R. F. A. (1987) *Biochim. Biophys. Acta 903*, 206-217.
Verhallen, P. F. J., Bevers, E. M., Comfurius, P., & Zwaal, R. F. A. (1988) *Biochim. Biophys. Acta 942*, 150-158.
Weiss, H. J., Vicic, W. J., Lages, B. A., & Rogers, J. (1979) *Am. J. Med. 67*, 206-213.
White, G. C., II (1980) *Biochim. Biophys. Acta 631*, 130-138.
Wiedmer, T., & Sims, P. J. (1985) *J. Biol. Chem. 260*, 8014-8019.
Wiedmer, T., Esmon, C. T., & Sims, P. J. (1986a) *Blood 68*, 875-880.
Wiedmer, T., Esmon, C. T., & Sims, P. J. (1986b) *J. Biol. Chem. 261*, 14587-14592.
Wiedmer, T., Ando, B., & Sims, P. J. (1987) *J. Biol. Chem. 262*, 13674-13681.
Zhuang, Q. Q., Rosenberg, S., Lawrence, J., & Stracher, A. (1984) *Biochem. Biophys. Res. Commun. 118*, 508-513.

Identification and Assignment of Base Pairs in Four Helical Segments of *Bacillus megaterium* Ribosomal 5S RNA and Its Ribonuclease T1 Cleavage Fragments by Means of 500-MHz Proton Homonuclear Overhauser Enhancements†

Jong Hwa Kim and Alan G. Marshall*,‡
Department of Chemistry, The Ohio State University, Columbus, Ohio 43210
Received June 6, 1989; Revised Manuscript Received September 19, 1989

ABSTRACT: Three different fragments of *Bacillus megaterium* ribosomal 5S RNA have been produced by enzymatic cleavage with ribonuclease T1. Fragment A consists of helices II and III, fragment B contains helix IV, and fragment C contains helix I of the universal 5S rRNA secondary structure. All (eight) imino proton resonances in the downfield region (9–15 ppm) of the 500-MHz proton FT NMR spectrum of fragment B have been identified and assigned as G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88} and three unpaired U's (U_{85}, U_{86}, and U_{87}) in helix IV by proton homonuclear Overhauser enhancement connectivities. The secondary structure in helix IV of the prokaryotic loop is completely demonstrated spectroscopically for the first time in any native or enzyme-cleaved 5S rRNA. In addition, G_{21}·C_{58}-A_{20}·U_{59}-G_{19}·C_{60}-A_{18}·U_{61} in helix II, U_{32}·A_{46}-G_{31}·C_{47}-C_{30}·G_{48}-C_{29}·G_{49} in helix III, and G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in the terminal stem (helix I) have been assigned by means of NOE experiments on intact 5S rRNA and its fragments A and C. Base pairs in helices I–IV of the universal secondary structure of *B. megaterium* 5S RNA are described.

Ribosomal 5S RNA is known to be an important structural and functional component of the large subunit of all ribosomes. Because of its essential role in protein synthesis in the ribosome (Pieler et al., 1984) and its small size among ribosomal RNA molecules, 5S rRNA has been studied extensively for 2 decades. Since the late 1960s, a large number of 5S rRNAs from many different prokaryotic and eukaryotic ribosomes have been sequenced (Erdmann & Wolters, 1986). Following the general belief that there should be a universal secondary structure for 5S rRNAs, many universal secondary structure models have been proposed on the basis of comparative sequence analysis (Fox & Woese, 1975; Leuhrsen & Fox, 1981; Studnicka et al., 1981; De Wachter et al., 1982; Delihas & Andersen, 1982), enzymatic cleavage and chemical modification (Nishikawa & Takemura, 1974; Pieler & Erdmann, 1982), and physical measurements (Kearns & Wong, 1974; Osterberg et al., 1976; Luoma & Marshall, 1978a,b; Chang et al., 1984).

Of the techniques used for the structural studies of RNA, proton homonuclear Overhauser enhancement (NOE) experiments are the most powerful for the determination of base-paired sequences in nucleic acids (Johnston & Redfield, 1978). Supplemented by the complete structural determination of tRNA by X-ray crystallography (Kim, 1976; Rich, 1977), NOE experiments have successfully identified (as A·U, G·C, or G·U) and assigned (to specific primary sequence bases) virtually all of the secondary and tertiary base pairs in several tRNAs in aqueous solution (Schimmel & Redfield, 1980; Reid, 1981; Johnston & Redfield, 1981; Heerschap et al., 1983a,b; Roy & Redfield, 1983).

