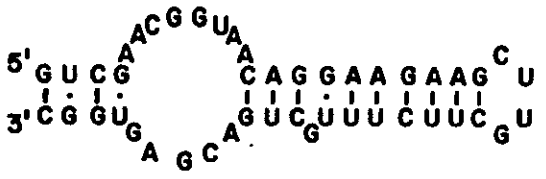
In summary, helix 6 of *Frankia* AcN14a and ORS020606 16S rRNAs differ from that of *E. coli*. The observed differences probably result from the deletion of about half of the nucleotides of the internal loop as well as a portion of the terminal stem structure of the corresponding *E. coli* 16S rRNA helix. This region of the *Frankia* strains AcN14a and ORS020606 rRNAs differ by two neutral substitutions and that of the latter contains one additional nucleotide. Deletion in this region of the 16S rRNA is not unique to *Frankia*; similar, although smaller deletions are observed in at least three other eubacteria. In addition, size variation in helix 6 has been observed in archaebacterial and organellar 16S rRNA sequences as well as in the corresponding eukaryotic small subunit rRNAs (Neefs *et al.*, 1991).

The second variable region observed between *Frankia* AcN14a and *E. coli* encompasses helices 10 and 11. This region corresponds to nucleotides 166 to 206 of *Frankia* AcN14a 16S rRNA. In *E. coli* 16S rRNA, the first of these helices is short whereas the second is long. In contrast, the 16S rRNA of *Frankia* AcN14a has an extended version of helix 10 and a short helix 11 (Figure 12). Thus in comparison to *E. coli*, *Frankia* AcN14a 16S rRNA appears to contain an addition in helix 10 and a deletion in helix 11.

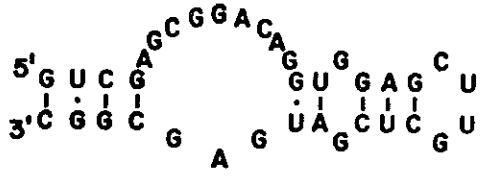
Similar secondary structures have been proposed for the corresponding regions of *S. ambofaciens* (Pernodet *et al.*, 1989) and two other *Frankia* strains, Ag45/Mut15 and AgB1.9 (Hahn *et al.*, 1989a). In addition, the corresponding regions of two other Gram positive eubacteria, *B. subtilis* (Green *et al.*, 1985) and *M. bovis* (Suzuki *et al.*, 1988) can be folded into similar secondary structures. While the primary sequence of the region of *Frankia* strain

**FIGURE 11 PROPOSED SECONDARY STRUCTURES OF HELIX 6 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED EUBACTERIA**

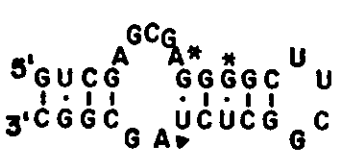
*Frankia* AcN14a 16S rRNA nucleotide sequence corresponding to positions 61 to 89 (see Figure 10) as well as the corresponding regions of *B. subtilis*, *M. bovis* and *S. ambofaciens* 16S rRNAs were folded with the help of the program RNAFOLD. In all cases, the *Frankia* AcN14a nucleotide sequence has been folded into structures showing maximal similarity to the prokaryotic model. The secondary structure proposed for the corresponding region of *E. coli* has been included as a reference (Neefs *et al.*, 1991). --- indicates possible additional base-pairings, the triangle indicates the position of the additional cytosine residue in *Frankia* strain ORS020606 and the asterisks indicate positions found to vary between *Frankia* AcN14a and ORS20606.



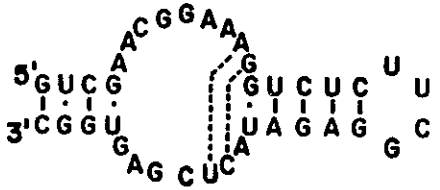
E. coli



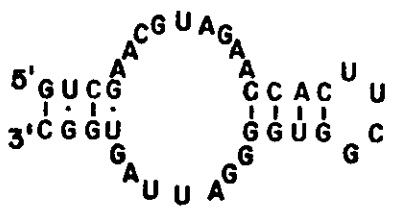
B. subtilis



Frankia sp. AcN14a



M. bovis



S. ambofaciens

ORS020606 16S rRNA could be folded in a secondary structure identical to that proposed for AcN14a, some difficulties were encountered in our attempt to fold the corresponding sequences of two other *Frankia* strains, AcNI1 and PtI1 (Harry *et al.*, 1991). For these two strains, structures similar to helices 9 and 10 of the three other *Frankia* strains could be formed. However, the secondary structure of the region corresponding to helix 11 could not be determined. Alignment of the *Frankia* AcN14a with those of the other *Frankia* strains suggests the deletion of about 15 nucleotides in the AcNI1 and PtI1 sequences. For this reason, the sequences of the last two strains have not be used in the comparison analysis of helix 11.

The first of the two helices proposed for *Frankia* AcN14a, contains ten pairs of nucleotides. This number is found to vary between the corresponding regions of other Gram positive eubacteria (Figure 12). Nine and ten base pairs are found in the corresponding regions of *B. subtilis* and *M. bovis*, respectively whereas the proposed stem structure of *S. ambofaciens* contains eleven base pairs and has an internal bulge. Interestingly, the length of helix 10 is not constant among the five *Frankia* strains; they contain ten or twelve pairs of nucleotides (Figure 13). This size variation depends on the presence of a guanosine or a cytosine residue at position 164. This position is indicated with an arrow in Figures 12 and 13. Like *M. bovis*, the two *Frankia* strains isolated from *Alnus crispa*, AcN14a and AcNI1, as well as ORS020606 (not shown) have a guanosine residue at this position. The 16S rRNAs of the three other *Frankia* strains and *Streptomyces* are capable of forming a longer hairpin structure.

Interestingly, the nucleotide composition of helix 10 varies between the six *Frankia* strains. Among the 27 nucleotides of this helix, there are 8 variable positions. These positions are not distributed throughout out the helix, but at the contrary, are limited to the region corresponding to the first 13 nucleotides. The variable positions are indicated with asterisks in Figure 13.

The secondary structure proposed for helix 10 of *Frankia* AcN14a is supported by several compensatory nucleotide substitutions present in the corresponding regions of *S. ambofaciens* and *M. bovis*. In contrast, all

nucleotide substitutions observed in helix 10 of the six *Frankia* strains are neutral since in all cases, a second substitution is not required to maintain the base-pairing. Helix 10 does not have to contain a perfect stem since a mismatch is observed in these regions of *Frankia* AgB1.9 and *Steptomyces* molecules (Pernodet *et al.*, 1989; Baylis and Bibb, 1987).

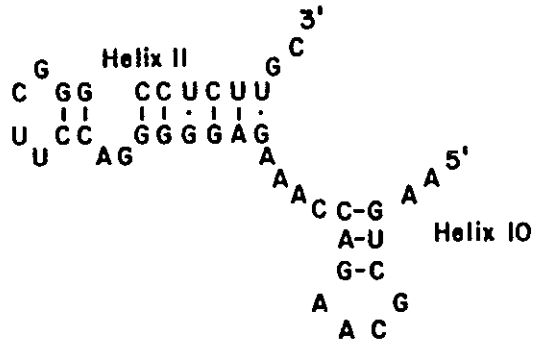
In contrast to the stem structure, the nucleotide composition (GCAU) of the terminal loop of helix 10 appears to be reasonably well conserved in members of these four genera. In addition, the adenosine triplet between helices 10 and 11 is conserved in most of these Gram positive eubacterial sequences as well as in the corresponding region of *E. coli*; however, the number of nucleotides in this bulge varies.

The nucleotide composition of the second helix, helix 11, is identical in the four *Frankia* strains (AcN14a, Ag45/Mut15, AgB1.9 and ORS020606). In contrast to the extended version of helix 10 proposed for *B. subtilis*, a secondary structure similar to that proposed for *Frankia* has been proposed for *S. ambofaciens* 16S rRNA. In addition, the corresponding sequence of *M. bovis* can also be folded into a similar structure. Compensatory nucleotide substitutions observed in the sequences of *M. bovis* and *S. ambofaciens* support the structure proposed for *Frankia* AcN14a.

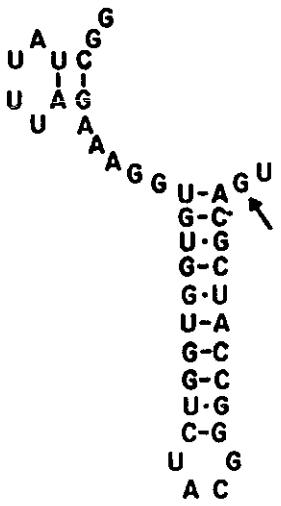
Similar secondary structures have been proposed for the corresponding regions of other eubacteria (Woese, 1987) and archaebacteria (Haas *et al.*, 1990; Neefs *et al.*, 1991). The extended version of helix 10 is common in the eubacterial kingdom. It is observed in all three subdivisions ($\alpha$, $\beta$ and $\gamma$) of purple bacteria, in cyanobacteria and in planctomyces (Woese, 1987). In contrast, the shortened version is less commonly found. For this reason, the extended version of helix 10 is likely to be the structure present in the ancestral organism. The shortened version of helix 11 is more common among eubacteria and archaebacteria and probably represents the ancestral form (Woese, 1987).

**FIGURE 12 PROPOSED SECONDARY STRUCTURES OF HELICES 10 AND 11 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED EUBACTERIA**

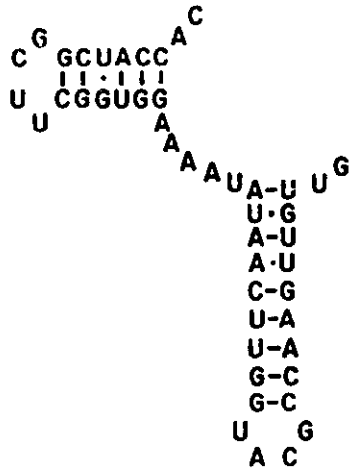
*Frankia* AcN14a 16S rRNA nucleotide sequence corresponding to positions 166 to 206 (see Figure 10) as well as the corresponding regions of *B. subtilis*, and *M. bovis* 16S rRNAs were folded with the help of the program RNAFOLD. The secondary structures proposed for the corresponding regions of *E. coli* and *S. ambofaciens* have been included as references (Neefs *et al.*, 1991; Pernodet *et al.*, 1989). The arrow indicates residue 164 discussed in text.



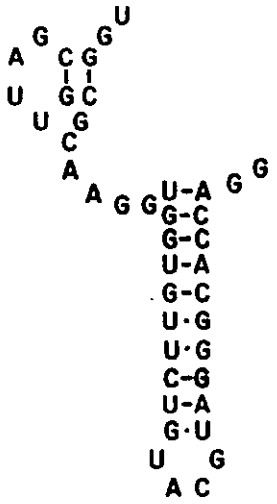

E. coli



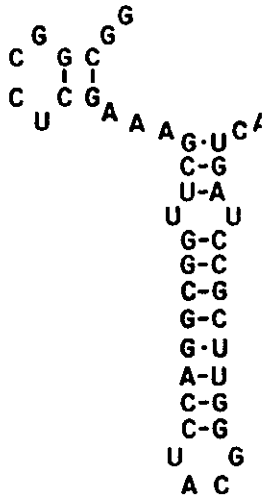
F. sp. AcN14a



B. subtilis



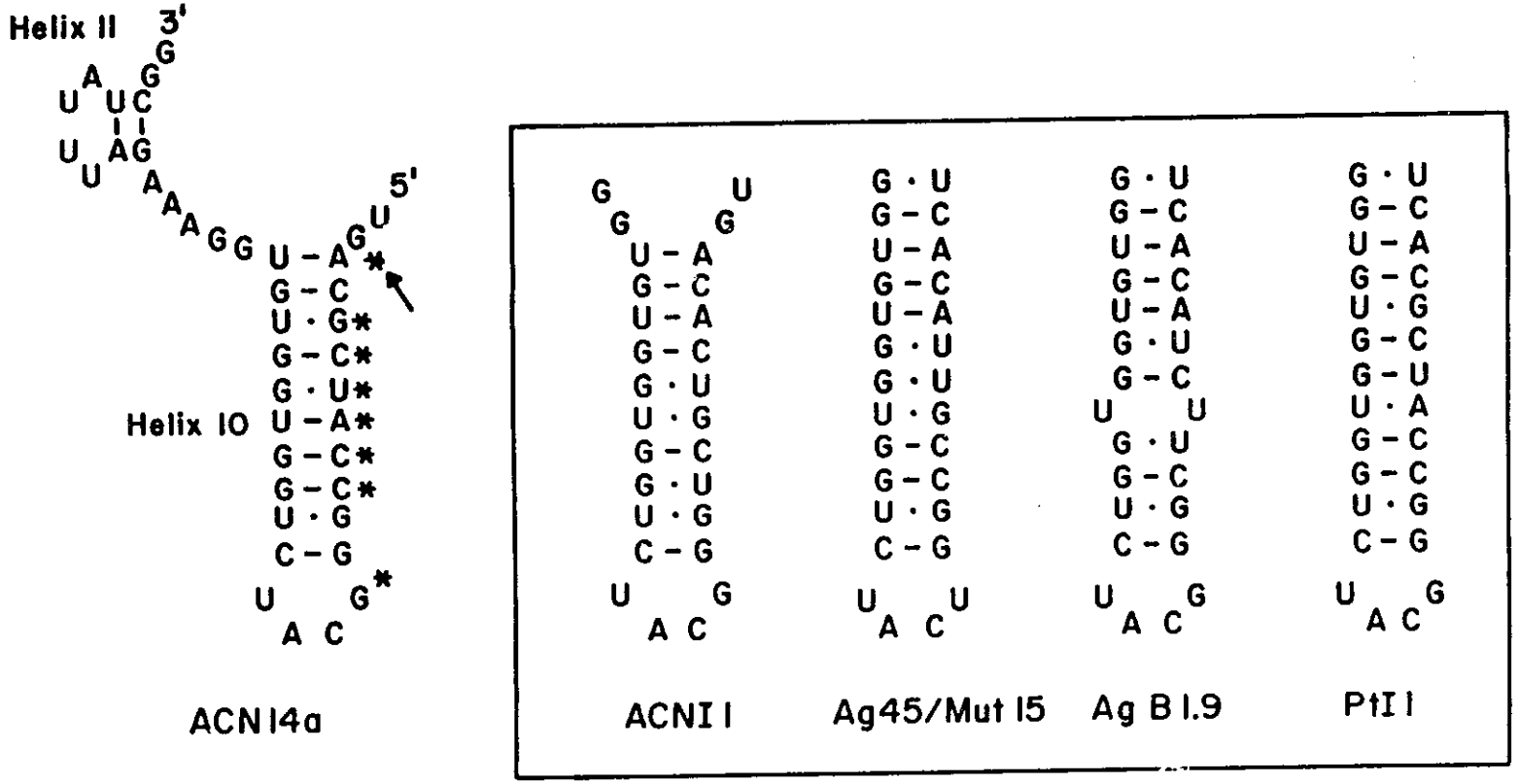
M. bovis



S. ambofaciens

**FIGURE 13 PROPOSED SECONDARY STRUCTURES OF HELICES 10 AND 11 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED SPECIES OF *FRANKIA*.**

*Frankia* AcN14a 16S rRNA nucleotide sequence corresponding to positions 166 to 206 (see Figure 10) as well as the regions corresponding to helix 10 of four other strains are compared. The secondary structures drawn for the strains Ag45/Mut15 and AgB1.9 have been proposed by Hahn *et al.* (1989a). The corresponding regions of the three other strains were folded with the help of the computer program RNAFOLD. The arrow and asterisks indicate residues discussed in text. *Frankia* strain ORS020606 differs from AcN14a by having an adenosine, a thymidine and a guanosine residues at positions 170, 171 and 173, respectively.


Helix II
3'
5'
Helix IO
ACN14a
ACNI1
Ag45/Mut15
Ag B1.9
Pt11

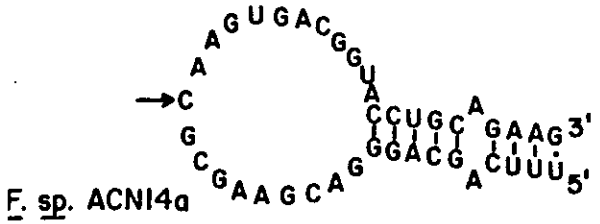
Thus, the region of *Frankia* AcN14a 16S rRNA which encompass helices 10 and 11 has been found to differ from that of *E. coli*. The corresponding regions of other Gram positive eubacteria such as *M. bovis* and *S. ambofaciens* may also be folded into similar secondary structures but minor variations are also observed. Variations were also observed between the corresponding regions of different *Frankia* strains. Although helices 10 of the two strains isolated from the same host plant species, AcN14a and AcNI1, as well as the strain isolated from *Casuarina equisetifolia*, ORS020606 were found to share the same number of base pairs, the primary sequences in this region of AcN14a shares more homology with that of PtI1 isolated from *Purshia tridentata* (Baker, 1987). These results contrast with the phylogenetic tree constructed by Nazaret *et al.* (1991) in which the strain PtI1 is shown to branch far apart from all the other strains studied. For their study, nucleotide sequence of domain III of *Frankia* 16S rDNAs were compared.

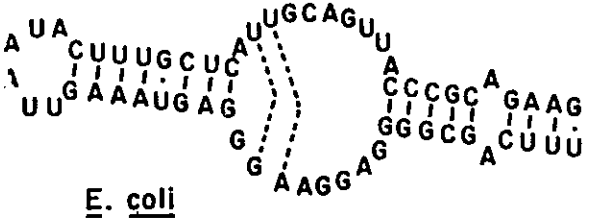
The third region found to differ between *E. coli* and *Frankia* AcN14a is located in helix 18 around nucleotide 441 in *Frankia* 16S rRNA. Most of the long terminal hairpin structure present in *E.coli* is deleted in *Frankia* AcN14a and in the other four *Frankia* strains (Figure 14). The nucleotide sequence of this helix is conserved in members of the genus *Frankia*. Only one or two substitutions (positions 441 and 442) are observed between the different *Frankia* strains. AcN14a, AcNI1 and PtI1 have a cytosine residue at position 441 whereas Ag45/Mut15, AgB1.9 and ORS020606 have an adenosine residue. This residue is indicated by an arrow in Figure 14. In addition, ORS020606 differs from all other strains; it has a guanosine residue at position 442, whereas other *Frankia* strains have an adenosine residue at the corresponding position.

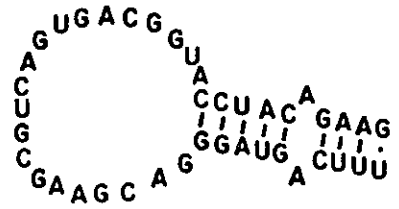
The shortened version of helix 18 is not unique to *Frankia*. The terminal stem structure of this helix is also absent in *Streptomyces* (Pernodet *et al.*, 1989; Baylis and Bibb, 1987). The nucleotide sequence of this region is highly conserved between the two genera, it varies at only two positions (positions 433 and 441).

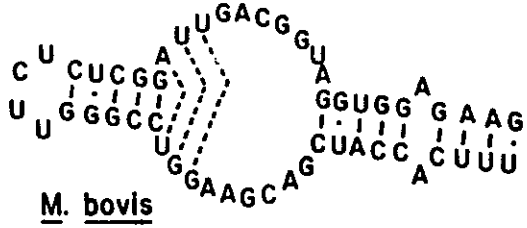
**FIGURE 14 PROPOSED SECONDARY STRUCTURES OF HELIX 18 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED EUBACTERIA**

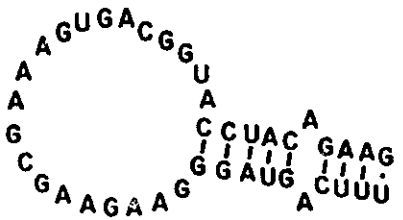
*Frankia* AcN14a 16S rRNA secondary structure of region corresponding to positions 422 to 462 (see Figure 10) and the corresponding regions of five other eubacterial species are presented. The secondary structures for *E. coli* (Neefs *et al.*, 1991), *S. ambofaciens* (Pernodet *et al.*, 1989) and the two *Mycobacteria* species (Stackebrandt and Smida, 1988) have been included as references. The corresponding regions of *Frankia* and *P. thermophila* were folded with the help of the computer program RNAFOLD. The arrow indicates residue 441 discussed in text.

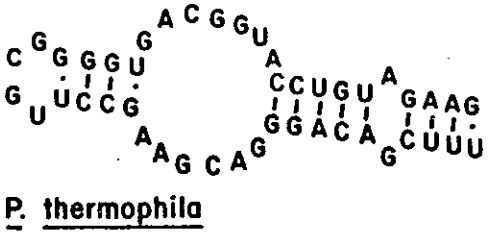

F. sp. ACN14a


E. coli


M. phlei


M. bovis


S. ambofaciens


P. thermophila

Besides *E. coli*, other eubacteria such as *B. subtilis* and some slow growing *Mycobacteria* species, such as *M. bovis* can form a similar hairpin structure. In contrast, the fast growing *Mycobacteria* species exhibit an extended version of this helix (Stackebrandt and Smida, 1988). Interestingly, the formation of hairpin structures of an intermediate length are possible in the corresponding regions of other actinomycete 16S rRNAs, such as *Pseudonocardia thermophila* and *Saccharopolyspora hirsuta* (Embley *et al.*, 1988). Thus, the length of the terminal structure appears to be highly variable among actinomycetes.

In those organisms that exhibit the extended version of helix 18, the nucleotide composition of the entire terminal hairpin structure varies. In contrast, the proximal portion of helix 18 appears quite conserved among the six eubacterial species. The proximal stem structure is not always perfect but generally contains ten paired nucleotides. Sequence conservation is also observed in the adjacent loop.

Helix 18 is quite variable in the eubacterial kingdom. In the chloroplast 16S rRNAs, the inner stem is deleted (Woese *et al.*, 1983). In addition, helix 18 appears to be differently structured in members of the different kingdoms (Woese *et al.*, 1983; Neefs *et al.*, 1991). In *Halobacterium cutirubrum*, this helix consists of a series of 12 base paired nucleotides and a very small terminal loop. In the model proposed for the eukaryotic small rRNAs, helix 18 is short, imperfect and its terminal loop comprises only four nucleotides (Neefs *et al.*, 1991).

Thus, the primary and secondary structures of helix 18 in *Frankia* AcN14a and other strains resemble those reported for *Streptomyces* (Pernodet *et al.*, 1989; Baylis and Bibb, 1987) and some species of *Mycobacteria* (Stackebrandt and Smida, 1988). The regions corresponding to helix 18 in these organisms was found to vary from that of *E. coli* ; a terminal helix present in *E. coli* has been deleted in these actinomycete species. The size of this putative helix appears to be variable among actinomycetes. Members of the genera *Frankia*, *Streptomyces* and some *Mycobacteria* species do not have the terminal

stem structure whereas some other *Mycobacteria* species have structures similar to that of *E. coli*. In addition, other actinomycetes such as *P. thermophila* and *S. hirsuta* can form short terminal stem structures (Embley *et al.*, 1988).

The three regions that have been discussed so far were first chosen because most of the size difference observed between the 16S rRNAs of *Frankia* AcN14a and *E. coli* could be attributed to changes in the number of nucleotides present in these areas. In these regions, *Frankia* AcN14a sequences could be folded into secondary structures similar to those proposed for *S. ambofaciens* (Pernodet *et al.*, 1989). The primary structure of one of these regions, the region corresponding to helix 10 was found to vary among different *Frankia* strains and could potentially be useful for their identification.

In contrast, the next two variable regions are located in domain III and will be described for different reasons. The fourth and fifth variable regions were chosen because of differences in primary and secondary structures respectively, observed between *Frankia* AcN14a and *S. ambofaciens*. Recently, Nazaret *et al.* (1991) have established the phylogenetic relationships among different *Frankia* strains and most genomic species on the basis of the nucleotide sequence of domain III of their 16S rDNA. Most of the variable positions were localized to these same two regions.

The fourth region of interest corresponds to nucleotides 964 to 1008 in *Frankia* AcN14a 16S rRNA. The nucleotide sequence of this region can form a composite helix in eubacteria whereas a simple helix is observed in archaebacteria and eukaryotes (Neefs *et al.*, 1991). The eubacterial structure is subdivided into three substructures: helices 35, P35-1 and P35-2. This region has been described as highly variable by Woese *et al.* (1983). In agreement with this statement, many substitutions are observed between distantly related species such as *Frankia* and *E. coli* as well as between the various *Frankia* strains (Figures 16 and 17).

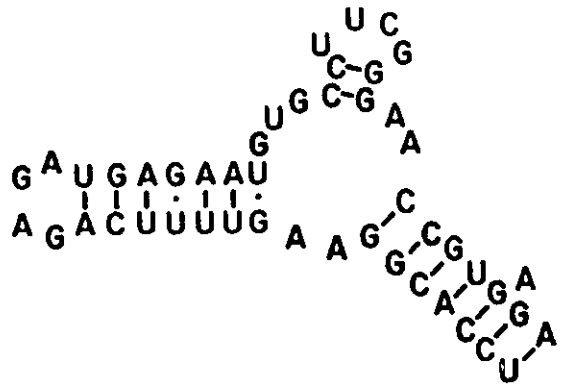
Primary structure comparison of *Frankia* AcN14a sequence to those of selected *Frankia* strains (Nazaret *et al.*, 1991, Hahn *et al.*, 1989a) has allowed identification of twenty-two variable positions. At least seven sequence patterns (A to G) could be identified for this region. The different *Frankia* genomic species can be categorized into three groups according to their sequence patterns (patterns A, F and G). In addition, the sequence of other *Frankia* strains have been reported to have unique sequence patterns. *Frankia* AcN14a sequence pattern (pattern A) is identical to that reported for the corresponding regions of several other strains comprised in genomic species 1 and 9 (Figure 16).