Application of NOE techniques to the larger 5S rRNA molecule presents several problems, as noted in a recent review of ^1H, ^{13}C, ^{15}N, ^{19}F, and ^{31}P NMR of 5S rRNAs (Marshall & Wu, 1989). The main obstacles are broader peak widths and more overlap of the proton resonances in the downfield region resulting from more base-paired imino protons than for tRNA. Moreover, the secondary structure and the tertiary folding patterns of 5S rRNA have not yet been established from X-ray diffraction. Although some of the imino proton resonances in the downfield proton NMR spectrum of 5S rRNA have been assigned to specific base pairs in the universal secondary structure models [e.g., Chang and Marshall (1986a)], NOE experiments must be supported by other experiments that shift or simplify the proton NMR spectrum. Among such methods are site-specific spin labeling (Lee & Marshall, 1987) and the use of enzymatic cleavage fragments (Kime & Moore, 1983a,b; Kime et al., 1984; Li & Marshall, 1986; Chen & Marshall, 1986; Li et al., 1987). Proton NMR studies of enzyme-cleaved fragments of 5S rRNA are especially attractive because of greatly reduced peak overlap resulting from fewer, narrower proton resonances.

The secondary structure of prokaryotic 5S rRNA has been studied in *Escherichia coli* (Gram negative) and *Bacillus*

† This work was supported by grants (to A.G.M.) from the U.S. Public Health Service (NIH GM-29274 and NIH RR-01458) and The Ohio State University.
‡ Also a member of the Department of Biochemistry.

0006-2960/90/0429-632$02.50/0

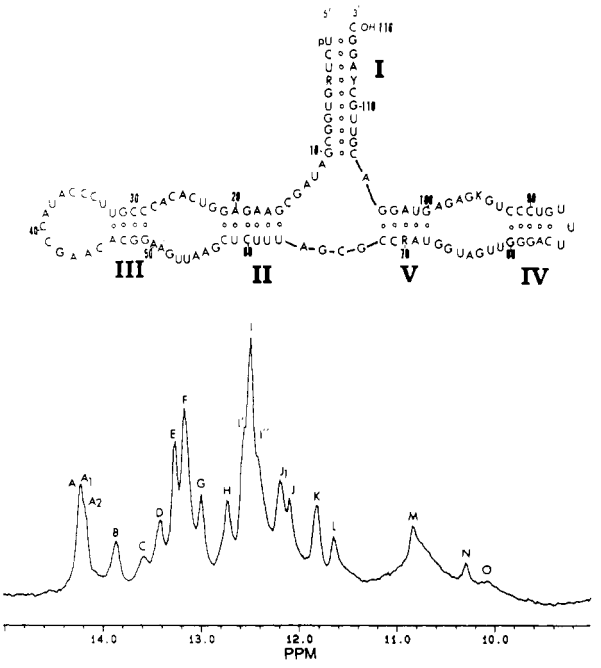

FIGURE 1: (Top) Proposed secondary structure (Fox and Woese model with an additional helix V) adapted to the primary base sequence of *B. megaterium* 5S rRNA. Heterogeneity of bases in the sequence is denoted as follows: R is G or A, K is G or U, and Y is C or U. (Bottom) 500-MHz downfield proton NMR spectrum of *B. megaterium* 5S rRNA at 23 °C in 10 mM cacodylic acid, 0.1 M NaCl, 1 mM EDTA, pH 7.0, and 95%:5% H_2O/D_2O. The peaks are labeled from A to O for convenience.

subtilis (Gram positive) bacteria by NMR techniques. Among a number of proposed models for the secondary structure of 5S rRNAs, the Fox and Woese (1975) model, with an additional helix V, is believed to best describe prokaryotic 5S rRNAs. Comparison of the imino proton NMR spectra of two or more 5S rRNAs that have similar base sequences can facilitate interpretation of the spectra and thus help to establish the secondary structure (Chen & Marshall, 1986). NMR analysis of *Bacillus megaterium* 5S rRNA is particularly useful, because its primary base sequence (and probably its secondary structure) is very similar to that of previously studied *B. subtilis*. Moreover, enzyme-digested fragments of *B. megaterium* which retain corresponding segments of the original 5S rRNA structure are especially amenable to NMR analysis, for the reasons noted above.

In this paper, several base-paired segments of *B. megaterium* 5S rRNA are identified from proton NMR experiments by correspondence to the universal base-pairing model adapted to the primary nucleotide sequence shown in Figure 1. The 5S rRNA was extracted and purified from (Gram positive) *B. megaterium* cells. Three different fragments (denoted as A, B, and C), which contain helices II and III, V, and I (see Figure 1), were produced by enzymatic cleavage with RNase T1. The fragments appear to retain the original structure of the intact 5S rRNA molecule. Eight imino proton resonances in the downfield proton NMR spectrum of the second fragment (fragment B) can be assigned as $(G{\cdot}C)_3(A{\cdot}U)(G{\cdot}C)$ and three unpaired U's in the "prokaryotic loop" (helix IV in Figure 1) by observed NOE connectivities and ring current calculations. The prokaryotic loop is reportedly involved in the binding of 5S rRNA to ribosomal protein L25 (Kime & Moore, 1983b). This paper presents the first complete NMR-based secondary structure of the prokaryotic loop in any 5S rRNA and its fragments. (Although base pairs from this helix have previously been identified by NOE connectivities in *E. coli* 5S rRNA, those assignments were not complete.) Base-pair sequences in helices I–III can then be assigned in the other two fragments and in intact 5S rRNA by NOE connectivities and variable-temperature NMR.