In spite of the variation in their primary structures, the secondary structures of the composite helices is conserved among the different *Frankia* strains. This helix complex generally contains 45 nucleotides. The strains AgB1.9 and Ag45/Mut15 are exceptions; they respectively have 44 and 43 nucleotides. In strain Ag45/Mut15, one pair of nucleotides is absent from helix 35 and accounts for the size difference. The corresponding pair of nucleotides of *Frankia* AcN14a is boxed in Figure 15. In contrast, only one nucleotide is missing in the corresponding helix of the strain AgB1.9 and two additional nucleotides are present in helix P35-2. The position corresponding to the deletion and addition in AgB1.9 are indicated respectively, with a star and a triangle in the secondary structure proposed for *Frankia* AcN14a (Figure 15).

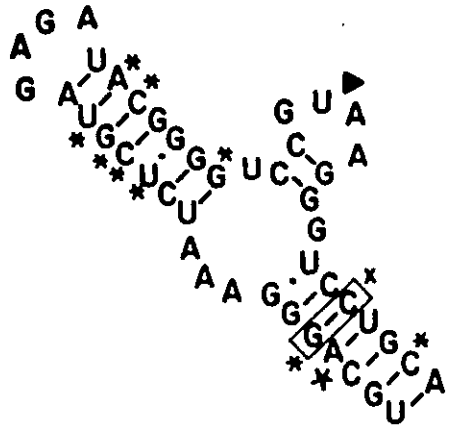
The secondary structures proposed for *E. coli*, *M. bovis*, *Streptomyces* and *Frankia* are similar. Most of the nucleotide changes are compensatory in nature or located in unpaired regions and as a result, the main difference between the three secondary structures presented in Figure 15 is the size of the internal loop regions. In these organisms, six or seven pairs of nucleotides (and possibly eight for *M. bovis*) may be used to form helix 35. In *E. coli* and in the two Gram positive organisms with a low G+C content (*B. subtilis* and *Mycoplasma pinum*) an unpaired nucleotide is present near the base of this stem. Helix P35-1 generally contains seven to eight pairs of nucleotides. In contrast, helix P35-2 is shorter. Conserved features of this region are the nucleotide composition of the terminal loop (GAGA) and the last pair of

**FIGURE 15 PROPOSED SECONDARY STRUCTURES OF THE COMPOSITE HELIX 35 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED EUBACTERIA**

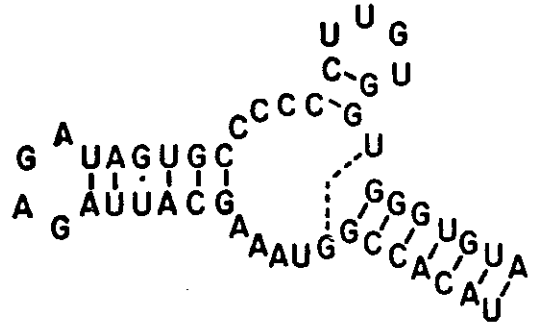
*Frankia* AcN14a 16S rRNA secondary structure of region corresponding to positions 964 to 1008 (see Figure 10) is presented. The corresponding regions of *E. coli* (Neefs *et al.*, 1991) and *S. ambofaciens* (Pernodet *et al.*, 1989) have been included as references. The nucleotide sequence of *Frankia* AcN14a was folded into a secondary structure with the help of the computer program RNAFOLD. The star, triangle, asterisks and box indicate positions discussed in text.



E. coli



F. sp. ACN14a



S. ambofaciens

**FIGURE 16 COMPARISON OF THE PRIMARY STRUCTURES OF THE COMPOSITE HELIX 35 OF VARIOUS *FRANKIA* STRAINS**

Aligned sequences of selected 16S rRNA patterns observed in the composite helix P35 of various *Frankia* strains. In this region, *Frankia* AcN14a sequence is identical to those of genomic species 1 (*F. alni*) and 9, the latter being represented by ORS020606. Only the nucleotides that differ from those of *Frankia* AcN14a are shown. Identities are denoted by dots, and deletions are marked by hyphens. This figure is an adapted version of that published by Nazaret *et al.*, (1991). The AgB1.9 sequence has been reported by Hahn *et al.* (1989a).

| PATTERNS | STRAINS | GENOMIC SPECIES | 968 | | | | 1008 |
|---|---|---|---|---|---|---|---|
| | | | UGCAGGG | AAAUCUCGUA | GAGAUACGGG | GUCCGU-AA- | GGGUCCUGCA |
| A | AcN14a, ORS020606 | 1, 9 | | | | | |
| B | Ag45/Mut15 | | ....-.. | ....UCUCC. | .....GG... | A..... | .....-..U. |
| C | AgB1.9 | | ...-... | .....CAUC. | .....GGU.. | ....U.U..A | ......C.U. |
| D | M2 | | ..... | .....C.... | ..... | ..... | ..... |
| E | PtI1 | | ...G... | .....CAUC. | .....GGU.. | ...U....G. | ......C.UG |
| F | | 2, 3 | ..... | .....CUCC. | .....GG... | ..... | ..... |
| G | | 4, 5, 6, 7 | .....A. | .....CU... | ......U... | ...U....G. | ...CU..... |

nucleotides (A-U) of the P35-1 helix as well as the two pairs of nucleotides of the P35-2 helix.

Thus, in this region corresponding to the helix 35 complex, considerable primary structure variation is observed among eubacteria. Their secondary structures are however conserved. Twenty-two variable positions present in this region allow categorization of selected *Frankia* strains into seven groups. Most of the *Frankia* genomic species could be divided into three groups and individual strains were found to have unique patterns. On the basis of helix 35 complex nucleotide sequence, *Frankia* AcN14a has been placed in the same group as genomic species 1 and 9.

The last region found to vary between the different eubacterial 16S rRNAs is located in helix 41 (Neefs *et al.*, 1991). It corresponds to nucleotides 1082 to 1110 in *Frankia* AcN14a 16S rRNA sequence. The primary and the secondary structures of the proximal portion of this helix are conserved between distantly related eubacteria; it comprises seven pairs of nucleotides (Figure 17). The nucleotide composition of the middle region is also reasonably well conserved. The nucleotide sequence and the secondary structure of the terminal hairpin structure are for their part, more variable. The length of the hairpin structure observed in *Frankia* is comparable to that of *E. coli* and *B. subtilis*. However, this structure is absent from *Mycoplasma pinum* 16S rRNA, whereas a longer version is observed in *Streptomyces*. Surprisingly, the sequences of *M. bovis* and *Pseudonocardia sp.* are identical in this region (Figure 17).

Substantial variation was reported in the terminal loop and the preceding stem of different *Frankia* strains (Nazaret *et al.*, 1991). Of the eight nucleotides (positions 1098 to 1105) present in the terminal loop and in the last pair of nucleotides in the stem structure, seven were found to vary between various *Frankia* strains, allowing identification of ten sequence patterns (Figure 18). This region is more informative than the helix P35 complex since most genomic species have an unique sequence pattern. According to its nucleotide sequence, *Frankia* AcN14a has been placed into genomic species 1.

**FIGURE 17 PROPOSED SECONDARY STRUCTURES OF HELIX 41 WITHIN THE 16S RIBOSOMAL RNAS OF SELECTED EUBACTERIA**

*Frankia* AcN14a 16S rRNA secondary structure of region corresponding to positions 1082 to 1110 (see Figure 10) is presented. The corresponding regions of *E. coli* (Neefs *et al.*, 1991) and *S. ambofaciens* (Pernodet *et al.*, 1989) have been included as references. The nucleotide sequence of *Frankia* AcN14a was folded into a secondary structure with the help of the computer program RNAFOLD. Asterisks indicate positions found to vary between AcN14a and ORS020606.

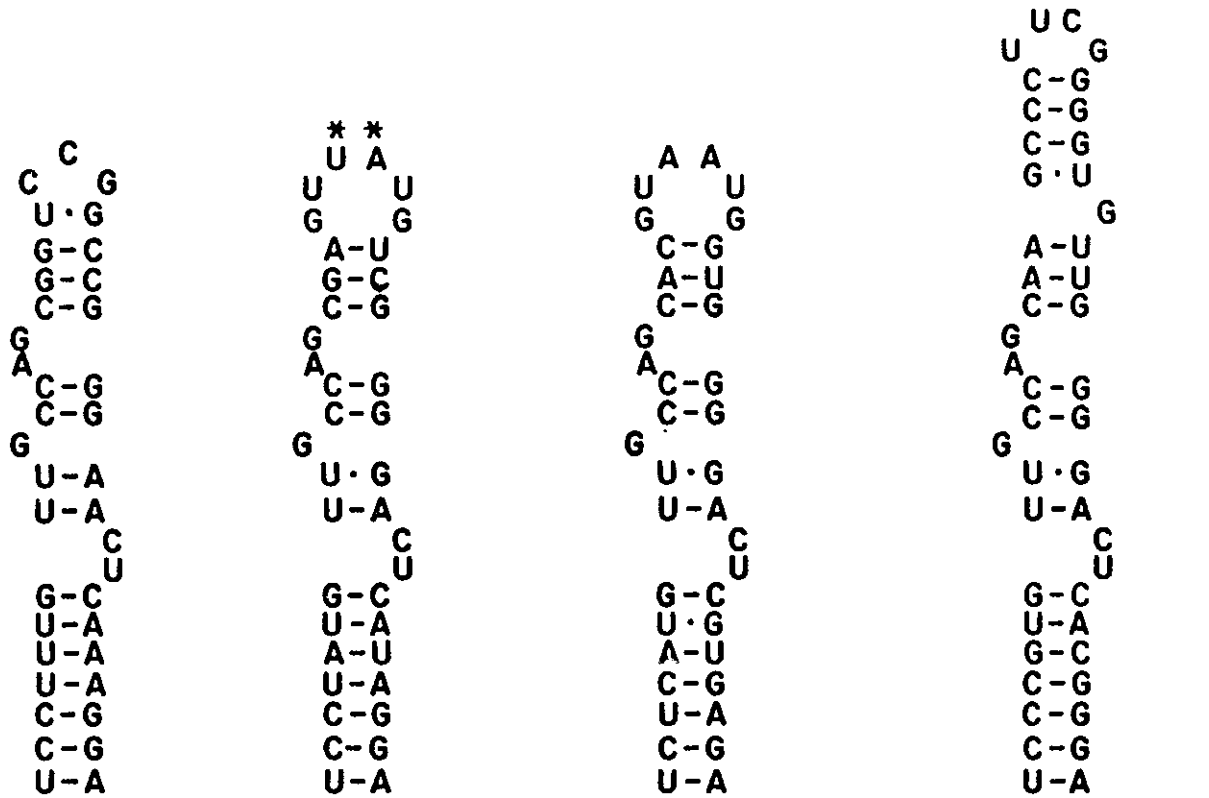

E. coli
F. sp. ACN14a
M. bovis
S. ambofaciens

**FIGURE 18 COMPARISON OF THE PRIMARY STRUCTURES OF THE HELIX 41 OF VARIOUS *FRANKIA* STRAINS**

Aligned sequences of selected 16S rDNA patterns observed in helix 41 of various *Frankia* strains. *Frankia* AcN14a sequence is identical to that of genomic species 1 (also called *F. alni* ). Only the nucleotides that differ from those of *Frankia* AcN14a are shown. Identities are denoted by dots, and deletions are marked by hyphens. Ambiguous nucleotide positions are denoted by n. This figure is an adapted version of that published by Nazaret *et al.*, (1991).

| PATTERNS | STRAINS | GENOMIC SPECIES * | 1081 | | | 1109 |
|---|---|---|---|---|---|---|
| A | AcN14a | 1 (*F. alni* ) | UCCUAUGUUG | CCAGCGAGUU | AUGUCGGGGA | CUCAUAGGA |
| B | Ag45/Mut15 | | .......... | .....-...C | ....n..... | ......... |
| C | Col | | .......... | ......C.G. | ...G...... | ......... |
| D | CH | | .......... | ......C..C | ...G...... | ......... |
| E | D11 | | .......... | ......C... | ...G...... | ......... |
| F | | 2 | .......... | .........A | .......... | ......... |
| G | | 3 | .......... | .........C | GA........ | ......... |
| H | | 4,5 | .......... | .........C | .......... | ......... |
| I | | 7 | .......... | ......C..A | UA.G...... | ......... |
| J | ORS020606 | 9 | .......... | .........C | G......... | ......... |

In addition, unique patterns have been reported for some other strains.

Thus, the terminal portion of helix 41 is not well conserved among eubacteria nor it is among members of the genus *Frankia*. At least ten *Frankia* sequence patterns could be identified. Interestingly, most *Frankia* genomic species have a specific sequence pattern in this region of their 16S rRNAs. *Frankia* AcN14a was grouped with genomic species 1.

In summary, comparison of *Frankia* AcN14a 16S rRNA to that of other eubacteria has allowed the identification of three relatively small deletions of fourteen to twenty nucleotides each and one addition of fifteen nucleotides which are all located in domain I. Although the number of nucleotides and the secondary structures of these regions were found to be conserved among members of the genus *Frankia*, variability at the primary structure level was observed. In addition, two other regions previously defined as variable among *Frankia* strains have been described. The latter are located in domain III of the 16S rRNA molecule and correspond to helices 35(P35-1 and P35-2) and 41.

We have PCR amplified, cloned and sequenced part of both copies of the 16S rRNA genes (positions 851 to 1175) present in two genomic clones. Since the two copies could not be differentiated, we expect both rRNA genes of *Frankia* AcN14a to be undistinguishable.

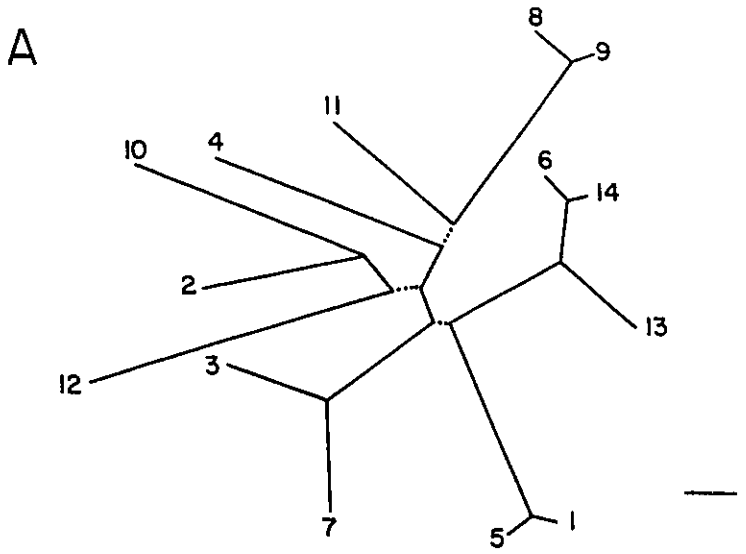
Partial 16S rRNA sequences are often used to construct phylogenetic trees (e.g. Embley *et al.*, 1988). We have compared a tree based upon four of the five variable regions analyzed to a tree based upon complete actinomycete 16S rRNA sequences. Data used for the construction of the trees are presented in Appendices 2 and 3. The variable regions selected for the construction of the unrooted tree are those corresponding to helices 6, 10, P35 and 41. Both unrooted trees are presented in Figure 19. Although differences are observed between the two trees, both trees cluster the different species of a given genus. Members of the genera *Frankia* (#1 and #5), *Streptomyces* (#6, #13 and #14) and *Mycobacterium* (#8 and #9), form three separated groups. A close relationship is shown for *Arthrobacter*

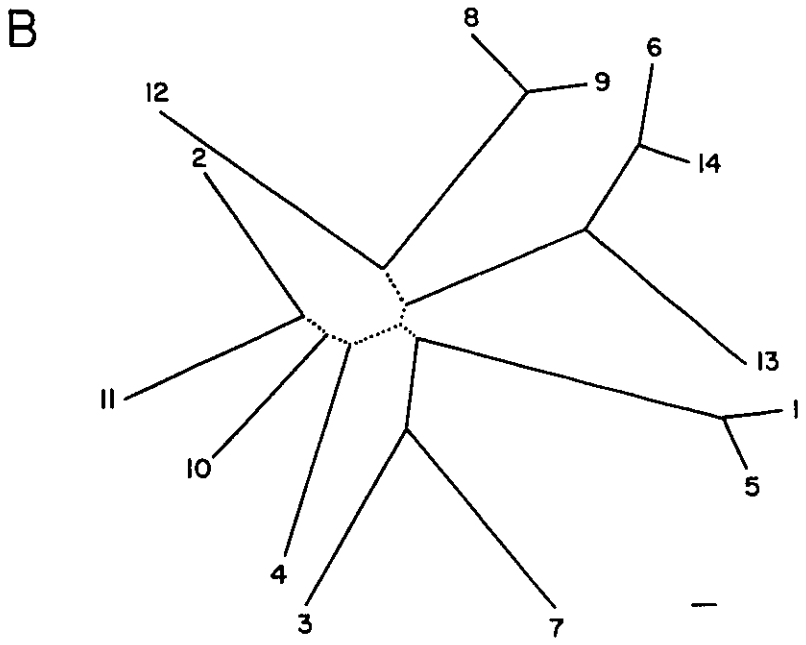
**FIGURE 19 UNROOTED PHYLOGENETIC TREES SHOWING THE INTERGENERIC RELATIONSHIPS OF *FRANKIA* AND SELECTED ACTINOMYCETE GENERA**

Unrooted phylogenetical trees based upon **A.** complete 16S rRNA sequences and **B.** four variable regions of the 16S rRNAs. The actinomycetes included in these trees are:

(1) *Frankia sp.* AcN14a
(2) *Aureobacterium erythreus*
(3) *Arthrobacter globiformis*
(4) *Corynebacterium variabilis*
(5) *Frankia sp.* ORS020606
(6) *Streptomyces ambofaciens*
(7) *Micrococcus luteus*
(8) *Mycobacterium kansasii*
(9) *Mycobacterium bovis*
(10) *Nocardioides luteus*
(11) *Nocardia asteroides*
(12) *Propionobacterium freudenreichii*
(13) *Streptomyces griseus*
(14) *Streptomyces lividans*

- indicates over 88% occurrence in a bootstrap confidence limits test. •• indicates less than 88% occurrences in a bootstrap confidence limits test. The bar at the bottom of each tree indicates distance representing 0.01 substitution/site.


A
8
9
11
10
4
6
14
2
13
12
3
7
5
1


B
8
6
9
12
14
2
13
11
1
10
5
4
3
7

(#3) and *Micrococcus* (#7). Because of the nature of the sequences analyzed, the branching of the tree based upon the variable regions is longer than that of the tree based upon complete rRNA sequences. This reflects the different rates at which the various portions of the 16S rRNAs change; the mutation rate being higher in variable regions. The relationships between the genera are different in the two trees. Interestingly, the tree based upon complete 16S rRNA sequences places the mycolic acid-containing taxa, *Mycobacterium* (#8 and #9), *Nocardia* (#11) and *Corynebacterium* (#4), together.

### Analysis of the 23S rRNA gene - primary and secondary structures

The predicted length of *Frankia* AcN14a 23S rRNA is 3105 nucleotides which is 200 nucleotides longer than that of *E. coli*. *Frankia* AcN14a 23S rRNA is relatively large compared to other prokaryotic species. Interestingly, the longest eubacterial 23S rRNAs to have been entirely sequenced, are those of three other actinomycetes, *S. ambofaciens* (Pernodet *et al.*, 1989), *Frankia* strain ORS020606 (Normand, unpublished) and *Micrococcus luteus* (Regensburger *et al.*, 1988) which are 3120, 3099 and 3094 nucleotides in length respectively. A comparable length has also been reported for the archaebacterium *Desulfurococcus mobilis* 23S rRNA (3077 nucleotides, Leffers *et al.*, 1987). In contrast, the eukaryotic cytoplasmic species are much longer, having 3377 to 5025 nucleotides, whereas mitochondrial ones show extreme variations with sizes ranging from 1141 to 3546 nucleotides (reviewed by Raué *et al.*, 1988).

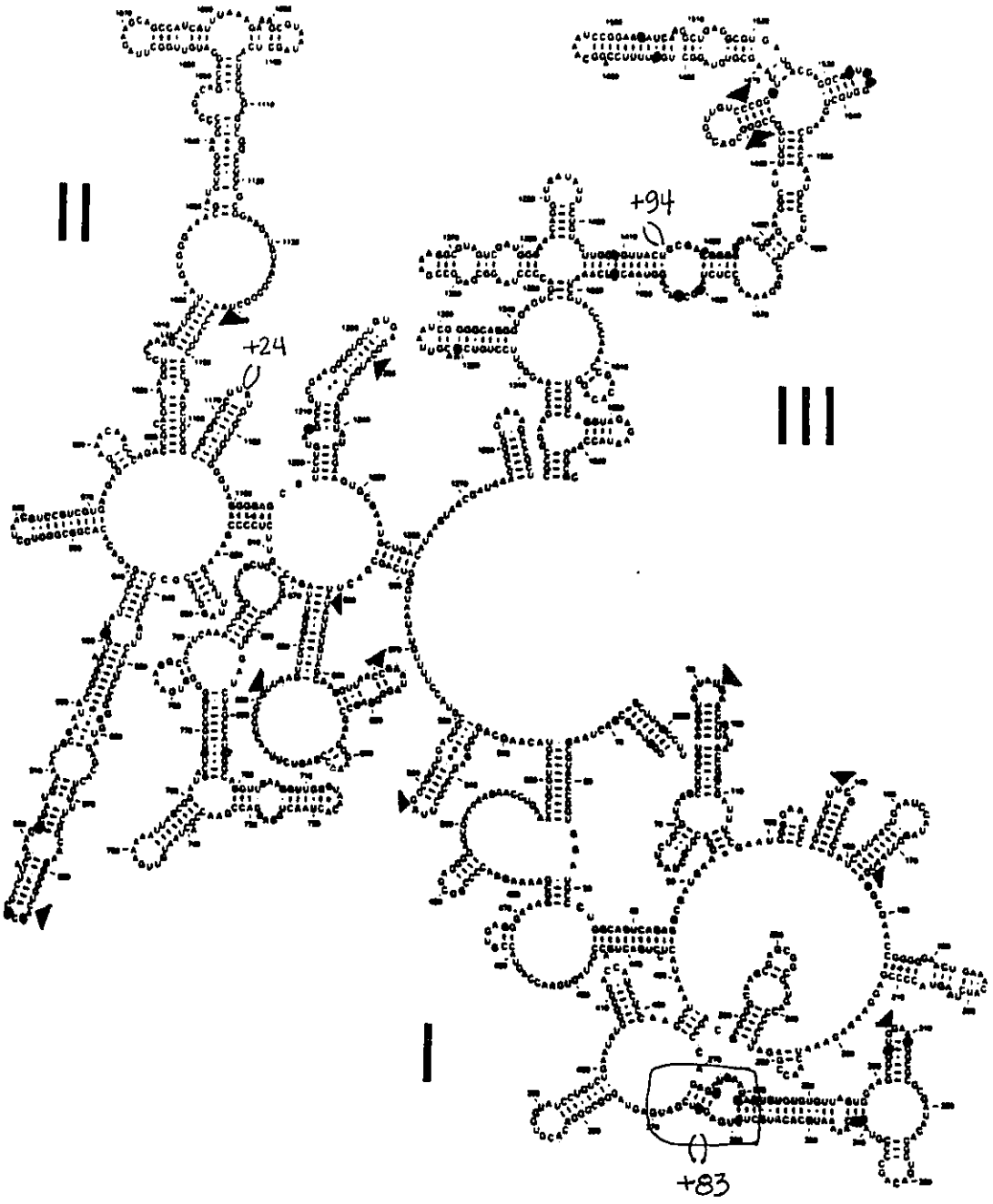
*Frankia* AcN14a 23S rRNA sequence was compared to other eubacterial sequences. Its sequence is 98% similar to that of the ORS020606 strain and shares 85% and 86% identity with the corresponding molecules of *M. luteus* and *S. ambofaciens* respectively and it is 80% similar to *E. coli* 23S rRNA sequence. Figure 20 illustrates the differences observed between the primary and secondary structures of *Frankia* AcN14a and *E. coli* 23S rRNAs. Like the two other actinomycete species, the size difference in the *Frankia* AcN14a and

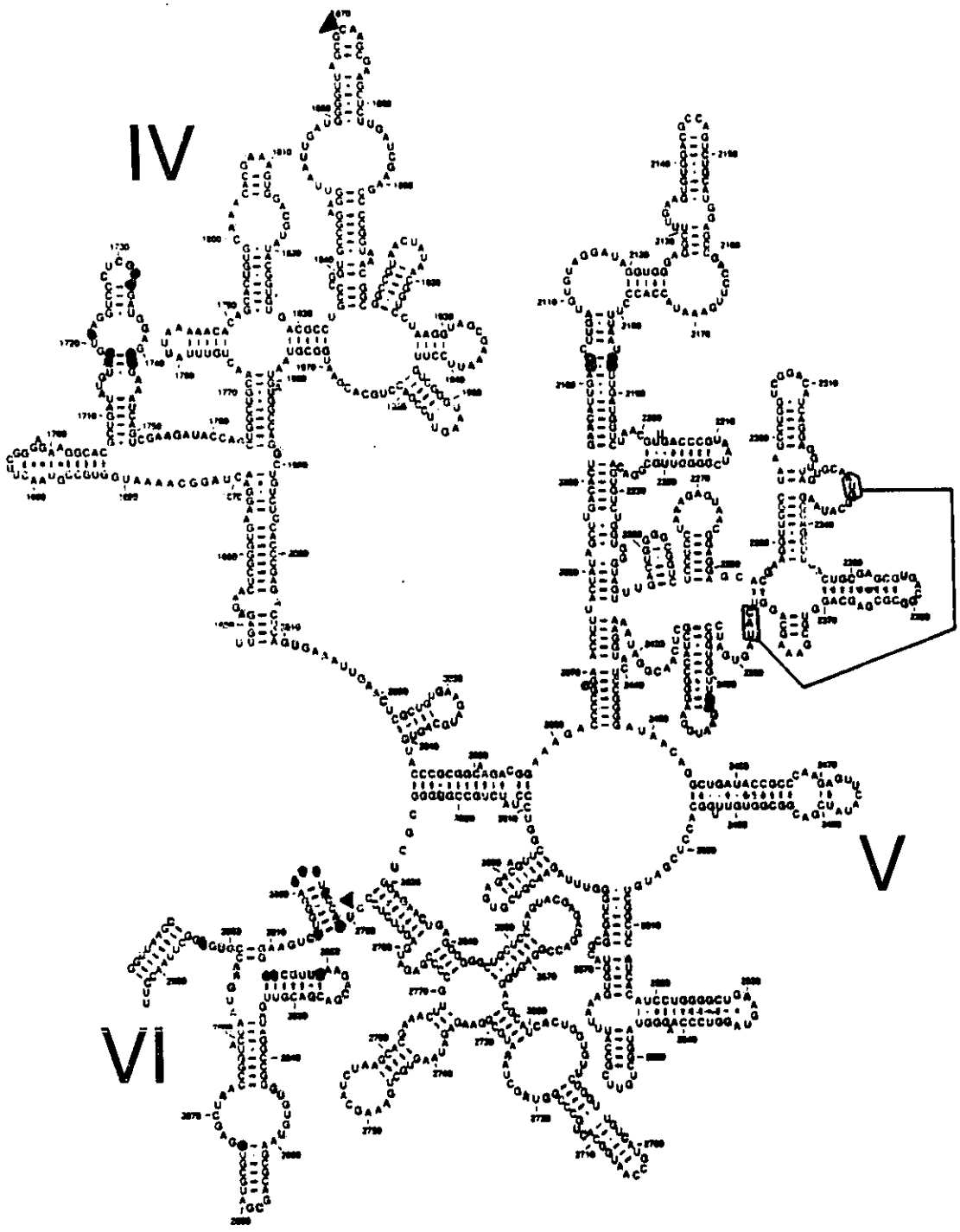
**FIGURE 20 PRIMARY AND SECONDARY STRUCTURE VARIATION BETWEEN *FRANKIA* ACN14A AND *E. COLI* 23S RIBOSOMAL RNAS.**

*Frankia* AcN14a 23S rRNA sequence deduced from the corresponding gene sequence is compared to that of *E. coli* (Raué *et al.*, 1988). The positions conserved between the *Frankia* AcN14a and *E. coli* 23S rRNA sequences are indicated by the corresponding nucleotides and the positions that vary between these two organisms are circled. Helices are numbered in the order of occurrence from 5' to 3' terminus. For the *E. coli* sequence, every 10th position is marked; every 100th position is numbered when possible. Domains are indicated by roman numbers. The locations and size of the large insertions present in *Frankia* AcN14a sequence are indicated by thick lines. The smaller insertions, the deletions and the different substitutions are indicated as follow:

Color code

| | |
|---|---|
| yellow circle: | compensatory substitution |
| orange circle: | neutral substitution |
| blue circle: | noncompensatory substitution |
| green circle: | deletion |
| black triangle: | small insertion (1 to 4 nucleotides) |


II
III
I
+24
+94
+83


IV
V
VI

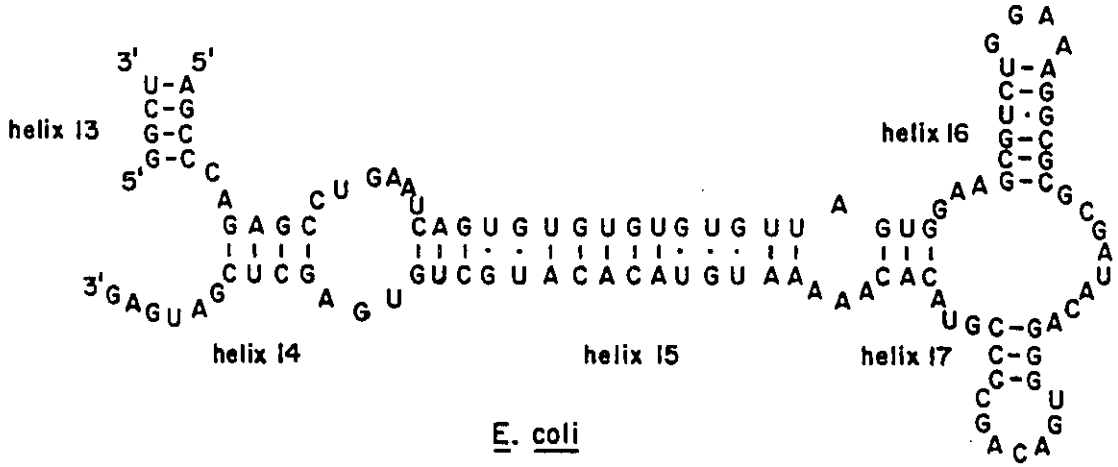
ORS020606 23S rRNA genes compared with that of *E.coli* is mostly due to two important insertions in the first half of the molecule. In addition, a shorter insertion segment, specific only to *Frankia* AcN14a 23S rRNA, is also observed in this portion of the rRNA molecule.

Interestingly, these three insertion segments are located in variable regions in which intervening sequences (IVS) have been found in some organisms. In these organisms, excision of intervening sequences during rRNA maturation results in fragmented large subunit rRNAs. This is in contrast to other cases in which the integrity of the rRNA molecule is restored by splicing after the excision of the intron. Such fragmentation of large subunit rRNA has been recognized in some chloroplasts, in eubacteria such as *Salmonella* and *Yersinia*, in insects, and in trypanosomes (reviewed by Raué *et al.*, 1988 and Pace and Burgin, 1989; Skurnik and Toivanen, 1991). *Frankia* AcN14a 23S rRNA is probably not fragmented since a single band corresponding to the full size molecule could be observed in *Frankia* AcN14a RNA preparation by gel electrophoresis (Figure 8).

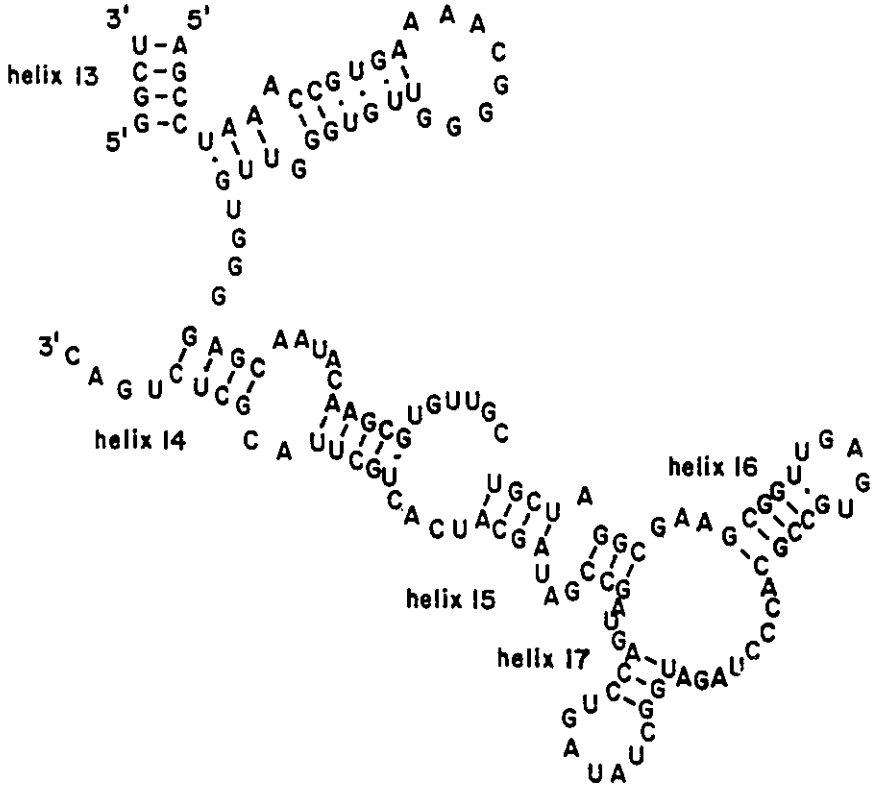
The first of the two large insertion segments is located in domain I, between helices 13 and 14 of *E. coli*. This region of *Frankia* AcN14a and ORS020606 23S rRNAs (nucleotides 273 to 456) contain 83 additional nucleotides in comparison to the corresponding region of *E. coli*. Similarly, 90 and 91 additional nucleotides are present in the same location in *S. ambofaciens* (Pernodet *et al.*, 1989) and *M. luteus* (Regensburgers *et al.*, 1988). The primary structure of the insertion varies between the different actinomycetes. The positions that vary between the two *Frankia* strains are indicated by asterisks in Figure 21. These sequences differ from each other by twelve substitutions, whereas a lot more are observed when the four actinomycete sequences are compared. Although the primary structure of this insertion diverges between species of these three actinomycete genera, analysis of *Frankia* and *M. luteus* sequences generally supports the folding model proposed for *S. ambofaciens*. As a result of this insertion, two additional helical structures can be formed in the bulge separating the corresponding helices 13 and 14 of *E. coli* (Figure 21). Base pairing of the first 49 nucleotides of the insertion segment

**FIGURE 21 PROPOSED SECONDARY STRUCTURE OF THE 23S RIBOSOMAL RNA REGION LOCATED IN DOMAIN I INCLUDING HELICES 13 TO 17.**

*Frankia* AcN14a 23S rRNA secondary structure proposed for region corresponding to positions 269 to 470 (see Figure 20) is presented. The corresponding regions of *E. coli* (Neefs *et al.*, 1991) and *Zea mays* (Gutell and Fox, 1988) have been included as references. The nucleotide sequence of *Frankia* AcN14a was folded into a secondary structure showing maximal similarity to *E. coli* and *S. ambofaciens* models with the help of the computer program RNAFOLD. Asterisks indicate position that vary between *Frankia* AcN14a and ORS020606.

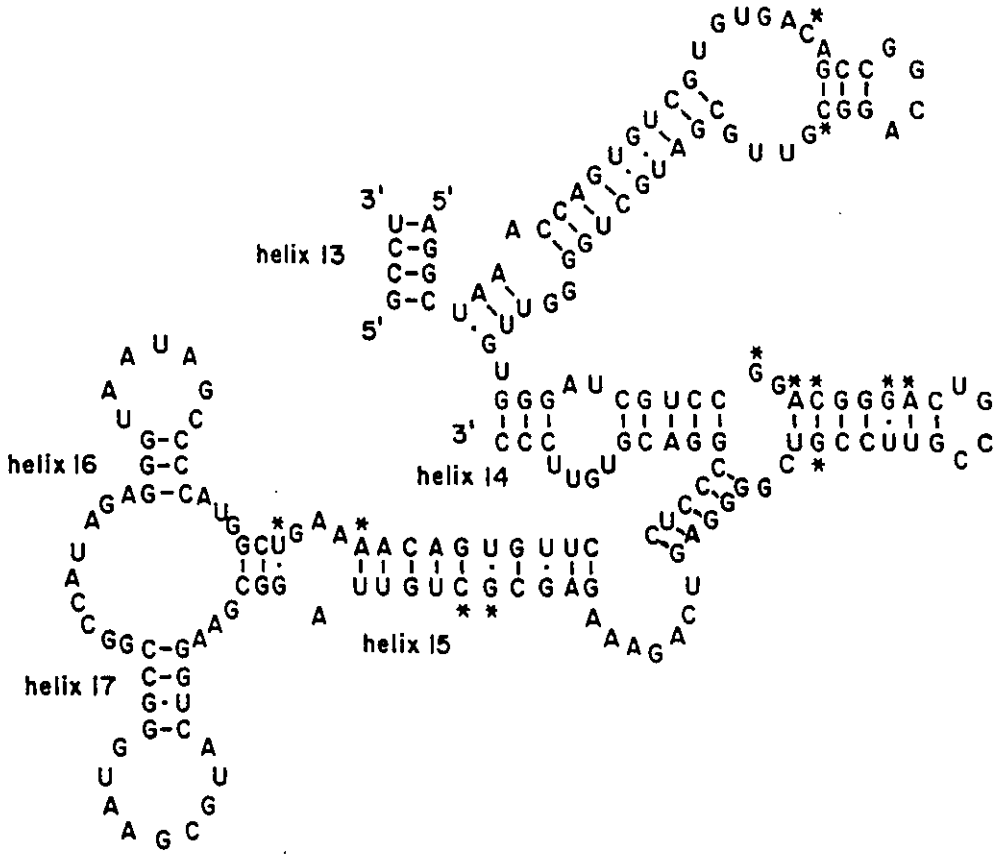



E. coli




A

Z. mays




B Frankia AcN14a

allows the formation of an imperfect helix (Figure 21). Apart from the middle part of the helix where many compensatory base substitutions are observed between *Frankia* AcN14a, *M. luteus* and *S. ambofaciens*, the remaining primary sequences are almost identical. The remaining 3' portion of the insertion segment can be folded into a second helical structure which enlarges and changes the relative position of the helix complex composed of helices 14, 15, 16 and 17 of *E. coli*. *Frankia* AcN14a primary sequence of this portion of the insertion segment and that of helix 15 differs greatly from those of the two other actinomycete species. The majority of the numerous substitutions observed between the two *Frankia* species and the other actinomycete species represent compensatory or neutral nucleotide changes and generally supports the secondary structure folding proposed for *S. ambofaciens* (Pernodet *et al.*, 1989).

The 23S rRNAs of many other eubacteria as well as those of archaebacteria and chloroplasts also contain additional nucleotides in this region (Gutell and Fox, 1988; Raué *et al.*, 1988; Höpfl *et al.*, 1989). In comparison to these three actinomycetes species, generally fewer nucleotides, between 30 and 45 are present in the eubacteria and chloroplast molecules. In these organisms, formation of a similar, but shorter helix is possible (Figure 21). Enlargements in the 23S rRNA of *Desulfurococcus mobilis* and the large cytoplasmic rRNA of eukaryotic organisms also occur in this region (Gutell and Fox, 1988; Raué *et al.*, 1988). The corresponding region of these organisms, can be folded in a more complex structure similar to that described for the actinomycete species.

Thus, compared to *E. coli*, most large subunit rRNA molecules contain additional nucleotides in this region. In some organisms, an insertion segment is located between helices 13 and 14 while in some others, such as *Frankia* AcN14a, an insertion segment is found and further size and shape variations are also observed for helices 14 and 15. The insertion segments within domain I of various actinomycete 23S rRNAs probably represent variants of the insertion sequence that was present in the large subunit rRNA molecules of their common ancestor. Even though the nucleotide sequences of the insertion

segments have been found to vary among the actinomycetes, their secondary structures are maintained.

The second large insertion segment of comparable size comprises 94 nucleotides and is located in domain III, in the bulge separating helices 50 and 51 of the *E. coli* molecule. This insertion segment has been shown to be present in both copies of the 23S rRNA gene (Gleeson, 1992) and corresponds to nucleotides 1529-1622 of *Frankia* AcN14a 23S rRNA (Figure 20). Similar insertions have been reported for *Frankia* strain ORS020606 and other actinomycetes including *M. luteus* (Höpfl *et al.*, 1988), *M. leprae* (Liesack *et al.*, 1990) *P. thermophila* and *Streptomyces* species (Pernodet *et al.*, 1989; Stackebrandt *et al.*, 1991).

The size of this region is relatively conserved between these actinomycete species. Among the actinomycetes, *Frankia* 23S rRNAs contain the smallest insertions; AcN14a contains 94 nucleotides and ORS020606 contains 96 nucleotides in the corresponding region of their 23S rRNAs. Inserts of 100, 101 and 103 nucleotides are found in *P. thermophila*, *M. luteus* and *S. ambofaciens*, respectively. In contrast, only few nucleotides are present in the corresponding region of members of the genus *Bacillus* (Gutell and Fox, 1988). Insertions in this area are also found in eukaryotic large rRNA subunits (Gutell and Fox, 1988).

Interestingly, the primary structure of this region is highly variable among actinomycete species. Alignment of the corresponding regions of *Frankia* AcN14a, *Frankia* ORS020606, *P. thermophilus* (Stackebrandt *et al.*, 1991) *M. luteus* (Regensburger *et al.*, 1988) and *S. ambofaciens* (Pernodet *et al.*, 1989) is presented in Figure 22. While the two *Frankia* strains share 93% identity, the other actinomycete species share only between 50% to 56% sequence homology with *Frankia* AcN14a insertion sequence. These homology values contrast with those obtained with overall comparisons of 23S and 16S rDNA sequences. Interestingly, this region has also been found to vary among different *Streptomyces* species, with homology values as low as 70% being observed (Stackebrandt *et al.*, 1991).

**FIGURE 22 NUCLEOTIDE SEQUENCE OF THE INSERTION SEGMENTS PRESENT IN DOMAIN III OF 23S RIBOSOMAL RNAS OF *FRANKIA* ACN14A AND FOUR OTHER ACTINOMYCETE SPECIES.**

Fourteen gaps (-) were introduced in the *Frankia* AcN14a sequence to allow the best alignment with those of other actinomycete species. Only the nucleotides that differ from those of *Frankia* AcN14a are shown. Identities are denoted by dots and sequence gaps are marked by hyphens. Nucleotide positions that are ambiguous are denoted by n. The corresponding regions of two nonactinomycete species are included as reference. Numbers refer to positions in the *Frankia* AcN14a sequence. This figure is an adapted version from the one published by Stackebrandt *et al.* (1991).

1529

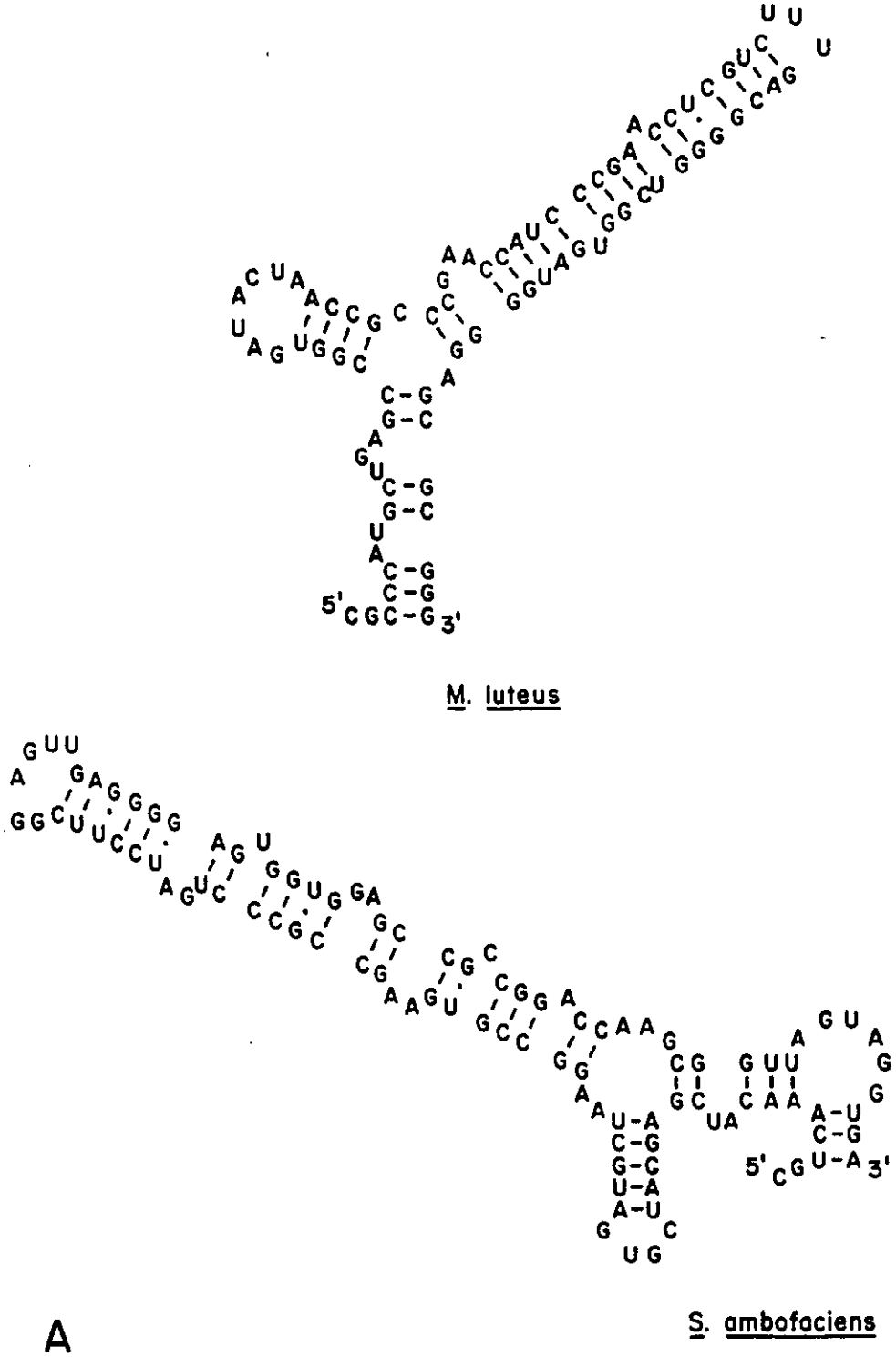
| | | | | | | |
|---|---|---|---|---|---|---|
| *Frankia* AcN14a | GGCGCUGACG | CGCCC-AUGC | UGAACCUGG- | UUGUGCUAAC | CAUCUGAU-- | ----CGGAUG |
| *Frankia* ORS20606 | ....U..... | ..G..-.... | .........U | .......... | .......... | .......... |
| *P. thermophila* | C.U.ACaGU. | .....-G..U | C..GG.C..- | -GA....G.. | ....C..-AC | CUCCUUCC.. |
| *M. luteus* | A.U.AA..AC | .....-.... | ...G..G-.- | -GA.A..... | .GC.C..-AC | CAUC.C..AC |
| *S. ambofaciens* | C..UG...A. | ..U.AA.CAU | C..G.A.C.- | -GA......G | GCCG...AGC | CGCC.U...C |
| *B. subtilis* | AC.U.CUCAC | ---------- | ---------- | ---------- | ---------- | ---------- |
| *E. coli* | U.GUG.U..U | ---------- | ---------- | ---------- | ---------- | ---------- |

1622

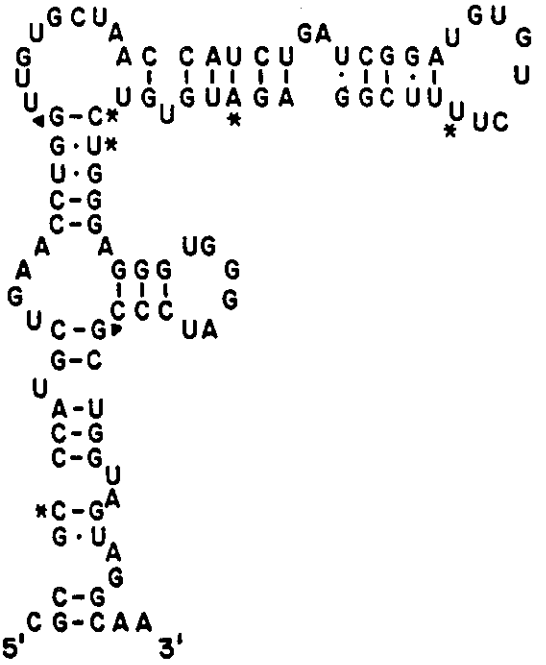
| | | | | | | |
|---|---|---|---|---|---|---|
| *Frankia* AcN14a | U--GUCUUU- | --UCGGAGAU | GUGUCUGGGA | GGGUGGGAUC | CC-GCUGGUA | GUAGGCAAGC |
| Frankia ORS20606 | .......C.. | .........G. | ....UC.... | .......... | ..G....... | .......... |
| *P. thermophila* | G..--...CG | GGCUn.GUUG | UG.GGG.--. | .C.n....C. | .GAU.C.... | .........CG |
| *M. luteus* | CUC......G | ACGG..UCGG | UGA.GG..-. | .C.C....C. | UGAA.C..GG | .......... |
| *S. ambofaciens* | ...--...CG | GAGUU...GG | .A GGU... | .CCGCC.GA. | .AA..G.U.. | .....UG.U. |
| *B. subtilis* | ---------- | ---------- | ---------- | ---------- | ---------- | --.UUUG... |
| *E. coli* | ---------- | ---------- | ---------- | ---------- | ---------- | ---------.. |

**FIGURE 23 PROPOSED SECONDARY STRUCTURE OF THE REGION WITHIN DOMAIN III OF THE 23S RIBOSOMAL RNA OF *FRANKIA* ACN14A AND TWO OTHER ACTINOMYCETES.**

*Frankia* AcN14a 23S rRNA secondary structure proposed for region corresponding to positions 1528 to 1621 of domain III (see Figure 20) is presented. The nucleotide sequence of *Frankia* AcN14a was folded into a secondary structure with the help of the computer program RNAFOLD. Asterisks indicate positions found to vary between *Frankia* strains AcN14a and ORS020606. Triangles indicates the positions of additional nucleotides present in ORS020606.




A



B Frankia sp. ACN14a

The primary sequence of *Frankia* AcN14a as well as those of the other actinomycete species, do not support the folding proposed for the corresponding region of *S. ambofaciens* (Pernodet *et al.*, 1989). Computer analysis of this region suggests the formation of one or two stem structures in all these organisms (data not shown). This region of the actinomycete 23S rRNAs appears to have diverge extensively, and as a result, we are unable to propose a common secondary structure model. Figure 23 represents one possible secondary structure of *Frankia* AcN14a insertion segment. Positions found to vary between the two *Frankia* strains are indicated by asterisks. However, the high degree of divergence observed between the different actinomycetes also makes the evaluation of the putative secondary structure a difficult task; the nucleotide sequences of more closely related organisms, such as those of other *Frankia* strains will probably be required.

In summary, a large insertion segment is present within domain III of the *Frankia* AcN14a 23S rRNA as well as in those of other actinomycetes. The fact that insertion segments, comparable in size and location, are present in other actinomycete species suggests that this extended version of domain III was present in their common ancestor and is now specific to this group of eubacteria. This insertion segment is also present in the 23S rRNA of *Frankia* strain ORS020606 and it is thus expected to be present in all *Frankia* strains. The degree of conservation of the primary structure of this region of the 23S rRNA genes among the different *Frankia* strains is still unknown, but members of the genus *Frankia* are likely to have nucleotide sequences that are more similar to each other than they have with sequences of members of other genera. Supporting that view, *Frankia* strains AcN14a and ORS020606 were found to share 93% identity in that region of their 23S rRNAs, whereas actinomycetes belonging to other genera were found to have less than 56% homology with *Frankia* AcN14a. Of interest is the site for the restriction endonuclease BamH1 (position 1600) located within the insertion segments in the 23S rRNAs of both *Frankia* strains but not in those of the three other actinomycetes. RFLP analysis of selected *Frankia* isolates indicates that this restriction site is conserved (see RFLP section). To our knowledge no such sites have been reported in any other actinomycete sequences and consequently,

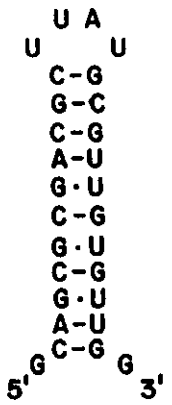
could represent a genus-specific signature. Thus this insertion segment could possibly be useful for the design of genus or strains specific probes depending upon its nucleotide sequence conservation among *Frankia* strains.

The third insertion segments is located in helix 41, which in *Frankia* AcN14a corresponds to nucleotides 1254 to 1299. In *Frankia* strains AcN14a and ORS020606, this helix contains 46 and 45 nucleotides, respectively, whereas only 22 are present in the corresponding region of *E. coli*. Other eubacteria and chloroplast species generally contain even fewer nucleotides (Raué *et al.*, 1988). In contrast, this region contains approximately 20 to 30 nucleotides in archaebacteria, 30 to 50 in mitochondria, whereas up to 200 nucleotides have been reported for the corresponding region of various eukaryotic cytoplasmic rRNA species (Gutell and Fox, 1988; Raué *et al.*, 1988). Interestingly, the size of this region varies between the three actinomycete species. The corresponding region of *M. luteus* and *S. ambofaciens* are smaller than that of *Frankia* AcN14a; they contain 19 and 34 nucleotides, respectively.

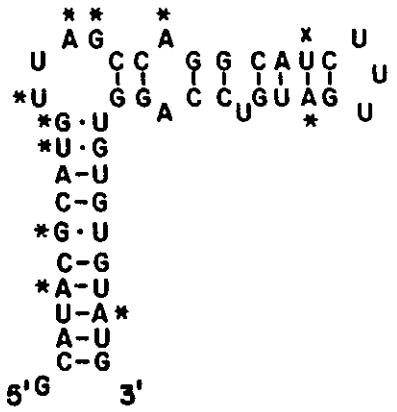
As it could be expected, the primary structure in this region of the 23S rRNAs of the two *Frankia* strains is more similar to each other than each of them is to those of *M. luteus* or *S. ambofaciens*. The two *Frankia* strains share 75% identity in this region whereas only between 40 and 56% identity is shared between either *Frankia* strain and the two actinomycete species. Thus, in contrast to the first two insertion segments found to be quite well conserved between the two *Frankia* strains, this third region appears as highly variable. In prokaryotes as well as in mitochondria and chloroplasts, the nucleotide sequence of this region can be folded into a single, perfect or slightly imperfect helix. Depending on the organism, a terminal loop of 3 to 8 nucleotides has been proposed (Gutell and Fox, 1988). The proposed secondary structures for the corresponding helix of *Frankia* AcN14a and *M. luteus* are presented in Figure 24. As a reference, the secondary structure proposed by different groups for some other eubacteria are included. A regular helical structure is proposed for the organisms having a shortened version of this helix. *Frankia* AcN14a sequence can be folded in a structure similar to that proposed for *S. ambofaciens* (Pernodet *et al.*, 1989). The base pairing of the

**FIGURE 24 PROPOSED SECONDARY STRUCTURES OF THE REGION IN DOMAIN II OF THE 23S RIBOSOMAL RNAS OF *FRANKIA* ACN14A,*M. LUTEUS* AND *S.AMBOFACIENS***

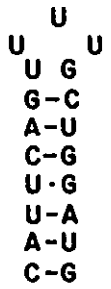
*Frankia* AcN14a 23S rRNA nucleotide sequence corresponding to positions 1529 to 1622 (see Figure 20) as well as the regions corresponding to helix 41 of other eubacteria are presented. The nucleotide sequence of *Frankia* AcN14a and *M. luteus* (Regensburger *et al.*, 1988) were folded into a secondary structure with the help of the computer program RNAFOLD. The secondary structures proposed for *E. coli* (Neefs *et al.*, 1991) and *S. ambofaciens* (Pernodet *et al.*, 1989) have been included as references. Asterisks indicate positions found to vary between *Frankia* strains AcN14a and ORS020606. X indicates positions of a deletion in ORS020606.



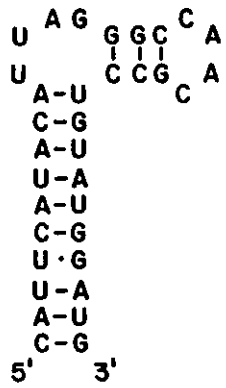

E. coli




Frankia ACN14a




M. luteus




S. ambofaciens

proximal stem structure can be considered proven since many compensatory substitutions are found in various prokaryotic sequences (Gutell and Fox, 1988). However, the validity of the folding proposed for the distal part of the structure remains unknown.

In contrast to the first two regions that have been described, this extended version of helix 41 appears to be specific to members of the genus *Frankia* since no comparable insertions are found in either *S. ambofaciens* or *M. luteus* large subunit rRNAs. Sequence comparison of this region in two *Frankia* strains, AcN14a and ORS020606 suggests that this helix is highly variable among members of this genus. Thus depending on the conservation of this region among members of the genus *Frankia*, this region of the rRNA molecule could be used to derive genus and/or strains specific probes.

In summary, we have identified regions of *Frankia* AcN14a 16S and 23S rRNAs which differ from other eubacterial counterparts. In contrast to the 16S rRNA where no large insertions were observed, two large and one small insertion segment were found within *Frankia* AcN14a 23S rRNA. Two types of variable regions have been observed: some are highly variable within the genus *Frankia* and others highly variable among actinomycetes, but more or less conserved within the genus *Frankia*. Interestingly, the regions exhibiting size variation were clustered in the 5' portions of the small and large subunit rRNAs.

**Analysis of the 5S rRNA gene - primary and secondary structures**

The region coding for 5S rRNA was determined by comparison with the sequence of the 5S rRNA from *E. coli*. Figure 25 represents the 120 nucleotide long *Frankia* AcN14a mature 5S rRNA derived from the corresponding gene sequence (nucleotides 5706-5825). *Frankia* AcN14a 5S rRNA nucleotide sequence is more than 99% identical to that of *Frankia* strain ORS020606 and resembles that of other actinomycetes; they usually share more than 78% identity.

Comparison of *Frankia* AcN14a 5S rRNA sequence to other selected eubacterial sequences is presented in Figure 26. *Frankia* ORS020606, *S. ambofaciens* and *M. luteus* 5S rRNAs differ from that of *Frankia* AcN14a by one (99%), seventeen (86%) and twenty-five nucleotides (79%), respectively. In contrast, the similarity between *Frankia* AcN14a 5S rRNA with those of *B. subtilis* and *E. coli* is 75% and 57%, respectively. The *Frankia* AcN14a 5S rRNA sequence could be folded according to the secondary-structure model proposed by Erdmann and Wolters (1986). Most of the substitutions are either neutral or compensatory in nature, and as a consequence, the overall 5S rRNA secondary structure is conserved among these organisms.

## Analysis of the putative primary transcript

### Putative primary transcript

The three types of rRNA genes have been found to be closely linked and transcribed as a single long precursor molecule in many eubacteria (King *et al.*, 1986). In *Frankia*, the specific sites for synthesis of the precursor molecule have not been studied. A preliminary search for promoter sequences upstream from the 16S rRNA gene and for terminator sequences downstream from the 5S rRNA gene was tentatively done. Four regions resembling the consensus prokaryotic promoter(s) were localized approximately 550 bp upstream from the 16S rRNA gene (Figure 9). Similarly, four putative promoters have been found in the corresponding regions of *S. ambofaciens* (Pernodet *et al.*, 1989) and *S. coelicolor* A3(2) (Baylis and Bibb, 1988b). A putative transcriptional terminator consisting of a perfect inverted repeat (16 nucleotides) followed by 6 thymines has been localized six nucleotides downstream from the 5S rRNA gene. A similar structure has been observed in *S. ambofaciens* (Pernodet *et al.*, 1989). This structure is characteristic of the eubacterial Rho-independent terminators (Lewin, 1987).

Further processing events are necessary to separate and produce the final mature rRNAs. The possibility of extensive base pairing between sequences flanking the mature 16S and 23S rRNAs to form two stems has been

**FIGURE 25 PROPOSED SECONDARY STRUCTURE OF THE *FRANKIA* ACN14A 5S RIBOSOMAL RNA**

The *Frankia* AcN14a 5S rRNA deduced from the corresponding gene sequence (nucleotides 5707-5826 in Figure 9) could be folded in a secondary structure similar to the 5S rRNA prokaryotic model (Erdmann and Wolters, 1986).

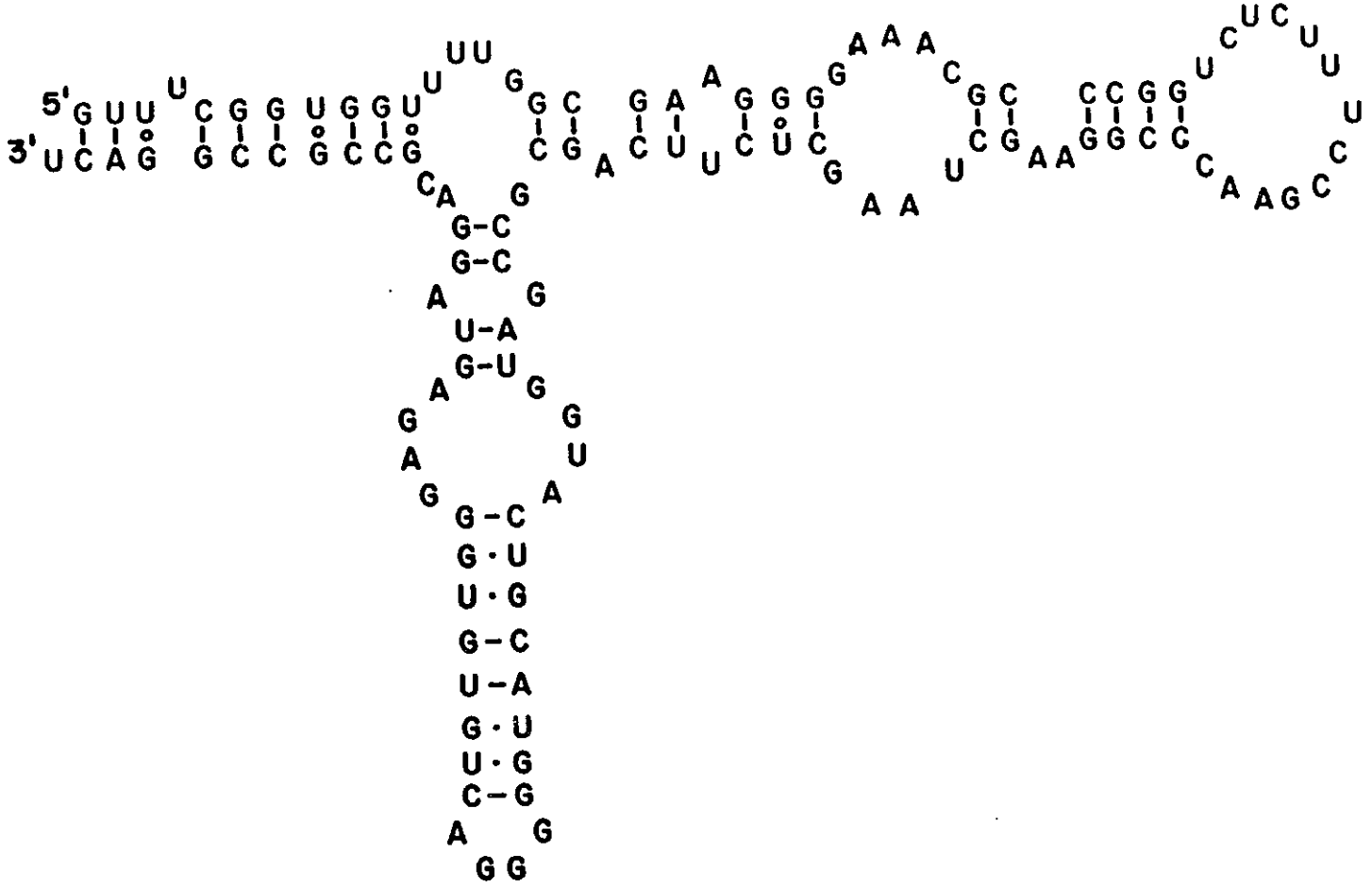

5'
3'
G-C
G-C
U-A
G-U
G·U
U·G
G-C
U-A
G·U
U·G
C-G

**FIGURE 26 COMPARISON OF THE NUCLEOTIDE SEQUENCES OF *FRANKIA* ACN14A AND OTHER SELECTED EUBACTERIAL 5S RIBOSOMAL RNAS.**

The nucleotide sequence of *Frankia* AcN14a and other selected eubacterial 5S rRNAs were aligned (Specht *et al.*, 1990; Pernodet *et al.*, 1989). Only the nucleotides that differ from those of *Frankia* AcN14a are shown. Identities are denoted by dots and sequence gaps are marked by hyphens. The *Frankia* AcN14a 5S rRNA sequence was deduced from the 5S rRNA gene sequence which corresponds to nucleotides 5707-5826 in Figure 9.

| | 1 | | | | | |
|---|---|---|---|---|---|---|
| *Frankia* AcN14a | GUUUCGGUGG | UUUUGGCGAA | GGGGAAACGC | CCGGUCUCUU | UCCGAACCCG | GAAGCUAAGC |
| *Frankia* ORS20606 | .......... | .......... | .......... | .......... | .......... | .......... |
| *S. ambofaciens* | –......... | .CA.A...UG | A......... | .....UA.A. | .......... | .......... |
| *M. luteus* | –G.CU..C.. | CCA.A...UG | .......... | ......C.A. | C......... | .......... |
| *B. subtilis* | —..U..... | CGA.A..... | .A..UC..A. | ...U..C.A. | G......A.. | ....U..... |
| *E. coli* | UGCCU..C.. | CCG.A..GCG | .U..UCC.A. | .U.AC.C.A. | G......U.A | ....UG..A. |

| | 61 | | | | | |
|---|---|---|---|---|---|---|
| *Frankia* AcN14a | UCUUCAGCGC | CGAUGGUACU | GCAUGGGGGA | CUGUGUGGGA | GAGUAGGACG | CCGCCGGACU– |
| *Frankia* ORS20606 | .......... | .......... | .......... | .......... | .......... | .........G– |
| *S. ambofaciens* | CU.A...... | .......... | ...G...... | .CC....... | .......... | ......A..A– |
| *M. luteus* | C.CAU..... | .......... | ...ACC.... | GGU....... | .......U.. | .........A– |
| *B. subtilis* | .......... | ........G. | CGGG..UUU– | .CCCC..U.. | .......... | .....AAG.–– |
| *E. coli* | G.CGU..... | ........G. | .UGG..UCU– | .CCCA..C.. | .......GAA | .U...A.G.AU |

reported for *E. coli* and several other prokaryotes (King *et al.*, 1986). The overall folding observed in the corresponding regions of different organisms are either simple as it is the case in *B. subtilis* (Loughney *et al.* 1983) or more complex as in *E. coli* (Brosius *et al.*, 1981) or *S. ambofaciens* (Pernodet *et al.*, 1989). The structures proposed for *S. ambofaciens* is presented in Figure 27A. Such structures called 16S and 23S RNA stems, can also be formed in the putative primary transcripts of *Frankia* strains AcN14a and ORS020606. The secondary structure proposed for the primary transcript of *Frankia* AcN14a is presented in Figure 27B. The folding is quite complex and, in that sense, resembles that proposed for *S. ambofaciens*. The overall primary structures of the rRNA gene flanking regions are not well conserved between *Frankia* AcN14a and *S. ambofaciens* but both core stems (identified by boxes) exhibit a high degree of conservation between the two organisms. Similarly, the majority of changes between the two *Frankia* strains are located in the secondary stems whereas the two core stems are almost identical. These conserved stem structures are thought to be recognized as processing signals during rRNA maturation (Baylis and Bibb, 1988b; Bram *et al.*, 1980; Loughney *et al.* 1983; Ogasawara *et al.* 1983). As in the primary transcripts of other prokaryotes, a common motif is observed in the two stem structures proposed for *Frankia* AcN14a. These repeats are bracketed in Figure 27B. In *B. subtilis*, they have been shown to be sites for RNAse III cleavage (Loughney *et al.* 1983; Ogasawara *et al.*, 1983). Thus in *Frankia* AcN14a, these regions probably play a role in processing of the primary transcript.

The *Frankia* AcN14a 16S-23S rRNA gene spacer is comparatively larger than that of other eubacteria lacking tRNA sequences (Pernodet *et al.*, 1989; Green *et al.*, 1985). About one third of the 16S-23S rRNA spacer in *Frankia* AcN14a is involved in the formation of the 16S and 23S RNA stem structures whereas the remaining nucleotides can potentially form a complex helical structure having the shape of the letter Y. No such complex structure has been proposed for eubacteria lacking tRNA sequences in this spacer and the importance of such structure is unknown. Analysis of this last portion of the 16S-23S rRNA gene spacer with a second computer program (DNASIS) reveals that it could also folded into two mutually exclusive tRNA-like structures (2403-

2488 and 2416-2493). However, they are probably not tRNAs, since the structures lack many features that are typical of tRNAs (Sprinzl *et al.*, 1989).

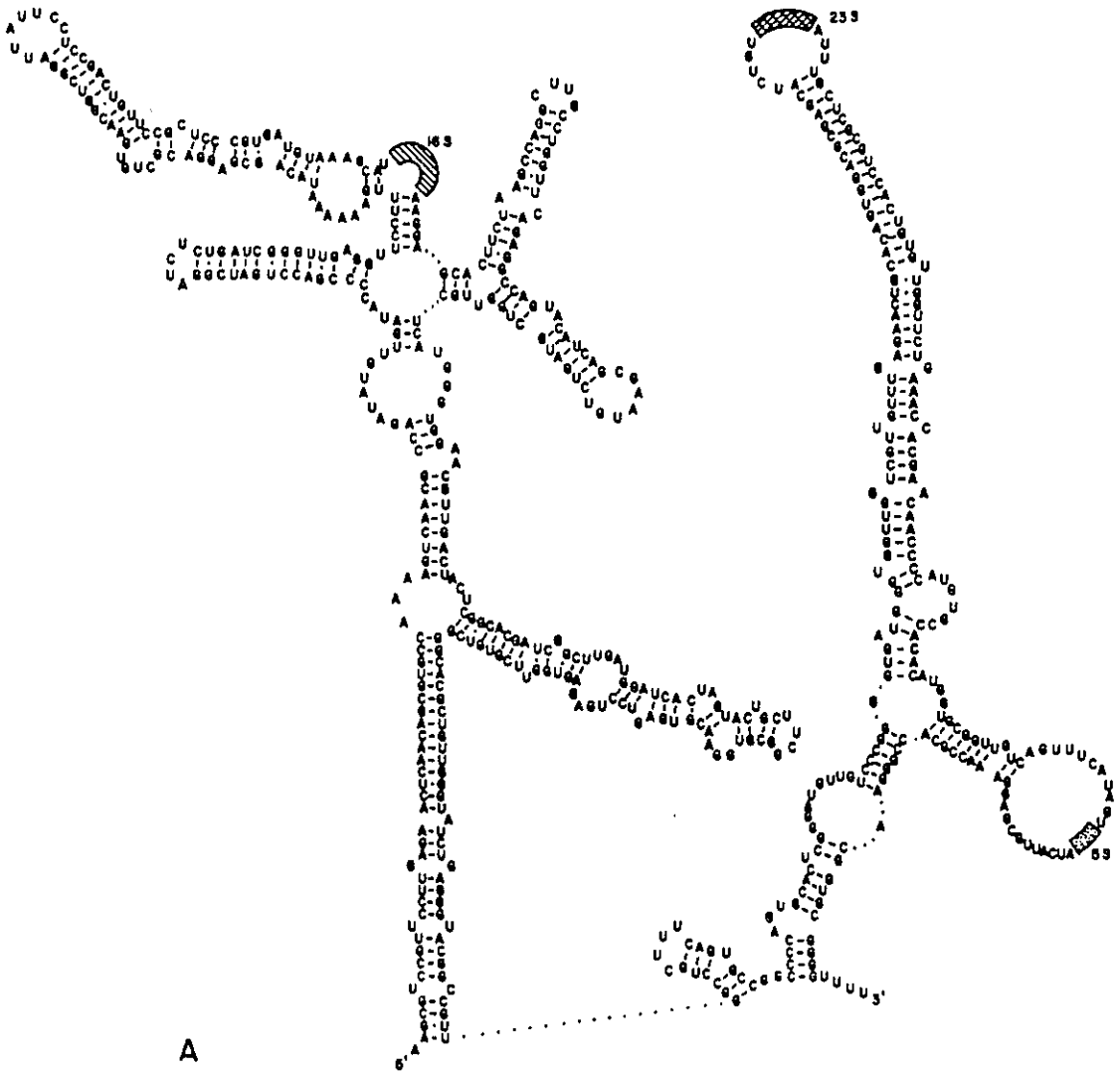
As in other prokaryotic primary transcripts, most nucleotides of the 23S-5S spacer (56 of the 73 nucleotides) of *Frankia* AcN14a rRNA operon are involved in the formation of the 23S RNA stem. However, as no base-pairing appears to be possible between the mature 5S rRNA flanking sequences, the future mature 5S rRNA sequence is apparently not contained in a loop, as it is in the model proposed for *S. ambofaciens* (Pernodet *et al.*, 1989). This feature is not unique to the primary transcript of *Frankia* AcN14a; a similar folding has been proposed for the *E. coli* primary transcript (Brosius *et al.*, 1981).

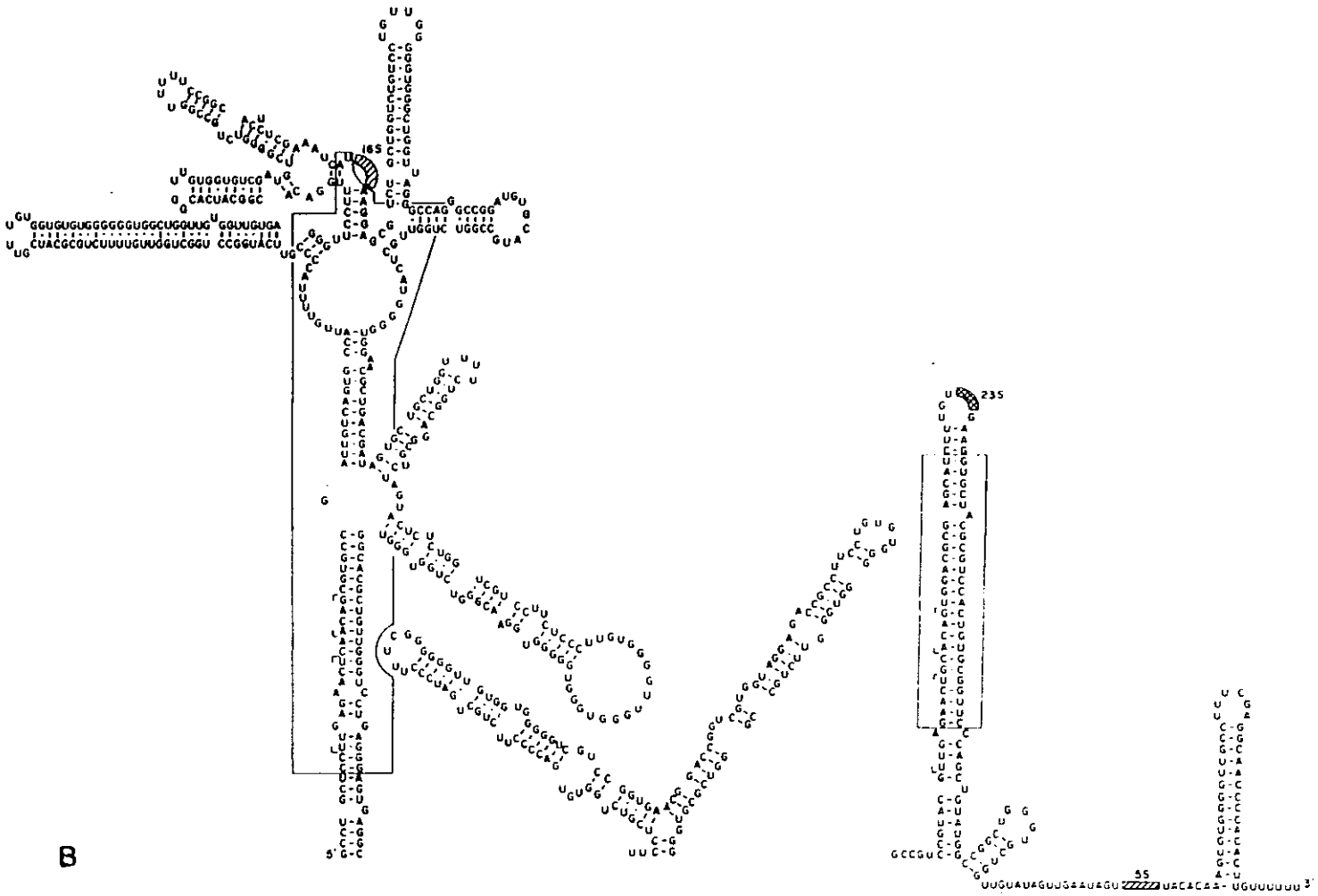
Another contrasting observation is the finding of a single putative termination signal, located only six nucleotides downstream from the future mature 5S rRNA sequence. In contrast to the *E. coli* and *B. subtilis* primary transcripts in which two putative termination signals have been found, no other inverted repeats were found in the two hundred forty nucleotide long region downstream from the mature 5S rRNA gene. Similar hairpin structures can be formed at the end of the primary transcript of *Frankia* ORS020606 and *S. ambofaciens*. In *Frankia* ORS020606, a long inverted repeat of twenty-three nucleotides can form a hairpin structure immediately downstream from the 5S rRNA gene whereas for *S. ambofaciens*, the putative termination signal is located fifteen nucleotides downstream from the 5S rRNA gene and involved an inverted repeat of seven nucleotides (Pernodet *et al.*, 1989).

Thus, the analysis of the *Frankia* AcN14a rRNA gene cluster and their flanking regions has revealed interesting features. As in most prokaryotes, the rRNAs are probably cotranscribed because there are no obvious terminator sequences until after the end of the 5S rRNA sequence. Other putative transcription and maturation signals have been pointed out. In addition, nucleotide sequence of the regions flanking the 16S and 23S rRNA gene enabled hypothetical stem structures to be drawn. Similarities were found between the rDNA organization in *Frankia* AcN14a and that of other organisms, notably *Frankia* strain ORS020606, *B. subtilis*, *E. coli* and

**FIGURE 27 PUTATIVE SECONDARY STRUCTURES OF THE ASSUMED PRIMARY TRANSCRIPTS OF *FRANKIA* ACN14A AND *S. AMBOFACIENS* RIBOSOMAL RNA OPERONS.**

Putative secondary structures of the assumed primary transcripts of the (**A**)*S. ambofaciens* (Pernodet *et al.* 1989) and (**B**) *Frankia* AcN14a rRNA gene operons. Only the nucleotides absent from the mature rRNAs (hatched) are indicated. The regions showing a high primary structure conservation are boxed and the common motif to both the 16S and 23S stems is bracketed.


A




B

*S. ambofaciens*.

**G+C content**

Members of the genus *Frankia* have a high G+C content; an average genomic G+C content of 70% has been reported (reviewed by Normand and Lalonde, 1986). The G+C content of rRNA genes of various eubacteria have been shown to differ from their total genomic base composition. For example, *S. ambofaciens* and *E. coli* which have genomic G+C content of 73% and 51%, have rRNA genes with an average of 58% (Pernodet *et al.*, 1989) and 56% G+C (Brosius *et al.*, 1981), respectively. We have calculated the G+C content of various regions of the *Frankia* AcN14a putative rRNA operon as well as of the regions flanking the rRNA gene cluster.

As expected, the putative primary transcript has an overall G+C content of 58% which is low compare to the values calculated for the flanking regions. Table 4 lists the G+C values of the different regions of the *Frankia* AcN14a rRNA gene cluster. The region between the first putative promoter and the 16S rRNA gene has a G+C content of 59%. A similar value of 60% was calculated for the region corresponding to the mature 16S rRNA. The spacer region between the 16S and 23S rRNAs contains 64% G+C whereas a lower value of 55% was calculated for the spacer region between the 23S and 5S rRNA genes. Values of 57% and 61% were calculated for the sequence corresponding to the 23S and 5S rRNA genes, respectively. Finally, the last small portion of the putative primary transcript between the 5S rRNA gene and the putative terminator, has a G+C content of 52%. In contrast, the regions flanking the rRNA gene cluster have a base composition similar to those reported for *Frankia* genomic DNA (reviewed by Normand and Lalonde, 1986). The regions upstream from the first putative promoter (70 nucleotides) and downstream from the putative terminator (192 nucleotides) contain 72% and 71% G+C, respectively.

Thus the G+C content of the *Frankia* rRNA genes has been found to be relatively low compare with the value reported for its genome. This observation is not unique to *Frankia*, since the G+C content of the rRNA gene sequences is relatively constant and appears to be independent from the

TABLE 4 COMPARISON OF THE G+C CONTENT OF SPECIFIC REGIONS OF THE RIBOSOMAL RNA OPERONS OF *FRANKIA* ACN14A AND SELECTED EUBACTERIA.

| Area and position | | % G+C calculated for specific regions of the *Frankia* AcN14a sequence | | % G+C of the putative rRNA primary transcripts and overall genomic% G+C reported for selected eubacteria | | |
|---|---|---|---|---|---|---|
| | | region | average | *Frankia* AcN14a [1] | *E. coli* [2] | *S. ambofaciens* [3] |
| Putative primary transcript | promoter (70-598) | 59 | 58 | 58 | 52 | 58 |
| | 16S rRNA (599-2110) | 60 | | | | |
| | 16S-23S rRNA spacer (2111-2527) | 64 | | | | |
| | 23S rRNA (2528-5626) | 57 | | | | |
| | 23S-5S rRNA spacer (5627-5705) | 55 | | | | |
| | 5S rRNA (5706-5825) | 61 | | | | |
| | terminator (5826-5867) | 52 | | | | |

| | | | | | | |
|---|---|---|---|---|---|---|
| Flanking regions | upstream sequence (1-69) | 72 | 72 | 70 | 51 | 73 |
| | downstream sequence (5868-6059) | 71 | | | | |

The G+C content of specific regions *Frankia* AcN14a rDNA, flanking sequences and total genome (Normand and Lalonde, 1986) are presented. In addition, values reported for selected eubacteria have been included as references.

1 The overall % G+C value for the primary transcript is an average of the values calculated for the specific regions of the putative primary transcript whereas the overall genomic % G+C value is that reported in Lalonde and Normand (1986).

2 The overall % G+C value for the primary transcript is an average of the values calculated for the specific regions of the primary transcript reported in Brosius *et al.* (1981) whereas the overall genomic % G+C value is that reported in Pernodet *et al.* (1989)

3 The overall % G+C value for the primary transcript is an average of the values calculated for the specific regions of the primary transcript reported in Pernodet *et al.* (1989) whereas the overall genomic % G+C value is that reported in Hunter (1985)

overall base composition of the genome.

### Post-transcriptional modifications and higher level interactions

Some residues of the *E. coli* 16S and 23S rRNAs have been shown to be post-transcriptionally modified. Even though the methylation state of the different positions of *Frankia* rRNA has not been determined, conservation of the nucleotide composition at the positions known to be modified in *E. coli* rRNAs suggests a similar modification patterns in the two species. Most of the twenty-three positions post-transcriptionally modified in *E. coli* are conserved between the two eubacterial species. Only two residues, the two cytosine residues at positions 1171 and 1481 of *Frankia* AcN14a predicted 16S rRNAs were found to have replaced the two guanosine residues present in the corresponding positions of the *E. coli* molecule.

Over the last few years, it has become evident that the rRNAs are the principal functional constituent of the ribosome. Supporting that view, phylogenetic comparison of a large number of rRNAs has revealed that despite often considerable differences in sequences, the secondary and probably the tertiary structures of the small and large-subunit rRNAs were extremely well conserved. In addition, the development of systems that allow the study of the effects of structural changes in rRNAs on ribosome function as well as the analysis of mutations causing antibiotic resistance have added direct evidence for a functional role of rRNAs. At present, several elements and regions of the prokaryotic rRNAs, mainly *E. coli* rRNAs, have been found to be intimately involved in specific ribosomal functions. Extrapolation of these data to other less well studied organisms, such as *Frankia* is, for the moment, limited to establishing wether the elements in question appeared to be conserved in *Frankia* rRNAs. *Frankia* AcN14a rRNA sequences were compared to those of *E. coli* to see if the positions possibly involved in tertiary folding or antibiotic sensitivity were conserved in these two eubacteria.

**TABLE 5 SEQUENCE VARIATION AT POSITIONS OF 16S AND 23S RIBOSOMAL RNA SEQUENCES INVOLVED IN TERTIARY INTERACTIONS**

| Eubacteria | 16S rRNAs | | | 23S rRNAs | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| *Frankia* AcN14a | $UCC_{18}$ $-GGA_{884}$ | $GGC_{471}$ $-GCC_{490}$ | $G_{534}$ $-C_{832}$ | $G_{65}$ $-U_{89}$ | $CCAC_{503}$ $-GUGG_{2614}$ | $A_{1375}$ $-U_{2222}$ | $U_{1456}$ $-A_{1515}$ | $U_{2231}$ $-A_{2263}$ | $AGU_{2534}$ $-ACU_{2591}$ |
| *E. coli* | UCA-UGA | GGC-GCC | G-C | U-A | CCAU-AUGG | A-U | U-A | U-A | AUG-CAU |
| *B. subtilis* | UCC-GGA | GGC-GCC | G-C | A-U | CCAU-AUGG | A-U | U-A | U-A | AAG-CUU |
| *M. luteus* | UCC-GGA | GGC-GCC | G-C | U-A | CCAC-GUGG | A-U | U-A | U-A | AGU-ACU |
| *S. ambofaciens* | UCC-GGA | GGC-GCC | G-C | G-U | CCAC-GUGG | A-U | U-A | U-A | AGU-ACU |

Nucleotide composition at positions proposed to be involved in tertiary structural interactions within 16S and 23S rRNAs. The corresponding nucleotide sequence of *Frankia* ORS20606 rRNAs is identical to that of *Frankia* AcN14a.

Phylogenetic evidence is available for the presence of three different tertiary interactions in the small subunit rRNA (reviewed in Pleij, 1990) and seven others in the large subunit rRNAs (Leffers *et al.*, 1987; Höpfl *et al.*, 1989). A comparison of *Frankia* AcN14a base composition at these positions with those of selected eubacteria is presented in Table 5. *Frankia* AcN14a 16S and 23S rRNA sequences are generally in concordance with the proposed tertiary interactions; only one of the interactions proposed for the 23S rRNA (not listed in Table 5) is disproven. These residues are located at positions 67 and 74 of *Frankia* AcN14a rRNA sequence. While a canonical pairing is observed in most eubacterial sequences (Höpfl *et al.*, 1989), two uracil residues are present in *Frankia* AcN14a and ORS020606 23S rRNAs.

The nucleotides which are target sites for the binding and/or for the action of various molecules such as elongation factors, tRNA or antibiotics have been determined by footprinting studies, in which the various ligands bound to ribosomes or their rRNA components have been shown to protect specific rRNA residues from chemical or nuclease attack (e.g. Moazed and Noller, 1987). In addition, attempts to link specific features of rRNAs to ribosomal functions have greatly been helped by the identification of naturally occurring mutations or modifications in rRNA causing antibiotic resistance. This is because it is assumed that the structurally altered region is likely to be involved in the ribosomal function perturbed by the given antibiotic.

Since all the positions implicated in antibiotic sensitivity are identical in *E. coli* and *Frankia* AcN14a, it suggests that the latter is sensitive to all these antibiotics. This is not surprising since these sites are generally located within highly conserved regions of the rRNAs. Thus by comparing sequences one could predict the sensitivity or resistance of *Frankia* to various antibiotics. This would then have to be verified *in vivo* since mechanisms other than mutation in the rRNAs such as mutation of ribosomal proteins, post-transcriptional modification of the rRNAs (reviewed by Cundliffe, 1989a) have also been implicated in antibiotic resistance.

## 2. RFLP ANALYSIS OF *FRANKIA* ISOLATES

A method is needed for the identification of closely related strains. Grouping of the most related *Frankia* strains could facilitate preliminary screening for the identification of the most efficient *Frankia-Alnus* association since the testing of one representative of each group would eliminate the duplication of strains selected for inoculation and effectiveness tests. Such a method is RFLP mapping.

A collection of *Frankia* isolates from across Canada and a standard strain, ArI3, were provided by PNFI. DNA was extracted, digested with BamH1 or HaeIII and subjected to Southern hybridization in order to detect RFLP in the rDNA region. Hybridization patterns were obtained using cloned fragments of *Frankia* AcN14a rDNA regions, thus specific for 23S and 16S rRNA sequences, as probes. All the isolates, except one were found to contain two copies of the rDNA region. Table 6 lists the sizes of the bands hybridizing to the probes. The data reported in this table comes from different gels and thus the sizes are subjected to gel-to-gel variation. As we can see, digestion of the various DNAs with HaeIII revealed little polymorphism. No RFLPs are observed with the 23S rDNA probe whereas only a few are detected with the 16S rDNA probe. In contrast, extensive polymorphism has been revealed by BamH1 digestion. A total of nine distinct 23S rDNA hybridizations patterns, designated L1 to L9, and ten distinct 16S rDNA hybridization patterns, designated S1 to S10, could be distinguished (Tables 7A and 7B). As shown in these Tables, 23S rDNA hybridization patterns consisted of two to three bands whereas 16S rDNA hybridization patterns consisted of one to three bands. The presence of two fragments of identical sizes or the lost of a BamH1 restriction site in the region upstream from one of the 16S rRNA gene could explain the detection by the 16S rDNA probe, of a single band in forteen isolates. While twenty-two of the thirty-nine isolates have 23S rDNA hybridization patterns identical to that of the ArI3 strain (1.9 and 10.4 kb fragments), most of them have 16S rDNA hybridization patterns different from that of the reference strain.

**TABLE 6 SUMMARY OF RFLP DATA FOR *FRANKIA* ISOLATES**

| Isolates | Sizes of bands hybridizing to the selected probes | | | |
|---|---|---|---|---|
| | 23S rDNA | | 16S rDNA | |
| | BamH1 | HaeIII[a] | BamH1 | HaeIII |
| ArI3 | 2.84, 1.93[b] | 1.04, 0.47[c] | 10.2, 4.87[d] | 0.84, 0.51[e] |
| 1 | 5.15, 2.78* | 0.92, 0.53 | 5.0* | 0.82, 0.57 |
| 2 | 5.15, 2.78* | 0.92, 0.53 | 5.0* | 0.82, 0.53 |
| 3 | 4.95, 2.65 | 1.06, 0.42 | 4.88 | 0.77, 0.45 |
| 5 | 2.65, 1.85 | 1.06, 0.42 | 4.75 | 0.77, 0.45 |
| 7 | 2.65, 1.85 | 1.06, 0.42 | (4.75) | 0.77, 0.45 |
| 8 | 5.15, 1.86* | 0.92, 0.53 | 5.10* | 0.82, 0.57 |
| 10 | (2.5, 1.9) | n.a. | (10.8) | n.a. |
| 11 | 2.65, 1.85 | 1.06, 0.42 | 5.60, 5.30 | 0.77, 0.53, 0.45 |
| 12 | (2.5, 1.9) | 0.92, 0.53 | n.a. | 0.82, 0.57 |
| 13 | 2.81, 1.99 | 0.92, 0.53 | 5.4 | 0.82, 0.57 |
| 14 | 4.90, 1.85 | 0.92, 0.53 | 12.1, 4.9* | 0.82, 0.57 |
| 15 | 2.65, 1.85 | 1.06, 0.42 | 4.75 | 0.77, 0.45 |
| 17 | 2.65, 1.85 | 1.06, 0.53 | 4.75 | 0.77, 0.45 |
| 18 | (2.5, 1.9) | n.a. | 4.75 | 0.82, 0.57 |
| 19 | 1.84, 1.72* | 0.92, 0.53 | 5.95, 4.88* | 0.77, 0.45 |
| 20 | 1.84, 1.72* | 1.06, 0.53 | 5.80, 4.75* | 0.77, 0.45 |
| 21 | 2.78, 1.95 | 1.06, 0.42 | 6.0, 4.8 | 0.45 |
| 22 | 5.83, 3.36* | 1.06, 0.42 | 5.41, 4.27* | 0.45 |
| 23 | 5.30, 3.40* | 0.92, 0.53 | 6.4, 4.85* | 0.82, 0.57 |
| 24 | 2.64, 1.84 | 0.92, 0.53 | 6.0, 4.9 | 0.82, 0.57 |
| 26 | 5.25, 2.65 | 1.06, 0.42 | 5.0* | 0.77, 0.45 |
| 28 | 2.72, 1.90 | 1.06, 0.42 | 10.9, 5.05 | 0.77, 0.45 |
| 29 | 2.72, 1.90 | 1.06, 0.42 | 9.7, 4.9 | 0.77, 0.45 |
| 30 | 5.25, 2.79* | 0.98, 0.34 | 5.0 | n.a. |

| | | | | |
|---|---|---|---|---|
| 33 | 3.72, 3.05, 2.09* | 0.98, 0.34 | 6.68, 5.93, 5.56* | n.a. |
| 34 | 8.0, 2.73* | 0.86, 0.35 | 5.66, 4.06* | n.a. |
| 35 | 2.80, 1.98 | 0.98, 0.34 | n.a. | n.a. |
| 36 | 2.72, 1.90 | 1.06, 0.42 | 6.0, 4.9 | 0.77, 0.45 |
| 37 | 2.60, 1.90 | n.a. | n.a. | n.a. |
| 38 | 2.89, 2.12* | 1.06, 0.52 | 6.03, 4.06* | 0.77, 0.45 |
| 39 | 2.72, 1.90 | 1.06, 0.42 | 4.9 | 0.77, 0.45 |
| 40 | 2.61, 1.79 | 0.98, 0.34 | 9.75, 4.70* | n.a. |
| 42 | 2.5, 1.67 | 0.98, 0.34 | 5.60, 4.85 | n.a. |
| 46 | 5.40, 2.77 | 0.98, 0.34 | 5.2 | n.a. |
| 48 | 2.75, 1.86 | 1.15, 0.57 | 10.6, 5.20 | 0.90, 0.52 |
| 49 | 2.75, 1.92 | 1.15, (0.57) | n.a. | n.a. |
| 50 | 10.2, 3.03* | 1.15, 0.52 | 4.6, 4.45* | n.a. |
| 52 | 2.65, 2.04* | 1.15, (0.52) | 4.29 | n.a. |

NOTES

[a] The observed variation is due to differences between gels. In every case, the observed fragments migrated with the same mobility as ArI3.

[b] average of 8 values: 2.84 +/- 0.21; 1.93 +/- 0.08; [c] average of 5 values; 1.04 +/- 0.08; 0.47 +/- 0.08; [d] average of 4 values; 10.2 +/- 0.81; 4.87 +/- 0.36; [e] average of 4 values; 0.84 +/- 0.05; 0.51 +/- 0.04.

* indicates repeated; ( ) indicates faint band at this position.

n.a. indicates that the data is not available.

Compilation of the two series of restriction patterns has permitted the classification of twenty-seven isolates into six groups (A-F). In addition, eight other isolates were found to have unique patterns. Table 8 summarizes the grouping of the selected isolates according to the hybridization patterns to both probes and their corresponding rDNA restriction patterns. Four of the groups contain between four to seven members each (A, B, C and D) while two members are present in each of the two other groups (E and F). Group A isolates comprises the standard strain ArI3 and at least four other isolates (#28, #29, #40, and #48). An identical hybridization pattern is observed for *Frankia* AcN14a from which one of the two operons, split between a 10.4 kb and a 1.9 kb BamH1 fragments, has been sequenced. From that we conclude that these strains also possess two rRNA operons; one contains BamH1 fragments of 2.8 kb and 4.9 kb while the second corresponds to 1.9 kb and 10.4 kb (8.0 kb) fragments. Group B isolates (#5, #7, #13, #15, #17, #18, and #39) also have the 2.8 kb and 1.9 kb fragments but only one fragment of 5.0 kb hybridizes to the 16S rDNA specific probe suggesting the lost of the large BamH1 fragment in members of this group. Similarly, group C isolates (#1, #2, #3, #26, #26,#30, and #46) have a single 5.0 kb 16S rDNA specific fragment but their 23S rDNA specific fragments differ from the latter since they have been replaced by 5.2 kb and 2.8 kb fragments. The last three groups have identical 16S rDNA specific patterns consisting of fragments of 4.8 kb and 5.9 kb, but they can be differentiated by their 23S rDNA patterns. Group D isolates (#21, #24, #36, and #42) have a 23S rDNA specific pattern identical to that of group A and B isolates (1.9 and 2.8 kb). Members of groups E (#19 and #20) have 23S rDNA specific fragments of 1.8 kb. Members of groups E (#19 and #20) have 23S rDNA specific fragments of 1.8 kb and 1.7 kb whereas the corresponding fragments of group F (#22 and #23) isolates have sizes of 5.6 kb and 3.4 kb. In addition to these groupings, there remains eight isolates (#8, #11, #14, #33, #34, #38, and #50) which have unique hybridization patterns. Since four fragments are detected in all except one isolate, we suggest that two rRNA operons are present in most *Frankia* strains. Isolates #33 probably has three rRNA operons. Figure 28 is an autoradiogram showing RFLP patterns observed for selected isolates.

**TABLE 7A GROUPING OF ISOLATES ACCORDING TO 23S rDNA RFLPS**

| Group | Isolate | Pattern (average values in kb) | | |
|---|---|---|---|---|
| L1 | ArI3, #5, #7, #10, #11, #12, #13, #15, #17, #18, #21, #24, #28, #29, #35, #36, #37, #38, #39, #40, #42, #48, #49 | | 2.8 | 1.9 |
| L2 | #1, #2, #3, #26, #30, #46 | 5.2 | 2.8 | |
| L3 | #8, #14 | 5.0 | | 1.9 |
| L4 | #19, #20 | | | 1.8 1.7 |
| L5 | #22, #23 | 5.6 | 3.4 | |
| L6 | #34 | 8.0 | 2.7 | |
| L7 | #50 | 10.2 | 3.0 | |
| L8 | #33 | | 3.7 3.1 | 2.1 |
| L9 | #52 | | 2.7 | 2.0 |

**TABLE 7B GROUPING OF ISOLATES ACCORDING TO 16S rDNA RFLPS**

| Group | Isolate | Pattern (average values in kb) | | |
|---|---|---|---|---|
| S1 | ArI3, #28, #29, #40, #48 | 10.3 | 4.9 | |
| S2 | #1, #2, #3, #5, #7, #8, #13, #15, #17, #18, #26, #30, #39, #46 | | 5.0 | |
| S3 | #19, #20, #21, #22, #23, #24, #36, #42 | | 5.9 4.8 | |
| S4 | #34, #38 | | 5.8 | 4.1 |
| S5 | #14 | 12.1 | 4.9 | |
| S6 | #10 | 10.8 | | |
| S7 | #33 | 6.7 | 5.9 5.6 | |
| S8 | #50 | | 4.6 | |
| | | | 4.45 | |
| S9 | #11 | | 5.6 5.3 | |
| S10 | #52 | | 4.3 | |

The values in these tables have been averaged.

**TABLE 8 ASSIGNMENT OF *FRANKIA* ISOLATES TO GROUPS BASED UPON 23S AND 16S rDNA RFLPS.**

| GROUP | 23S SIZE | | | 16S SIZE | | ISOLATE |
|---|---|---|---|---|---|---|
| A | | 2.8 | 1.9 | 10.3 | 4.9 | ArI3, #28, #29, #40, #48 |
| B | | 2.8 | 1.9 | | 4.9 | #5, #7, #13, #15, #17, #18, #39 |
| C | 5.2 | 2.8 | | | 4.9 | #1, #2, #3, #26, #26,#30, #46 |
| D | | 2.8 | 1.9 | 5.9 | 4.9 | #21, #24, #36, #42 |
| E | | 1.9 | 1.7 | 5.9 | 4.9 | #19, #20 |
| F | 5.6 | 3.4 | | 5.9 | 4.9 | #22, #23 |

The values have been averaged.

The six hybridization patterns could be related to each other by a succession of simple mutational events. Figure 29 illustrates the possible relationship between five of the six different groups. According to the model proposed, the BamH1 site present in the 23S rRNA gene sequence would appears to be conserved among *Frankia* isolates and the observed RFLPs would be the result of mutation in BamH1 sites located outside the rDNA sequences.

In summary, the present work demonstrates the usefulness of rRNA gene hybridization patterns for the categorization of *Frankia* isolates into groups. In addition, most of these groups could be related to each other by a series of simple mutational events.

**FIGURE 28 SOUTHERN TRANSFERS OF SELECTED *FRANKIA* ISOLATES**

Southern hybridization of BamH1 digest of total DNA from selected *Frankia* isolates. In **A.** and **B.** DNA samples were hybridized with the 23S rDNA specific probe (pnod4) and in **C.** the 16S rDNA probe. The order of the samples is as follow:

**A.** and **C.**: ArI3, #19, #20, #21, #22, #26, #28, #29 and #36.

**B.** : ArI3, #1, #14, #23, #30 and #34.

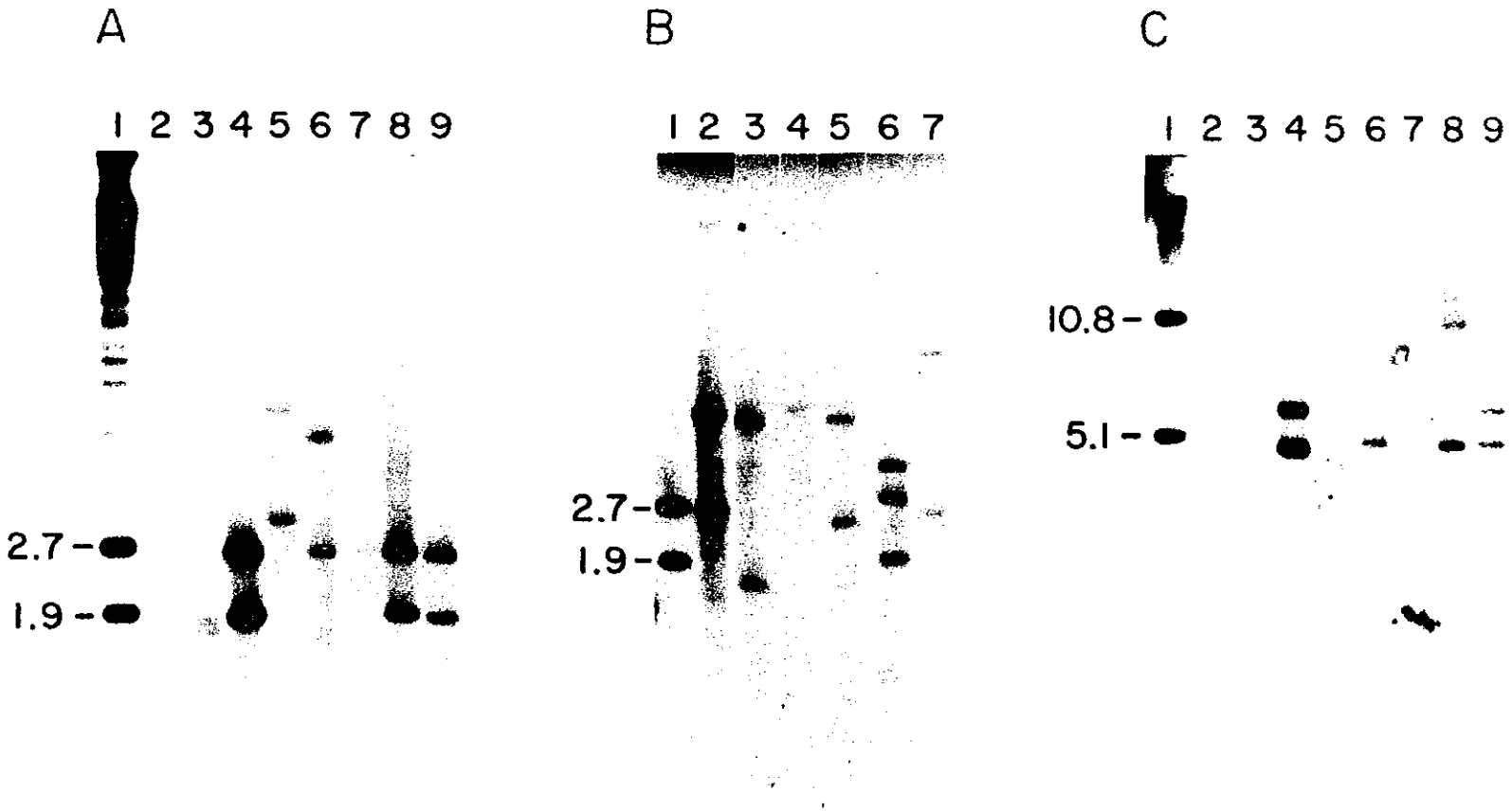
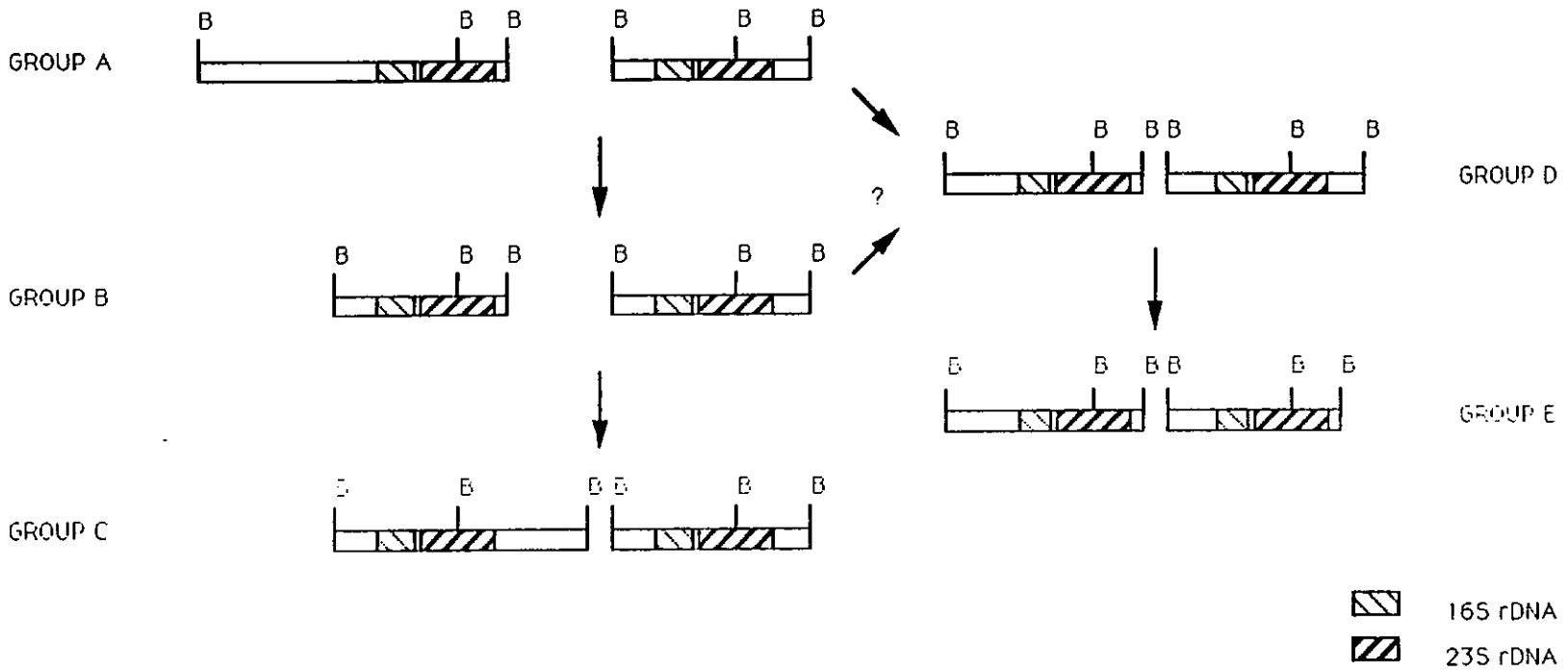

A
1 2 3 4 5 6 7 8 9
2.7
1.9
B
1 2 3 4 5 6 7
2.7
1.9
C
1 2 3 4 5 6 7 8 9
10.8
5.1

**FIGURE 29 POSSIBLE RELATIONSHIP BETWEEN RFLP PATTERNS OBSERVED FOR *FRANKIA* ISOLATES**

A possible relationship between five of the six patterns listed in Table 8 is proposed. The RFLP patterns were arranged in order to minimize the number of mutations necessary to convert member of one group into a member of a second group. RFLP pattern of Group F isolates which is not present in the figure can be derived from that of group D or E by two mutations.


GROUP A
GROUP B
GROUP C
GROUP D
GROUP E
?
B
16S rDNA
23S rDNA

# CHAPTER 4: DISCUSSION

## CLONING AND CHARACTERIZATION OF A *FRANKIA* AcN14a rRNA OPERON

The work in this thesis describes the cloning and the characterization of the ribosomal rRNA genes from *Frankia* AcN14a. To achieve these goals, a *Frankia* AcN14a genomic library was constructed and screened for the presence of clones containing the 23S rDNA sequence. Part of one genomic clone, λFr-8, containing all three rRNA genes has been sequenced (Figure 9). The second rRNA gene cluster of *Frankia* AcN14a has also been cloned but was not further analyzed.

As in most prokaryotic genomes, the *Frankia* AcN14a rRNA genes are closely linked with the order 5' 16S-23S-5S 3' suggesting their organization into an operon. However, in contrast to many other prokaryotes, tRNA genes are absent from the *Frankia* AcN14a rRNA gene cluster, as they are absent from that of *Frankia* strain ORS020606. This organization resembles that observed in *Streptomyces* which are considered close relatives of *Frankia* (Pernodet *et al.*, 1989).

We have proposed a possible folding for the putative primary rRNA gene transcript of *Frankia* AcN14a which has approximately 5700 bases in length and shares many features with that proposed for *S. ambofaciens* (Pernodet *et al.*, 1989). Putative transcription and maturation signals could be identified in the regions flanking the rRNA genes, supporting the proposal that the rRNA genes are organized into an operon (Figures 9 and 27B). Four putative promoters and a putative terminator sequence located upstream from the 16S rRNA gene and downstream from the 5S rRNA gene, respectively, have been identified. The -35 and -10 regions of two of these putative promoters contain the nucleotides that are the most highly conserved among the eubacterial consensus promoter, (TTGxxx) and (TAxxT). In addition, extensive base-pairing is possible between the regions flanking the 16S and 23S rRNAs

in the putative primary transcript, giving stem-loop structures with the 16S and 23S rRNAs in the loops. Similar stems have been shown to be processing sites during rRNA maturation in other gram positive eubacteria such as *B. subtilis* and *S. coelicolor* A3(2) (Loughney *et al.*, 1983; Baylis and Bibb, 1988b). Furthermore, two regions related in nucleotide sequence and thought to function as RNAse III cleavage sites have been identified in the 16S and 23S RNA stems of the putative primary transcript. The ends of the precursor rRNAs have been shown to be located in the corresponding regions of *B. subtilis* and *S. coelicolor* A3(2) (Loughney *et al.*, 1983; Baylis and Bibb, 1988b). In several of the *B. subtilis* operons, a set of opposed guanine residues is found in both stems and have been suggested to be the cleavage sites. In *Frankia* AcN14a, one of the sets of opposed guanine residues has changed to a G-C pair due to a guanine residue change to a cytosine residue in the 3' part of the 23S RNA stem. A similar and several other substitutions are present in the corresponding region of *Frankia* strain ORS020606. No such mutation is seen in the corresponding regions of *S. ambofaciens* (Pernodet *et al.*, 1989). Therefore it appears that the initial processing events in *Frankia* AcN14a, despite having strong resemblance to *B. subtilis* and *S. ambofaciens*, may show some different characteristics.

In summary, the clustering of the rRNA genes with the order 5' 16S-23S-5S 3' and the identification of putative promoter, terminator as well as maturation signals in the region corresponding to the putative primary rRNA transcript are indirect evidences for the organization of the *Frankia* AcN14a rRNA genes in an operon.

## PRIMARY AND SECONDARY STRUCTURES OF THE RIBOSOMAL RNA GENES - IDENTIFICATION OF VARIABLE REGIONS NOT PRESENT IN *E. COLI*

As expected, all three rRNA genes could easily be identified because of the high degree of primary sequence conservation of these genes among living organisms. The *Frankia* strain AcN14a rRNA genes are almost identical to those of *Frankia* strain ORS020606 and share a higher degree of similarity to

those of another actinomycete species, namely *S. ambofaciens* (Pernodet *et al.*, 1989) than to those of more distantly related eubacteria.

We would also expect that the nucleotide sequence of both copies of the rRNA genes within an isolate to be identical. Supporting this hypothesis, the two 16S rRNA genes in AcN14a could not be distinguished in the region corresponding to positions 851 to 1175.

**16S ribosomal RNA**

The predicted 16S rRNA of *Frankia sp.* strain AcN14a resembles other eubacterial rRNA sequences. At the nucleotide sequence level, it shares 99, 92 and 91% identity to those of *Frankia sp.* strain ORS020606, *Streptomyces ambofaciens* and *Mycobacterium bovis*, respectively. The predicted size of *Frankia* AcN14a 16S rRNA is in fact, approximately 30 nucleotides smaller than those of more distantly related eubacteria such as *E. coli* and *B. subtilis*. Most of this size variation is accommodated by changes in the number of nucleotides present in three specific regions of the rRNA molecule, which encompass helices 6, 10-11 and 18 (Figures 10, 11, 12 and 14). In comparison to the 16S rRNA of *E. coli*, *Frankia* AcN14a 16S rRNA contains one addition and three deletion segments, all located in the first domain of the rRNA. Helices 6 and 18 of the *Frankia* AcN14a 16S rRNA are comparatively shorter than those present in *E. coli* whereas the sizes of two helices, helices 10 and 11, are interchanged. *Frankia* 16S rRNA contains an extended version of helix 10 and a shortened version of helix 11 compared to the *E. coli* counterpart. In this regard, *Frankia sp.* strain AcN14a 16S rRNA resembles that of *Frankia sp.* strain ORS020606, *Streptomyces ambofaciens* (Pernodet *et al.*, 1989), and those of other actinomycetes, which also contain deletion and addition segments in the corresponding regions of their 16S rRNAs. Although the 16S rRNAs of these organisms share similar secondary structures in these specific regions, many nucleotide substitutions are observed among the different genera.

Apart from *Frankia sp.* strain ORS020606 from which the complete rRNA operon sequence has been determined, only a small number of partial 16S rRNA sequences have been published for this domain of the *Frankia* 16S

rRNAs. In spite of the low number of sequences available, the nucleotide composition of helix 10 has been reported as highly variable among the members of this genus whereas that of helix 18 has been described as quite conserved (Harry *et al.*, 1991). Our results support their statements since the nucleotide sequence in helix 10 of the *Frankia* AcN14a 16S rRNA was different from those already published (Figure 13). *Frankia* AcN14a could be distinguished from the four isolates analyzed in that study since one to six nucleotide substitutions were observed in the region corresponding to helix 10. *Frankia* AcN14a nucleotide sequence is more similar to those of AcNI1, ORS020606 and PtI1, differing by three, three and one substitutions, respectively. In contrast, the nucleotide sequence of helix 18 is conserved among *Frankia* strains; only two variable positions have been identified (Figure 14). *Frankia* strains AcN14a, AcNI1 and PtI1 differ from Ag45/Mut 15 and AgB1.9 by only one substitution (Harry *et al.*, 1991), while they differ from ORS020606 by two substitutions. We have also compared the nucleotide sequence of helix 6 of the 16S rRNAs of strain AcN14a to that of strain ORS020606; two substitutions and an additional nucleotide not found in AcN14a were observed in ORS020606 (Figure 11).

These three regions have already been defined as variable by Raué *et al.* (1988), since most of the differences in size observed between the various eubacterial 16S rRNAs are also located in these helices. Thus, these regions are probably not involved in any ribosomal function.

All the regions found to vary among members of the genus *Frankia* could be used to design synthetic oligonucleotide probes in hybridization experiments in order to discriminate between these isolates. Of these three regions, that corresponding to helix 10 is the best candidate since it contains more variable positions than the two others. Interestingly, all the variable positions are located on one side of this helix.

Apart from these three regions where size differences between *Frankia* and *E. coli* are concentrated, we have analyzed two additional regions where the variation was limited to differences in primary sequence (Figures 15, 16,

17, and 18). We have chosen to analyze these regions because various *Frankia* isolates have been categorized based upon the differences in the primary sequence observed into these regions of their 16S rRNAs (Nazaret *et al.*, 1991). Comparison of *Frankia* AcN14a sequence to the partial sequences published by this latter group allow categorization of AcN14a in the first genomic species as defined by Nazaret et al. (1991), also called *Frankia alni* (Lalonde *et al.*, 1988). However, comparison of other regions of the 16S rRNA not determined for a large number of isolates, such as helix 10 will probably permit further delineation of the isolates since *Frankia* AcN14a and five other strains AcN1, Ag45/Mut15, AgB1.9, ORS020606 and PtI1 (Harry *et al.*, 1991) could be identified based upon the primary sequence of this region of their 16S rRNAs. Thus comparison of the primary structure of helix 10 of the strains AcN14a and AcNI1 of genomic species 1 further divides the genomic species into genomic subspecies. Further comparison analysis could determined if variation of the nucleotide composition of helix 10 could be sufficient to identify any *Frankia* strains or isolates.

**23S ribosomal RNA**

According to its gene sequence, the predicted 23S rRNA of *Frankia* AcN14a is longer than those of most eubacteria. In that respect, it resembles those of other actinomycetes, such as *Frankia sp.* strain ORS020606, *Micrococcus luteus* and *S. ambofaciens* with which it shares 98, 86 and 85 % identity, respectively. As in the case of the 16S rRNA, the size variation between the 23S rRNAs of *Frankia* AcN14a and *E. coli* can be traced to three specific variable regions, in domains I, II and III of the first half of the molecule (Figures 20, 21, 22, 23, and 24). Compared to *E. coli*, *Frankia* AcN14a 23S rRNA contains two additions of approximately 90 nucleotides and a smaller addition of 24 nucleotides. Similar additions are present in the 23S rRNA of other actinomycete species.

The first insertion segment is located in domain I and appears to be specific to actinomycetes since two other members of this group, *Streptomyces* and *Micrococcus* as well as the other *Frankia sp.* strain, ORS020606, also have similar insertion segments in the corresponding region of their 23S rRNAs.

Primary structure variation is observed between the insertion segments of the three actinomycete species but the overall secondary structure is conserved (Figure 21). The majority of the extra nucleotides present are probably present as a helical structure between helices 13 and 14, as well as an increase in the length of the latter helix. Thus this insertion segment was probably present in the common ancestor of these three genera. Actually, most prokaryotic 23S rRNAs contain extra nucleotides in this region compared to the *E. coli* molecule. However, the number of extra nucleotides present in the corresponding regions of prokaryotes is generally lower than those reported for the three actinomycetes.

The second large insertion has occurred in the upper half of an internal bulge present between helices 50 and 51 of the *E. coli* 23S rRNA. It can fold into a long irregular helical stem (Figure 23). Similar insertion segments have been reported for other actinomycetes such as *Streptomyces* (Pernodet *et al.*, 1989; Stackebrandt *et al.*, 1991), *Micrococci* (Höpfl *et al.*, 1988), *Mycobacteria* (Liesack *et al.*, 1990) and *Pseudonocardia* (Stackebrandt *et al.*, 1991). The primary structure of this insertion varies greatly among the different actinomycete species and a common secondary structure model could not be proposed (Figure 22). However, we expect the various *Frankia* isolates to share higher degrees of homologies. This is supported by the detection of a common restriction site for BamH1 in numerous isolates. Once again, the fact that similar insertion segments are present in different groups of actinomycetes suggests that this is a feature specific to actinomycetes. Primary sequence variation would then reflect the number of mutations accumulated since these groups diverged.

In addition, helix 41 of *Frankia* AcN14a 23S rRNA was found to contain 24 extra nucleotides not present in *E. coli*. In contrast to the other insertion segments described in the 23S rRNA of *Frankia* AcN14a, no such extended version of helix 41 are observed in *Streptomyces* or *Micrococci* (Figure 24). The corresponding region of *Frankia* strain ORS020606 has a similar length but contains many substitutions when compared to its counterpart. Thus the

length of helix 41 could be genus-specific. Analysis of additional *Frankia* 23S rRNA sequences needs to be done to confirm this hypothesis.

### 5S ribosomal RNA

In contrast to *Frankia* AcN14a 16S and 23S rRNAs predicted to be somewhat different from the typical eubacterial rRNAs, *Frankia* AcN14a 5S rRNA shows no outstanding deviation from the eubacterial 5S rRNA sequence. *Frankia* AcN14a 5S rRNA is 120 nucleotides in length and can be folded into the secondary structure model proposed by Erdmann and Wolters (1986) (Figure 25). Comparison of *Frankia sp.* AcN14a 5S rRNA sequence to that of *Frankia sp.* strain ORS020606 5S rRNA sequence revealed that the two sequences were almost identical; only one residue, at position 120, was found to vary between the two organisms (Figure 26).

## POSSIBLE USES OF RIBOSOMAL RNA GENE SEQUENCES

As mentioned in the previous section, size variation has been observed in some regions of both the 16S and 23S rRNAs of *Frankia* AcN14a compared to those of *E. coli* models. Some other regions where variation was limited to the primary structure have also been identified. Since all these regions are located in areas already defined as variable (Raué *et al.*, 1988), it was thus not surprising to find that some of the regions first identified for their size differences also exhibit variable primary structures among actinomycetes and even among *Frankia* strains. Such poorly conserved regions are often used to allocate isolates to lower taxonomic ranks by sequence analysis or oligonucleotide probing (Stackebrandt *et al.*, 1991). We have identified the more promising regions for the targetting of strain- as well as species-specific oligonucleotide probes. We have also evaluated the usefulness of some of the variable regions identified in the *Frankia* 16S rRNA for the construction of a phylogenetic tree (Figure 19).

**Design of probes for the detection or identification of *Frankia***

Some regions of the 16S and 23S rRNAs showing promising degrees of variation have been identified. While some of these regions are predicted to be good targets for universal probes, others are expected to be specific to a single or a limited number of *Frankia* strains. Analysis of the region corresponding to helix 10 of the 16S rRNA revealed high levels of nucleotide divergence among *Frankia* strains (Harry *et al.*, 1991, Hahn *et al.*, 1989a, Hahn *et al.*, 1990). This region could be a good target site for a strain-specific oligonucleotide probe since six strains (AcN1, AcN14a, Ag45/Mut15, AgB1.9, ORS020606 and PtI1) could be distinguished by the nucleotide sequence of the helix 10 of their 16S rRNAs (Figure 13). Two of these strains, AcN14a and AcN1 belong to the same genomic species based upon the conservation of the nucleotide sequence between positions 886 and 1155 of their 16S rRNAs (Nazaret *et al.* 1991). Other variable regions of the 16S rRNAs analyzed in this thesis were found to vary to a different extent among *Frankia* strains. Based upon the limited sequence data available, the region corresponding to helices 6 and 18 appear as quite well conserved among *Frankia*, whereas regions corresponding to the helix complex P-35 and helix 41 show some intrageneric variation (Figures 11, 14, 15, and 17). Nazaret *et al.* (1991) have successfully categorized several *Frankia* strains into genomic species according to the nucleotide sequence of the portion of the 16S rRNA containing the last two variable regions. The corresponding regions have also been successfully used to identify *Streptomyces* isolates by Stackebrandt *et al.* (1991). Thus, species-specific oligonucleotides having these regions of the 16S rRNA as targets could be designed.

In recent years, oligonucleotide probes directed against 16S rRNA targets have played an important role in the identification and detection of microorganisms (e.g. Hahn *et al.*, 1990; Rossau *et al.*, 1989). In contrast, the potential of 23S rRNA sequences has yet rarely been used for the design of such probes. This was mainly due to the limited number of 23S rRNA sequences available. Comparison of *Frankia* AcN14a 23S rRNA sequence to other eubacterial sequences including some other actinomycetes as well as that of another *Frankia* strain, ORS020606 allowed identification of three insertion

segments compared to the *E. coli* model. Two of these large insertions have been proposed to be present in most if not all actinomycetes (Liesack *et al.*, 1990; Stackebrandt *et al.*, 1991). A highly specific oligonucleotide probe targetting a stretch contained in the domain III insertion segment has been developed for *Mycobacterium leprae* (Liesack *et al.*, 1990). We have also identified another insertion in helix 41 which is apparently only present in *Frankia*. These three regions are expected to be good candidates for the design of probes.

The specificity of such probes should be investigated by computer analysis and then tested by dot blot hybridization. We could also document the influence of the washing temperature on the specificity of the probes. Such probes usually discriminate between homologous and heterologous targets which differ by as few as one nucleotide (out of 15 to 25 nucleotides) (Stackebrandt *et al.*, 1991). These oligonucleotides can be synthesized with radioactive or fluorescent tags, allowing the detection and identification of specific *Frankia* strains or isolates and would thus facilitate the study of competitive interactions among *Frankia* strains, as well as other aspects of *Frankia* ecology.

**Construction of phylogenetic trees**

We have constructed two unrooted phylogenetic trees showing the relationships between *Frankia* AcN14a and other selected actinomycetes (Figure 19). The first tree is based upon complete or nearly complete 16S rRNA sequences of selected actinomycetes whereas the second has been constructed using regions corresponding to four of the five variable regions analyzed in the 16S rRNA of *Frankia* AcN14a. The variable region located near position 441 of *Frankia* AcN14a was not included in the second tree since it was also absent from most actinomycete sequences included in the analysis. Comparison of both trees revealed that all the distances calculated for the tree based upon the variable regions are larger than those calculated for the other tree. This reflects the different substitution rates of the type of sequences used in the analyses; variable regions of the 16S rRNAs change at a high rate. In spite of the different branching observed in the two trees, analysis of the complete as

well as the variable regions allows clustering of the different species of a given genus; members of the the genera *Frankia*, *Streptomyces* and *Mycobacterium* form three well separated groups in both trees. In addition, both trees show a close relationships between *Arthrobacter* and *Micrococcus*. However, the relationships between these three genera, the Arthrobacter-Micrococcus group as well as their relationships with more distantly related actinomycete species differ in both trees. As predicted for trees based upon almost complete 16S rRNA sequences published by Embley *et al.* (1988), Hahn *et al.* (1989b) and Stackebrandt and Charfreitag (1990), the tree based upon complete 16S rRNA sequences groups the mycolic acid-containing taxa in a cluster. The second tree based upon the 16S rRNA variable regions only, does not. For this reason, we consider trees based upon complete or nearly complete 16S rRNA sequences to be a more accurate reflection of the phylogeny of actinomycetes. A more detailed analysis with a larger number of actinomycetes need to be done since phylogenetic trees are dynamic, their topology changing when a larger selection of different genera is used (Stackebrandt and Charfreitag, 1990).

## CONSERVATION OF THE RIBOSOMAL RNA SECONDARY STRUCTURE - CONSEQUENCE OF THEIR FUNCTIONAL ROLE IN TRANSLATION

As in the case of the 16S rRNA, the structural variation among 23S rRNAs can be traced to a limited number of specific variable regions leaving a common core structure that must contain the elements indispensable for proper functioning of the rRNAs within the ribosome. Although no experimental attempts have been made to study post-translational modification and the tertiary structure of the *Frankia* AcN14a rRNAs, comparison of *Frankia* rRNA nucleotide sequences to those of *E. coli* and other prokaryotic sequences enables us to make some predictions.

To illustrate that point, conservation of the residues shown to be post-transcriptionally modified in *E. coli* 16S and 23S rRNAs was evaluated. Even though the methylation state of the different positions of *Frankia* rRNA has not been determined, conservation of the nucleotide composition at the

positions known to be modified in *E. coli* rRNAs suggests a similar modification patterns in the two species. Most of the twenty-three positions post-transcriptionally modified in *E. coli* were found to be conserved between the two eubacterial species. Only two residues, the two cytosine residues at positions 1171 and 1481 of *Frankia* AcN14a 16S rRNA gene were found to have replaced the two guanosine residues present in the corresponding positions of the *E. coli* molecule. In *E. coli*, these two positions are modified differently; the ribose component of residue 1207 (*E. coli* numbering) is methylated whereas the base component of residue 1515 is methylated. *Frankia* AcN14a is not the only organism having such substitutions at these two positions; at least seven other actinomycetes, such as other *Frankia* strains (e.g. Hahn *et al.*, 1989b), *Streptomyces* (e.g. Pernodet *et al.*, 1989), *Mycobacteria* and *Micrococci* have the same sequence. However, the substitution of the guanosine at position 1207 (*E. coli* numbering) to a cytosine or a uracil residue is lethal in *E. coli*, whereas substitution to an adenosine residue has no effect on its growth rate (Jemiolo *et al.*, 1991). The methylation state of these two cytosine residues in the 16S rRNAs of *Frankia* is not known.

Comparison of *Frankia* AcN14a rRNAs to other rRNA sequences has also allowed us to evaluate each of the tertiary interactions proposed for the 16S and 23S rRNAs (Table 5). By doing so, we have found that *Frankia* AcN14a rRNA sequences were generally in concordance with the proposed tertiary interactions; six of the seven contacts proposed by Leffers *et al.* (1987) could be formed within the 16S and 23S rRNAs of *Frankia* AcN14a. A single mismatch ($U_{67}/U_{74}$) was found within *Frankia* AcN14a 23S rRNA, disproving the putative interaction of residues 67 and 74. Such a mismatch is also observed in the 23S rRNA of *Frankia* strain ORS020606. Although disproven in eukaryotic sequences, this interaction is supported by many compensatory substitutions observed among prokaryotic sequences (Höpfl *et al.*, 1987). Even in actinomycetes species, such as *S. ambofaciens* and *M. luteus*, considered as close relatives of *Frankia*, the pairing of the corresponding residues is possible. The pairing of a cytosine to a guanosine residue (C-G) and the pairing of an uracil to an adenosine residue (U-A) can be formed in *S. ambofaciens* and *M. luteus* 23S rRNAs, respectively. In some eukaryotes, the presence of an

uracil and a cytosine at the corresponding positions does not support this interaction (Höpfl *et al.*, 1987). However, the secondary structure of this region in eukaryotes differs from that of prokaryotes and thus, a slightly different structure for eukaryotes remains a possibility. By a comparison of *Frankia* rRNA sequences to those of other eubacteria we predict the tertiary folding of *Frankia* AcN14a rRNAs to be similar to that of *E. coli* rRNA molecules. However, the observation of a non-compensatory substitution in the 23S rRNA of two *Frankia* strains disproves the tertiary contact proposed for that region.

## RFLP ANALYSIS OF *FRANKIA* ISOLATES

We also have evaluated the potential usefulness of rDNA RFLP analysis to categorize *Frankia* isolates (Tables 6, 7, and 8). The rDNA restriction patterns were detected using BamH1 fragments of *Frankia sp.* AcN14a or ArI3 rDNA as probes. These fragments contained 16S and/or 23S rDNA sequences. Hybridization of both 16S and 23S rDNA specific probes to total DNA digested with BamH1, revealed sufficient polymorphism to permit classification of twenty-six isolates into six groups (A to F). In addition, eight other isolates exhibited unique patterns. All these isolates, except one contain two copies of the rDNA region. Isolates #33 probably contains three rRNA operons.

Sequencing of the two copies of the 23S rRNA genes of *Frankia* AcN14a revealed that a BamH1 site was present in both 23S rRNA genes. Similarly, we expect a BamH1 site to be present in the corresponding region of most if not all members of the genus *Frankia* because the recognition site is located within the rRNA sequence. Thus, the observed RFLPs result of mutation in the BamH1 sites located outside the rDNA regions. Based on the assumption that the internal BamH1 site is conserved among *Frankia*, a possible relationships has been proposed for most of the six hybridization patterns. Five hybridization patterns, corresponding to groups A to E, could be related to each other by a succession of simple mutational events (Figure 29). In addition, the hybridization pattern of group F could be derived from that of group D or E by two mutations.

Thus from this preliminary study we can conclude that sufficient rDNA polymorphism exists to categorize *Frankia* isolates into groups. A more elaborate analysis with a larger number of isolates remains to be done. The assumption is that more closely related isolates are likely to share similar rDNA hybridization patterns as well as competitiveness and symbiotic activity characteristics. A reduced number of isolates, representatives of each group would have to be tested, thus eliminating the duplication of strains selected for extensive field trials.

Because it has not been possible to trace the strains used in this study to their site of origin, the possibility that the RFLP patterns reflects geographic origin cannot be assessed. Neither can be assessed, the relationships between isolates and species, since species delineation has not been defined. However, the fact that both *Frankia* AcN14a and ArI3, both categorized in the *Frankia alni* species (genomic species 1) by partial sequencing of their 16S rRNA gene (Nazareth et al., 1991) share an identical rDNA RFLP profile suggests that the rDNA hybridization pattern may be species-specific. This is observed in Staphylococci for which species-specific rDNA hybridization profiles have been detected. Different Staphylococcal strains of a particular species were found to share an identical or similar hybridization pattern (Thomson-Carter *et al.*, 1989). If the latter is true for members of the genus *Frankia*, strains and isolates of the same species would also share a similar rDNA hybridization pattern. This would have to be tested by rDNA RFLP analysis of many *Frankia* strains and isolates that have had their 16S rRNA gene already partially sequenced. If the latter is true, one could them assigned new isolates to a given species, based on the rDNA RFLP hybridization patterns.

## REFERENCES

Amikan, D., Razin,S., and Glaser, G. (1982). Ribosomal RNA genes in *Mycoplasma*. Nucl. Acids. Res. 10, 4215-4221.

An, C.S., Riggsby, W.S., and Mullin, B.C. (1985). Relationship of *Frankia* isolates based on deoxyribonucleic acid homology studies. Int. J. Syst. Bacteriol. 35: 140-146.

An, C.S., Wills, J.W., Riggsby, W.S., and Mullin, B.C. (1983). Deoxyribonucleic acid base composition of twelve *Frankia* isolates. Can J. Bot. 61, 2859-2862.

Bacot, C.M. and Reeves, R.H. (1991). Novel tRNA gene organization in the 16S-23S intergenic spacer of the *Streptococcus pneumoniae* rRNA gene cluster. J. Bacteriol. 173, 4234-4236.

Baird, S.D. and Johnson, D.A. (1991). Determination of the sizes of DNA fragments cloned into M13 without DNA purification. Biotechniques 10(2), 180.

Baker, D. (1987). Relationships among pure cultured strains of *Frankia* based on host-specificity. Physiol. Plant. 70, 245-248.

Baylis, H.A. and Bibb, M.J. (1987). The nucleotide sequence of a 16S rRNA gene from *Streptomyces coelicolor* A3(2). Nucl. Acids Res. 15, 7176.

Baylis, H.A. and Bibb, M.J. (1988a). Organisation of the ribosomal RNA genes in *Streptomyces coelicolor* A3(2). Mol. Gen. Genet. 211, 191-196.

Baylis, H.A. and Bibb, M.J. (1988b). Transcriptional analysis of a rRNA gene set (*rrn* D) from *Streptomyces coelicolor* A3(2). Mol. Microbiol. 2, 569-579.

Beaupied, H., Moiroud, A., Domenach, A-M., Kurdali, F. and Lensi, R. (1990). Ratio of fixed and assimilated nitrogen in a black alder (*Alnus glutinosa*) stand. Can. J. For. Res. 20, 1116-1119.

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

Binkley, D. (1981). Nodule biomass and acetylene reduction rates of red alder and Sitka alder on Vancouver Island, B.C. Can. J. For. Res. 11, 281-286

Bloom, R.A., Mullin, B.C., and Tate III, R.L. (1989). DNA restriction patterns and DNA-DNA solution hybridization studies of *Frankia* isolates from *Myrica pensylvanica* (Bayberry). Appl. Environ. Microbiol. 55: 2155-2160.

Boettger, E.C. (1989) Rapid determination of bacterial ribosomal RNA sequences by direct sequencing of enzymatically amplified DNA. FEMS Microbiol. Lett. 65, 171-176.

Bonen, L. and Doolittle, W.F. (1975). On the prokaryotic nature of the red algal chloroplasts. Proc. Natl. Acad. Sci. 72, 2310-2314.

Bousquet, J., Cheliak, W.M. and Lalonde M. (1987). Genetic diversity within and among 11 juvenile populations of green alder (*Alnus crispa*) in Canada. Physiol. Plantarum 70, 311-318.

Bram, R.J., Young, R.A. and Steitz, J.A. (1980). The ribonuclease III site flanking 23S sequences in the 30S ribosomal precursor RNA of *E. coli*. Cell 19, 393-401.

Brimacombe, R., Atmadja, J., Stiege, W. and Schüler, D. (1988). A detailed model of the 3-dimensional structure of *Escherichia coli* 16S ribosomal RNA *in situ* in the 30S subunit. J. Mol. Biol. 199, 115-136.

Brimacombe, R., Greuer, B., Mitchell, P., Osswald, M., Rinke-Appel, J., Schüler, D. and Stade K. (1989). Three dimensional structure and function of *Escherichia coli* 16S and 23S rRNAs as studied by cross-linking techniques, pp. 93-106. *In* W.E., Hill, A. Dahlberg, R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology, Washington, D.C.

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

Brozek, S. (1990). Effect of soil changes caused by red alder (*Alnus rubra*) on biomass and nutrient status of Douglas-fir (*Pseudotsuga menziesii*) seedlings. Can. J. For. Res. 20, 1320-1325.

Chen, X. and Finch, L.R. (1989). Novel arrangement of rRNA genes in *Mycoplasma gallisepticum*: separation of the 16S gene of one set from the 23S and 5S genes. J. Bacteriol. 171, 2876-2878.

Cundliffe, E. (1989). How antibiotic-producing organisms avoid suicide. Ann. Rev. Microbiol. 43, 207-233.

Dale, R.M.K., McClure, B.A., Houchins, J.P. (1985). A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid 13: 31-40.

Denhardt, D.T. (1966). A membrane filter technique for the detection of complementary DNA. Biochem Biophys Res Comm 23, 641-646.

Duester, G.L. and Holmes, W.M. (1980). The distal end of the ribosomal RNA *rrn* D of *Escherichia coli* contains a tRNA $^{Thr}$ gene, two 5S rRNA genes and a transcription terminator. Nucl. Acids Res. 8, 3793-3807.

Embley, M.T., Smida, J. and Stackebrandt, E. (1988). The phylogeny of mycolate-less wall chemotype IV actinomycetes and description of *Pseudonocardiaceae* fam. nov. System. Appl. Microbiol. 11, 44-52.

Engel, J.N. and Ganem, D. (1987). Chlamydial rRNA operons: gene organization and identification of putative tandem promoters. J. Bacteriol. 169, 5678-5685.

Erdmann, V.A. and Wolters, J. (1986). Collection of published 5S, 5.8S and 4.5S ribosomal RNA sequences. Nucl. Acids Res. 14(supplement), r1-r60.

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

Fernandez, M.P., Meugnier, H., Grimont, P.A.D. and Bardin., R. (1989). Deoxyribonucleic acid relatedness among members of the genus *Frankia*. Int. J. System. Bacteriol. 39, 424-429.

Fox, G.E., Stackebrandt, E., Hespell, R.B., Gibson, J., Maniloff,J., Dyer, T.A., Wolfe, R.S., Balch, W.E., Tanner, R.S., Magrum, L.J., Zablen, L.B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B.J., Stahl, D.A., Luehrsen, K.R., Chen, K.N. and Woese, C.R. (1980). The phylogeny of prokaryotes. Science 209: 457-463.

Fukunaga, M. and Mifuchi, I. (1989). Unique organization of *Leptospira interrogans* rRNA genes. J. Bacteriol. 171, 5763-5767.

Garrett, R.A., Dalgaard, J., Larsen, N., Kjems, J., and Mankin, S. (1991). Archeal rRNA operons. TIBS 16,22-26.

Gleeson, T.D. (1992). An attempt to isolate nodule-specific cDNA clones from *Alnus incana*, M.Sc. thesis, University of Ottawa, Ottawa.

Green, C.J., Streward, G.C., Hollis, M.A., Vold, B.S. and Bott, K.F. (1985). Nucleotide sequence of *Bacillus subtilis* ribosomal RNA operon, *rrn*B. Gene 37, 261-266.

Gutell, R.R. and Fox, G.E. (1988). A compilation of large subunit RNA sequences presented in a structural format. Nucl. Acids Res. 16, r175-r269.

Gutell, R.R. and Woese, C.R. (1990). Higher order structural elements in ribosomal RNAs: preudo-knots and the use of noncanonical pairs. Proc. Natl. Acad. Sci. 87, 663-667.

Haas, E.S., Brown, J.W., Daniels, C.J. and Reeve, J.N. (1990). Genes encoding the 7S RNA and $tRNA^{Ser}$ are linked to one of the two rRNA operons in the genome of the extremely thermophilic archeobacterium *Methanothermus fervidus*. Gene 90, 51-59.

Hahn, D., Dorsch, M., Stackebrandt, E. and Akkermans, A.D.L. (1989a). Synthetic oligonucleotide probes for identification of *Frankia* strains. Plant and Soil 118, 211-219.

Hahn, D., Lechevalier, M.P., Fischer, A. and Stackebrandt, E.(1989b). Evidence for a close phylogenetic relationship between members of the genera *Frankia, Geodermatophilus*, and "*Blastococcus* " and emendation of the family *Frankiaceae*. System. Appl. Microbiol. 11, 236-242.

Hahn, D., Starrenburg, M.J. and Akkermans, A.D.L. (1990). Oligonucleotide probes that hybridize with rRNA as a tool to study *Frankia* strains in root nodules. Appl. Environ. Microbiol. 56, 1342-1346.

Hall, R.B. and Burgess. (1990). Evaluation of *Alnus* species and hybrids. Biomass 22, 21-34.

Hanahan, D. (1985). Techniques for transformation of *E. coli*,, pp.109-135, *In* D.M. Glover (ed.), DNA cloning volume 1: a practical approach, IRL Press, Oxford, Washington DC.

Harry, D.E., Yang, D.C. and Dawson, J.O. (1991). Nucleotide sequence and diversity in 16S ribosomal RNA from *Frankia*. Plant and Soil 131, 143-146.

Hartmann, R.K. and Erdmann, V.A. (1989). *Thermus thermophilus* 16S rRNA is transcribed from an isolated transcription unit. J. Bacteriol. 171, 2933-2941.

Hartmann, R.K., Ulbrich, N. and Erdmann, V.A. (1987) An unusual rRNA operon constellation: in *Thermus thermophilus* HB8 the 23S\5S rRNA operon is a separate entity from the 16S rRNA operon. Biochimie 9, 1097-1104.

Heilman, P.E. (1982). Rates of nitrogen and organic matter accumulation in coal mine spoils supporting alder stand. Can. J. For. Res. 12, 809-813.

Heilman, P.E. (1989). Growth and N status of Populus in mixture with red alder on recent volcanic mudflow from Mount Saint Helens. Can. J. For. Res. 20, 84-90.

Hooker, J.E. and Wheeler, C.T. (1987). The effectivity of *Frankia* for nodulation and nitrogen fixation in *Alnus rubra* and *A. glutinosa*. Physiol. Plantarum 70, 333-341.

Höpfl, P., Ludwig, W., Schleifer, K.H. and Larsen, N. (1989). The 23S ribosomal RNA higher-order structure of *Pseudomonas cepacia* and other prokaryotes. Eur. J. Biochem. 185, 355-364.

Hunter, I.S. (1985). Gene cloning in *Streptomyces*, pp. 19-44, *In* Glover, D.M. (ed), DNA cloning: a practical approach, vol. II, IRL Press Limited, Washington, D.C.

Innis, M.A. and Gelfand, D.H. (1990). Optimization of PCRs, pp. 3-12, *In* Innis, M.A., Gelfand, D.H., Sninsky, J.J., and White, T.J. (eds), PCR protocols; a guide to methods and applications, Academic Press Inc.,

Jemiolo, D.K., Taurence, J. S. and Giese, S. (1991). Mutations in 16S rRNA in *Escherichia coli* at methyl-modified sites: G966, C967, and G1207. Nucl. Acids Res. 19, 4259-4264.

Jukes, T.H. and Cantor, C.R. (1969) Evolution of protein molecules, pp 21-132, *In* Munro, H.N., Mammalian protein metabolism, Acadamic Press, New York.

Kaiser, K. and Murray, N.E. (1985). The use of phage lambda replacement vectors in the construction of representative genomic DNA libraries, pp.1-48, *In* Glover, D.M. (ed), DNA cloning: a practical approach, vol. I, IRL Press Limited, Washington, D.C.

Kim, E., Kim, H., Kang, K.E., Kho, Y.H. and Park, Y-H. (1991). Complete nucleotide sequence of a 16S ribosomal RNA gene from *Streptomyces griseus* subsp. *griseus*. Nucl. Acids Res. 19, 1149.

King, T.C., Sirdeskmukh, R. and Schlessinger, D. (1986). Nucleolytic processing of ribonucleic transcripts in prokaryotes. Microbiol. Rev. 50, 428-451.

James, B.D., Olsen, G.J. and Pace N.R. (1989). Phylogenetic comparative analysis of RNA secondary structure. Meth. Enzymol.180, 227-239.

Lalonde, M., Simon, L., Bousquet, J. and Séguin, A. (1988). Advances in the taxonomy of *Frankia*: recognition of species *alni* and *elaeagni* and novel subspecies *pommerii* and *vandijkii*. p. 671-680. In Bothe/ de Bruijn/Newton (eds.), Nitrogen fixation: hundred years after. Gustav Fisher, Stuttgart, New York.

Lane, D.J., Pace, B., Olsen, G.J., Stahl, D.A., Sogin, M.L. and Pace, N.R. (1985). Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA 82, 6955-6959.

Lechevalier, M.P. (1984). The taxonomy of the genus *Frankia*. Plant and Soil 78, 1-6.

Leffers, H., Kjems, J., Ostergaard, L., Larsen, N. and Garrett, R.A. (1987). Evolutionary relationships amongst Archeobacteria. A comparative study of 23S ribosomal RNAs of sulphur-dependent extreme thermophilic methanogen. J. Mol. Biol. 195, 43-61.

Lehrach, H.D., Diamond, D., Wozney, J.M. and Boedtker, H. (1977). RNA molecular weight determinations by gel electrophoresis under denaturing conditions, a critical reexamination. Biochem. 16, 4743-4751.

Lévesque, M. (1990). Investigation of alder (*Alnus incana*) chloroplast ribosomal RNA genes, M.Sc. thesis, University of Ottawa, Ottawa.

Lewin, B. (1987). Control at termination: attenuation and antitermination. *Genes*, third edition. John Wiley & sons (ed.), New York, 248-268.

Liesack, W., Pitulle, C. and Stackebrandt. (1990). Development of a highly specific diagnostic 23S rDNA oligonucleotide probe for *Mycobacterium leprae*. Lett. Appl. Microbiol. 11, 96-99.

Ligon, J.M. and Nakas, J.P. (1987). Isolation and characterization of *Frankia* sp. strain FaC1 genes involved in nitrogen fixation. Appl. Environ. Microbiol. 53, 2321-2327.

Loughney, K., Lund, E. and Dahlberg, J.E. (1983). Ribosomal RNA precursors of *Bacillus subtilis*. Nucl. Acids Res. 11, 6709-6721.

Maniatis, T., Fritsch, E.F., Sambrook, J. (1982). Molecular cloning. Cold Spring Harbor laboratory: Cold Spring Harbor, N.Y.

Messing, J., Crea, R., Seeburg, P.H. (1981). A system for shotgun DNA sequencing. Nucl. Acids Res. 9, 309-322.

Moazed, D. and Noller, H.F. (1987). Chloramphenicol, erythromycin, carbomycin and vernamycin B protect overlapping sites in the peptidyl transferase region of 23S ribosomal RNA. Biochimie 69, 879-884.

Moazed, D., Stern, S. and Noller, H.F. (1986). Rapid chemical probing of conformation of 16S ribosomal RNA and 30S ribosomal subunits using primer extension. J. Mol. Biol. 187, 399-416.

Morelle, G. (1989). A plasmid extraction procedure on a miniprep scale. BRL Focus 11(1), 7-8.

Mullin, B.C., Swensen, S.M. and Goetting-Mineski, P. (1990). Hypothesies for the evolution of actinorhizal symbioses, p.781-787. *In* Gresshoff, Roth, Stacey and Newton (eds.) Nitrogen fixation: achievements and objectives. Chapman and Hall, New York/London.

Murray,P.A. and Zinder, S.H. (1984). Nitrogen fixation by a methanogenic archeobacterium. Nature 312, 284-286.

Nazaret, S., Cournoyer, B., Normand, P. and Simonet, P. (1991). Phylogenetic relationships among *Frankia* genomic species determined by use of amplified 16S rDNA sequences. J. Bacteriol. 173, 4072-4078.

Neefs, J-F., Van de Peer, Y., De Rijk, P., Goris, A. and De Wachter, R. (1991). Compilation of small ribosomal subunit RNA sequences. Nucl. Acids Res. 19, 1987-2004.

Newton, W.E. and Burgess, B.K. (1981). Nitrogen fixation: its scope and importance, pp. 1-23. *In* Müller, A. and Newton, W.E. (Eds.), Nitrogen fixation: the chemical-biochemic-genetic interface, Plenum Press, New York.

Nittayajarn, A. Mullin, B.C. and Baker, D.D. (1990). Screening of Symbiotic *Frankia* for host specificity by restriction fragment length polymorphism analysis. Appl. and Environ. Microbiol. 56, 1172-1174.

Noller, H.F. (1984). Structure of ribosomal RNA. Ann. Rev. Biochem.53, 119-162

Noller, H.F., Moazed, D., Stern, S., Powers, T., Allen, P.N., Robertson, J.M., Weiser, B. and Triman, K. (1989). Structure of rRNA and its functional interaction in translation, pp. 73-92. *In* W.E., Hill, A. Dahlberg, R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology., Washington, D.C.

Normand, P. and Lalonde, M. (1986). The genetics of actinorhizal *Frankia* : a review. Plant Soil 90, 429-453.

Normand, P., Simonet, P. and Bardin, R. (1988). Conservation of *nif* sequences in *Frankia*. Mol. Gen. Genet. 213, 238-246.

Normand,P., Simonet, P., Butour, J.L., Rosenberg, C., Moiroud, A and Lalonde, M. (1983). Plasmids in *Frankia*. J. Bacteriol. 155, 32-35.

Ogasawara, N.,Moriya, S. and Yoshikawa, H. (1983). Structure and organization of rRNA operons in the region of the replication origin of the *Bacillus subtilis* chromosome. Nucl. Acids Res. 11; 6301-6317.

Olsen, G.J., Lane, D.J.,Giovannoni, S.J., Pace N.R., and Stahl, D.A. (1986). Microbial ecology and evolution: a ribosomal RNA approach. Ann. Rev. Microbiol. 40: 337-365.

Pace, N.R. (1984). Protein-polynucleotide recognition and the RNA processing nucleases in prokaryotes, pp.1-30. *In* D. Apirion (ed.). Processing of RNA. CRC Press, Inc., Boca Raton, Fla.

Pace, N.R. and and Burgin, A.B. (1989). Processing and evolution of the rRNAs, pp.417-425. *In* W.E., Hill, A. Dahlberg, R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology., Washington, D.C.

Pace, N.R. and Pace, B. (1990). Ribosomal RNA terminal maturase: ribonuclease M5 from *Bacillus subtilis*. Methods Enzymol. 181, 366-374.

Pernodet, J-L., Boccard, F.,Alegre, M-T., Gagnat, J. and Guérineau. (1989). Organization and nucleotide sequence analysis of a ribosomal RNA gene cluster from *Steptomyces ambofaciens* . Gene 79, 33-46.

Pezeshki, S.R. and Chadwick, D.O. (1985). Early growth patterns of red alder and black cottonwood in mixed species plantations. Forest Sci. 31, 1990-200.

Pleij, C.W.A. (1990). Pseudoknots: a new motif in the RNA game. TIBS 15, 143-147.

Plohl, M. and Gamulin, V. (1991). Sequence of the 5S rRNA gene and organization of ribosomal RNA operons in *Streptomyces rimosus*. FEMS Microbiol. Lett. 77, 139-144.

Raleigh, E.A. (1987). Restriction *in vivo* by *E. coli* K12. Met. Enzymol. 152, 130-144.

Rasmussen, O.F., Frydenberg, J. and Christiansen, C. (1987). Analysis of the leader and spacer regions of the two rRNA operons of *Mycoplasma* PG50: two tRNA genes are located upstream of *rrn*A. Mol. Gen. Genet. 208, 23-29.

Raué, H.A., Klootwijk, J., Musters, W. (1988). Evolutionary conservation of structure and function of high molecular weight ribosomal RNA. Prog. Biophys. mol. Biol. 51, 77-129.

Raué, H.A., Musters, W., Rutgers, C.A., Riet, J.V. and Planta, R.J. (1989). rRNA: from structure to function, pp. 217-235. *In* W.E., Hill, A. Dahlberg, R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology., Washington, D.C.

Regensburger, A., Ludwig, W., Frank, R., Blöcker, H. and Schleifer, K. (1988). Complete nucleotide sequence of a 23S ribosomal RNA gene from *Micrococcus luteus* . Nucl. Acids Res. 16, 2344.

Rochefort, D.A. and Benson, D.R. (1990). Molecular cloning, sequencing, and expression of the glutamine synthase II (*glnII*) gene from the actinomycete root nodule symbiont *Frankia* sp. strain CpI1. J. Bacteriol. 172, 5335-5342.

Rossau,R., van Mechele, E., de Ley, J. and van Heuverswijn, H. (1989). Specific *Neisseria gonorrhoeae* DNA-probes derived from ribosomal RNA. J. Gen. Microbiol. 135, 1735-1745.

Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol. Evol. 4, 406-425.

Sanger, F., Nicklin, S., Coulson, A.R., (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci., 74, 5463-5467.

Schubert, K.R. (1986). Products of biological nitrogen fixation in higher plants: synthesis, transport, and metabolism. Ann. Rev. Plant Physiol. 37, 539-574.

Siliciano, P.G., Jones, M.H. and Guthrie. (1987). *Saccharomyces cerevisiae* has a U1-like small nuclear RNA with unexpected properties. Science 237, 1484-1487.

Silvester, W.B. (1977). Dinitrogen fixation by plant associations excluding legumes, pp. 141-190. In R.W.F. Hardy, A.H. Gibson (eds.), A treatise on dinitrogen fixation, Sect. 4. Wiley, New York.

Simonet, P., Normand, P., Moiroud, A. and Lalonde, M. (1985). Restriction enzyme digestion patterns of *Frankia* plasmids. Plant and Soil 87, 49-60.

Skurnik, M. and Toivanen, P. (1991). Intervening sequences (IVSs) in the ribosomal RNA genes of pathogenic *Yersinia enterocolitica* strains. The IVS in *Y. enterocolitica* and *Salmonella typhimurium* have a common origin. Mol. Microbiol. 5, 585-593.

Smolander, A. (1990). *Frankia* populations in soils under different tree species - with special emphasis on soils under *Betula pendula*. Plant and Soil 121, 1-10.

Specht, T., Wolters, J. and Erdmann, V.A. (1990). Compilation of 5S rRNA and 5S rRNA gene sequences. Nucleic Acids Res. 18 (suppl.), 2215-2230.

Sprent, J.I. and Sprent, P. (1990). Nitrogen fixing organisms: pure and applied aspects Chapman and Hall (eds.), 256 p., London.

Sprinzl, M., Hartmann, T., Meissner, F., Moll, J. and Vorderwülbecke. (1989). Compilation of tRNA sequences and sequences of tRNA genes. Nucl. Acids Res.17, suppl., r1-r172.

Srivastava, A.K. and Schlessinger, D. (1989). rRNA processing in Escherichia coli, pp.426-434. *In* W.E., Hill, A. Dahlberg, R.A. Garrett, P.B. Moore, D. Schlessinger, and J.R. Warner (ed.), The ribosome: structure, function and evolution. American Society for Microbiology., Washington, D.C.

Srivastava, A.K. and Schlessinger, D. (1990). Mechanism and regulation of bacterial ribosomal RNA processing. Annu. Rev. Microbiol. 44, 105-129.

Stackebrandt, E. and Charfreitag, O. (1990). Partial 16S rRNA primary structure of five *Actinomyces* species: phylogenetic implication and development of an *Actinomyces israelii* -specific oligonucleotide probe. J. Gen. Microbiol. 136, 37-43.

Stackebrandt, E. and Smida, J. (1988). The phylogeny of the genus *Mycobacterium* determined by 16S rRNA sequences, and development of DNA probes, p. 244-250. *In* Okami, Y., Beppu, T. and Ogawara, H., Biology of actinomycetes. Japan Scientific Societies Press, Tokyo.

Stackebrandt, E., Witt, D., Kemmerling, C., Kroppenstedt, R. and Liesack, W. (1991). Designation of Streptomycete 16S and 23S rRNA-based target regions for oligonucleotide probes. Appl. Environ. Microbiol. 57, 1468-1477.

Stern, S., Powers, T., Changchien, L.M. and Noller, N.F. (1989). RNA-protein interactions in 30S ribosomal subunit: folding and function of 16S rRNA. Science 244, 783-790)

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

Suzuki, Y., Ono, Y., Nagata, A. and Yamada, T. (1988a). Molecular cloning and characterization of an rRNA operon in *Streptomyces lividans* TK21. J. Bacteriol. 170, 1631-1636.

Suzuki, Y., Nagata, A., Ono, Y. and Yamada, T. (1988b). Complete nucleotide sequence of the 16S rRNA of *Mycobacterium bovis* BCG. J. Bacteriol. 170, 2886-2889.

Suzuki, Y., Yoshinaga, K., Ono, Y., Nagata, A., and Yamada, T. (1987). Organization of rRNA genes in *Mycobacterium bovis* BCG. J. Bacteriol. 169, 839-843.

Taschke, C., Klinkert, M.Q., Walters, J., and Herrmann, R. (1986). Organization of the ribosomal RNA genes in *Mycoplasma hyopneumoniae* : the 5S rRNA gene is separated from the 16S and 23S rRNA genes. Mol. Gen. Genet. 205, 428-433.

Thomson-Carter, F.M., Carter, P.E. and Pennington, T.H. (1989). Differentiation of Staphilococcal species and strains by ribosomal RNA gene restriction patterns. J. Gen. Microbiol. 135: 2093-2097.

Tjepkema, J.D., Schwintzer, C.R. and Benson, D.R. (1986). Physiology of actinorhizal nodules. Ann. Rev. Plant Physiol. 37: 209-232.

Triman, K., Becker, E., Dammel, C., Katz, J., Mori, H.,Douthwaite, S., Yapijakis, C., Yoast, S. and Noller, H.F. (1989). Isolation of temperature-sensitive mutants of 16S rRNA in *E. coli*. J. Mol. Biol. 209, 643-653.

Ulbrich, N., Kamagai, I., and Erdmann, V.A. (1984). The number of ribosomal RNA genes in *Thermus thermophilus* HB8. Nucl. Acids Res.12, 2055-2060.

van Wezel, G.P., Vijgenboom, E. and Bosch, L. (1991). A comparative study of the ribosomal RNA operons of *Streptomyces coelicolor* A3(2) and sequence analysis of *rrn*A. Nucl. Acids Res. 19, 4399-4403.

Weber, A., Nurmiaho-Lassila, E-L. and Sundman, V. (1987). Features of the intrageneric *Alnus-Frankia* specificity. Physiol. Plantarum 70, 289-296.

Woese, C.R. (1987). Bacterial evolution. Microbiological Rev. 51, 221-271.

Woese, C.R. and Fox, G.E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA 74: 5088-5090.

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

Wolters, J. and Erdmann, V.A. (1988). Compilation of 5S rRNA and 5S rDNA gene sequences. Nucl. Acids Res. 16, r1-r70.

Yamada, T., Suzuki, Y., Nagata, A., Ono, Y. and Yamanouch, T. (1986). Ribosomes and drug resistance in *Mycobacterium*, pp.335-340, *In* Szabo, G., Biro, S. and Goodfellow, M.(eds), Biological, biochemical and biomedical aspects of actinomycetes. Akademiai Kiado, Budapest.

Yannish-Perron, C., Vieira, J., Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene 33, 103-119.

Young, R.A. and Steitz, J.A. (1978). Complementarity sequences 1700 nucleotides apart from a ribonuclease III cleavage site in *Escherichia coli* ribosomal precursor RNA. Proc. Natl. Acad. Sci. USA 75, 3593-3597.

Zablen, L.B., Kissil, M.S., Woese, C.R. and Buetow, D.L. (1975). Phylogenetic origin of the chloroplast and prokaryotic nature of its ribosomal RNA. Proc. Natl. Acad. Sci. 72, 2418-2422.

**APPENDIX 1**

**SIZE OF *FRANKIA* ACN14A GENOMIC DNA FRAGMENTS DETECTED BY SOUTHERN HYBRIDIZATION**

| Restriction endonuclease | Source of probes used in experiments | | | | |
|---|---|---|---|---|---|
| | λ FR-8 | | | | pnod4 |
| | 8.0B | 2.1 E/H | 0.55 H/B | 1.9 B | 1.5 B/E |
| Acc1 | 5.0, 5.2 | 5.0, 5.2 | 5.0, 5.2 | 1.8, 4.0, 5.0, 5,2 | 1.8, 4.0, 5.0, 5,2 |
| BamH1 | 4.8, 8.0 | 4.8, 8.0 | 4.8, 8.0 | 2.0, 2.8 | 2.0, 2.8 |
| BglII | 15.0, 21.0 | 15.0, 21.0 | 15.0, 21.0 | 15.0, 21.0 | 15.0, 21.0 |
| BstEII | 0.6, 2.5*, 2.7*, 3.7, 4.9 | 0.6, 3.7, 4.9 | 0.6*, 2.5*, 2.7 | 2.5, 2.7 | 2.5, 2.7 |
| EcoR1 | 6.6 | - | - | - | - |
| HindIII | 2.4 | - | 2.4 | 2.4 | - |
| Pst1 | 1.8, 5.1 | 5.1 | 5.1 | 5.1 | 5.1 |
| Sal1 | 5.0, 5.2, 6.4, 6.9 | 5.0, 5.2, 6.4, 6.9 | 5.0, 5.2,6.4, 6.9 | 5.0, 5.2, 6.4, 6.9 | 5.0, 5.2, 6.4, 6.9 |
| Sma1 | 0.7*, 1.1, 1.3*, 1.6, 1.9* | 0.7, 1.1, 1.6 | 1.6, 1.9*, 2.8* | 0.4, 0.9, 1.6 | 0.4, 0.9, 1.6 |
| SstI | 0.4, 0.8*, 1.0*, 8.0, 13.0 | 8.0, 13.0 | 8.0, 13.0 | 8.0, 13.0 | 8.0, 13.0 |
| SstII | 0.7, 5.1, 5.4 | 5.1, 5.4 | 5.1, 5.2 | 5.1, 5.2 | 5.1, 5.2 |

* indicates that the hybridization signal for these fragments was less intense than for other fragments generated by the same restriction endonuclease.

## APPENDIX 2

### 16S rRNA SEQUENCES USED FOR THE CONSTRUCTION OF THE TREES

1. *Frankia species* strain AcN14a
2. *Aeromicrobium erythreus*
3. *Arthrobacter globiformis*
4. *Corynebacterium variabilis*
5. *Frankia species* strain ORS20606
6. *Streptomyces ambofaciens*
7. *Micrococcus luteus*
8. *Mycobacterium kansasir*
9. *Mycobacterium bovis*
10. *Nocardioides luteus*
11. *Nocardia asteroides*
12. *Propionobacterium freudenreichis*
13. *Streptomyces griseus*
14. *Streptomyces lividans*

### A1. DISTANCES CALCULATED FOR THE CONSTRUCTION OF AN UNROOTED TREE BASED UPON COMPLETE 16S rRNA SEQUENCES (Numbers in parentheses are branch lengths)

Cycle 1: SEQ 8 (0.00975) joins SEQ 9 (0.00208)

Cycle 2: SEQ 1 (0.00356) joins SEQ 5 (0.00504)

Cycle 3: SEQ 6 (0.00584) joins SEQ 14 (0.00462)

Cycle 4: NODE 6 (0.01133) joins SEQ 13 (0.02012)

Cycle 5: SEQ 3 (0.01930) joins SEQ 7 (0.002161)

Cycle 6: SEQ 2 (0.03207) joins SEQ 10 (0.04585)

Cycle 7: NODE 8 (0.03840) joins SEQ 11 (0.03128)

Cycle 8: SEQ 4 (0.04734) joins NODE 8 (0.00515)

Cycle 9: NODE 2 (0.00729) joins SEQ 12 (0.06333)

Cycle 10: NODE 2 (0.00435).joins NODE 4 (0.00886)

Cycle 11: NODE 2 (0.00737) joins NODE 3 (0.02450)

Cycle 12 (last cycle, trichotomy): NODE 1 (0.03978) joins NODE 2 (0.00298) joins NODE 6 (0.02527)

## A2. BOOTSTRAP CONFIDENCE LIMITS

DIAGRAMMATIC REPRESENTATION OF THE ABOVE TREE

SCORE(PER1000 TRIALS)

```
 1  2  3  4  5  6  7  8  9 10 11 12 13 14
 *  *  *  *  *  *  *  .  .  *  *  *  *  *     1000
 .  *  *  *  .  *  *  *  *  *  *  *  *  *     1000
 *  *  *  *  *  .  *  *  *  *  *  *  *  .     1000
 *  *  *  *  *  .  *  *  *  *  *  *  .  .     1000
 *  *  .  *  *  *  .  *  *  *  *  *  *  *     1000
 *  .  *  *  *  *  *  *  *  .  *  *  *  *      963
 *  *  *  *  *  *  *  .  .  *  .  *  *  *      701
 *  *  *  .  *  *  *  .  .  *  .  *  *  *      880
 *  .  *  *  *  *  *  *  *  .  *  .  *  *      573
 *  .  *  .  *  *  *  .  .  .  .  .  *  *      885
 *  .  .  .  *  *  .  .  .  .  .  .  *  *      469
```

## B1. DISTANCES CALCULATED FOR THE CONSTRUCTION OF AN UNROOTED TREE BASED UPON SELECTED VARIABLE REGIONS OF 16S rRNA SEQUENCES

(Numbers in parentheses are branch lengths)

Cycle 1: SEQ 1 (0.02804) joins SEQ 5 (0.01741)

Cycle 2: SEQ 8 (0.03624) joins SEQ 9 (0.03042)

Cycle 3: SEQ 6 (0.03945) joins SEQ 14 (0.02488)

Cycle 4: NODE 6 (0.04660) joins SEQ 13 (0.09875)

Cycle 5: SEQ 3 (0.10594) joins SEQ 7 (0.11483)

Cycle 6: NODE 8 (0.10422) joins SEQ 12 (0.12531)

Cycle 7: SEQ 2 (0.07551) joins SEQ 11 (0.08801)

Cycle 8: NODE 2 (0.01232) joins SEQ 10 (0.08063)

Cycle 9: NODE 2 (0.00729) joins SEQ 12 (0.06333)

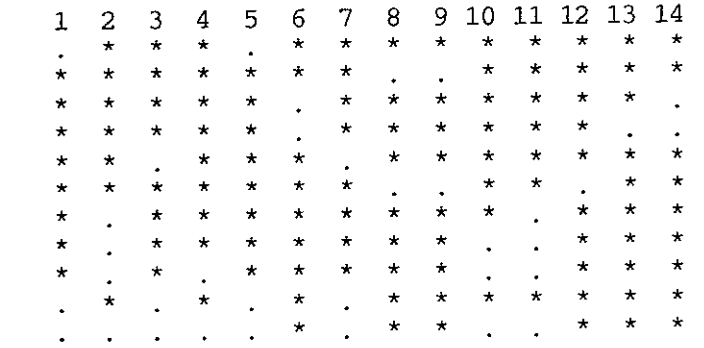
Cycle 10: NODE 2 (0.00996).joins SEQ 4 (0.10576)

Cycle 11: NODE 1 (0.01114) joins NODE 2 (0.02410)

Cycle 12 (last cycle, trichotomy):NODE 1 (0.00739) joins
NODE 6 (0.09597) joins
NODE 8 (0.01944)

## B2. BOOTSTRAP CONFIDENCE LIMITS

DIAGRAMMATIC REPRESENTATION OF THE ABOVE TREE




SCORE(PER1000 TRIALS)

1000
1000
999
998
884
516
548
482
462
360
257

**APPENDIX 3**

## ALIGNMENT OF THE VARIABLE REGIONS OF THE 16S RIBOSOMAL RNA USED FOR THE CONSTRUCTION OF THE PHYLOGENETIC TREE.

The sequence corresponds to nucleotides (59-89) + (164-205) + (961-1009) +(1079-1111) of *Frankia* AcN14a 16S rDNA. A list of the scientific names is given in Appendix 2. Ambiguous positions are denoted by X.

```
 1.ACN16SRD A-GTCGAGCGAGGGGCTTC---------GGCTCT--AGCGGCGA-TGACGCTACCGGGCA
 2.AERRRAA  AAGTCGAGCGGTAAGGCCC--TTTCGGGGGAACACGAGCGGCGA-TGACCCCCTGTCGCA
 3.ARGSSRNA AAGTCGAACGATGATCCGGTGCTT-------------GCACCGA-TGACTCCTCATCGCA
 4.COR16S   AAGTCGAACGGAAAGGCCCTGCTT-------------GCAGGGA-GGACCAT-CGTTTAA
 5 FRARRNA  AAGTCGAGCGGGGAGCTTC---------GGCTCTC-AGCGGCGA-TGACATTGCCGGGCA
 6 M27245   AAGTCGAACGATGAACCACTTCGGTGGGGATT----AGTGGCGACTGATCCGCTTGGGCA
 7.MLURRSS  AAGTCGAACGATGAAGCCCAGCTT-------------GC-TGGA-GGAGCGTCCACCGCA
 8.MSGRR16  AAGTCGAACGGAAAGGTCT--CTTCGGAGACACTCGAGTGGCGA-GGACCACTTGGCGCA
 9.MSGTGDA  AAGTCGAACGGAAAGGTCT--CTTCGGAGACACTCGAGTGGCGA-GGACCACGGGATGCA
10.NOA16U   AAGTCGAGCGGAAAGGXXC--------GGGGTCC------GGCGAC-GACAACCGATTGCA
11.NOC16SRN AAGTCGAGCGGTAAGGCCC--TTC---GGGTACACGAGCGGCGA-TGACCTTACATCGCA
12.PRS16S  AAGTCGGACGGTXAGGCCTTTXXXXGGGGGTXCTCGAGTGGCGA-TGAGCCTXGCCTGCA
13.STMDRNA AAGTCGAACGATGAAGCCTTTCGGGGTGGATT----AGTGGCGAACACTCTGTCCCG-CA
14.STMRRNB AAGTCGAACGATGAACCACTTCGGTGGGGATT----AGTGGCGACTGACCCTCGCAGGCA
           * ****  **                                **              *
```

```
 1.ACN16SRD TCTGGTGGTG-TGGAAAGATTTAT--CGGATGCAGGGAAATCTCGTAGAGATACGGGGTC
 2.AERRRAA  TGGTG-GGGGGTGGAAAGCTCTGGCG-GAATATGCCGGAAAGCTGCAGAGATGTGGCCCX
 3.ARGSSRNA TGGTG-GGGGGTGGAAAGCTTT-TTGTGGATGGACCGGACCGCCGCAGAAATGTGGTTTC
.4.COR16S   TG-TC-GGTGGTGGAAAGTTTT-T--CGGATATGCCGGATCGXCGCAGAGATGCGTXTTC
 5.FRARRNA TCTGGTGGTG-TGGAAAGATTTAT--CGGATGCAGGGAAATCTCGTAGAGATACGGGGTC
 6.M2745   TCCAG-GCGGTTCGAAAGCTCCG--GCGGATACACCGGAAAGCATTAGAGATAGTGCCCC
 7.MLURRSS TGGTG-GGTGTTGGAAAGATTT-AT-CGGATGTTCTCGATCGCCGTAGAGATACGGTTTC
 8.SGRR16  TGCCT-TGTGGTGGAAAGC--TTTTGCGGATGCACAGGACGCGTCTAGAGATAGGCGTTC
 9. MSGTGDA TGTCT-TGTGGTGGAAAGCGCTTTAGCGGATGCACAGGACGCGTCTAGAGATAGGCGTTC
10.NOA16U  TGATCTGGTTGTGGAAAGTTTTT---CGGGTACACCGGAAAGCTGCAGAGATGTAG---C
11.NOC16SRN TGGTG-TTTGGTGGAAAGATTTATCG-GA-TACACCGGAAACCTGCAGAGATGTAGGCCC
12.PRS16S  TXGGT-CGGGTTGGAAAGCTTTAT-GCGGATGTACTGGAAGCGTTCAGAGATGGGCGTGC
13.TMDRNA  TGGGA-CGGGGTTAAAAGCTCCG--GCGGATATACCGGAAAGCATCAGAGATGGTGCCCC
14. STMRRNB TCTGC-GAGGTTCGAAAGCTCCG--GCGGATACACCGGAAAGCATCAGAGATGGTGCCCC
            *          *  ****         * *       *      *** **
```

```
 1.ACN16SRD CG---TAAGGGTCC-TGCACACGTCCTATGTTGCCAGCGAGTT-------ATGTCGGGGA
 2.AERRRAA  XC---TTGTGGTCGGTATACACGTCCTATGTTGCCAGCACGT-------GATGGTGGGXA
 3.ARGSSRNA TCCTTTTGGGGCCGGTTCACACGTTCCATGTTGCCAGCGCGTA-------ATGGCGGGGA
 4.COR16S   CC---TXGTXGTCGGTATACAXGTCTTGTGTTGCCAGCACGTT-------ATGGTGGGGA
 5.FRARRNA  CG---TAAGGGTCC-TGCACACGTCCTATGTTGCCAGCGAGTC-------GTGTCGGGGA
 6.M2745    CC---TTGTGGTCGGTGTACATGTCCCGTGTTGCCAGCAAGCCCTTCGGGGTGTTGGGGA
 7.MLURRSS  CCCTTT-GGGXCGGGTTCACACGTTCCATGTTGCCAGCACGTC-------GTGGTGGGGA
 8.MSGRR16  CC---TTGTGGCCTGTGTGCATGTCTCATGTTGCCAGCGGGT-------AATGCCGGGGA
 9.MSGTGDA  CC---TTGTGGCCTGTGTGCATGTCTCATGTTGCCAGCACGT-------AATGGTGGGGA
10.NOA16U   CCCTXTTX--GTCGGTGTACACGTCCTATGTTGCCAGCAA----TTCGG----TTGGGGA
11.NOC16SRN CC---TTGTGGTCGGTGTACATATCTTATGTTGCCAGCGCGT-------AATGGCGGGGA
12.PRS16S  CT---TTTTGGCTGGTACACACGTCCAATGTTGCCAGCA-GT-------TCGGCTGGGGA
13.STMDRNA CC---TTGTGGTCGGTATACATGTTCTGTGTTGCCAGCATGCC-TTCGGGGTGATGGGGA
14.STMRRNB CC---TTGTGGTCGGTGTACATGTCCCGTGTTGCCAGCAAGCCCTTCGGGGTGTTGGGGA
                 *         *   **  *    **********                    ***

 1.ACN16SRD C
 2.AERRRAA  C
 3.ARGSSRNA C
 4.COR16S   C
 5.FRARRNA  C
 6.M2745    C
 7.MLURRSS  C
 8.MSGRR16  C
 9.MSGTGDA  C
10.NOA16U   C
11.NOC16SRN C
12.PRS16S   C
13.STMDRNA  C
14.STMRRNB  C
```