Materials and Methods

Isolation and Purification of B. megaterium 5S rRNA. B. megaterium strain KM (ATCC 13632) cells were provided and grown by the Fermentation Laboratory of The Ohio State University. The procedure for isolation and purification of 5S rRNA was nearly the same as that for *B. subtilis* 5S rRNA described previously by Li et al. (1984). The cells were suspended in 0.01 M sodium acetate, 0.1 M NaCl, and 0.25% sodium dodecyl sulfate (SDS), pH 5.0. The same volume of phenol was added and the mixture stirred for 1.5 h at room temperature. The aqueous layer was separated from the phenol layer by centrifugation at 11000*g* for 15 min, and RNA was allowed to precipitate overnight at −20 °C after the addition of 2.5 volumes of cold 95% ethanol. Purification of 5S rRNA was done by means of ion-exchange (DEAE-32) and gel filtration (Sephadex G-75) column chromatography.

Production, Isolation, and Purification of RNase T1 Cleaved Fragments of B. megaterium 5S rRNA. B. megaterium 5S rRNA was dissolved in a solution containing 0.01 M tris(hydroxymethyl)aminomethane (Tris) base and 0.3 M NaCl, pH 7.5, at a concentration of A_{260} = 20/mL. RNase T1 (Boehringer Mannheim) was added at 2.5 μL (=250 units) per milliliter of 5S rRNA, and the reaction was allowed to proceed for 25 min at room temperature (~24 °C). The reaction was stopped by the addition of an equal volume of 5% SDS. Phenol extraction was done twice to remove RNase T1, and the fragments of the 5S rRNA were recovered from the aqueous layer by precipitation with 2.5 volumes of cold (−20 °C) ethanol. The purification of RNase T1 cleaved fragments was achieved by nondenaturing Sephadex G-50 gel filtration chromatography (150 × 2.5 cm column) with aqueous elution buffer containing 0.1 M NaCl and 10 mM Tris base, pH 7.0.

Gel Electrophoresis. The characterization and purity test of the RNase T1 digestion products was done by polyacrylamide gel electrophoresis under nondenaturing conditions. The gel consisted of 10% acrylamide, 0.5% bis(acrylamide), 0.08% TEMED, and 0.5% ammonium sulfate. The gels were run at 3 V/cm for 18 h at room temperature with 0.1 M KCl, 5 mM $MgCl_2$, and 50 mM Tris–borate, pH 7.8, as the electrophoresis buffer; RNA was visualized with methylene blue.

The primary nucleotide sequences of RNase T1 digested fragments A–C of *B. megaterium* 5S rRNA were confirmed by an RNA sequencing procedure (BRL RNA sequencing system, Bethesda Research Laboratories). This procedure consisted of initial 3′-dephosphorylation with calf intestine alkaline phosphatase followed by 3′-end labeling with [5′-^{32}P]pCp with T4 RNA ligase, purification of the end-labeled fragments, RNA T1 partial digestion, and autoradiography of a subsequent sequencing electrophoresis gel.

NMR Samples. B. megaterium 5S rRNA or RNase T1 cleaved fragments of *B. megaterium* 5S rRNA were dissolved in 10 mM EDTA, 100 mM NaCl, and 10 mM cacodylic acid, pH 7.0, and then dialyzed against 500 volumes of 1 mM EDTA, 100 mM NaCl, and 10 mM cacodylic acid, pH 7.0, at 5 °C for 3 h. The samples were then dialyzed against fresh buffer overnight. The samples were concentrated by ultrafiltration by use of an Amicon Centricon apparatus. D_2O was added to a final 5% (v/v) concentration to provide a ^{2}H field-frequency lock signal. The final 5S rRNA concentration of the NMR sample was 28.7 mg/mL. The concentrations of fragments A–C in NMR samples were 225, 144, and 140

A_{260} units in 0.5 mL, respectively. All NMR samples exhibited at least 95% purity as estimated from electrophoresis gel scans obtained under nondenaturing conditions.

NMR Spectroscopy. All ^1H NMR spectra were obtained with a Bruker AM-500 FT NMR spectrometer, with phase-cycled quadrature detection, without sample spinning. Water suppression in NMR spectra was achieved by use of a 1331 hard-pulse sequence (Hore, 1983) with the carrier frequency centered at the H_2O resonance. Chemical shifts were measured relative to the H_2O resonance and referenced to DSS [3-(trimethylsilyl)-1-propanesulfonic acid] from an independent calibration. Downfield shifts are defined as positive.

The nuclear Overhauser enhancement (NOE) difference spectra of 5S rRNA were obtained by use of a modified Redfield 214 pulse sequence (Redfield et al., 1975; Chang & Marshall, 1986) with the radio-frequency carrier centered at 15 ppm. NOE spectra for RNA fragments were produced with a 1331 hard-pulse sequence, with an acquisition period of 0.67 s [16K time domain data, typically 16000 time domain transients (8000 on-resonance and 8000 off-resonance)] and 0.5-s preirradiation (0.2-mW decoupler power) of the resonance of interest, with off-resonance irradiation at 18 ppm. All NOE experiments for fragments were carried out at 23 °C.

Results and Discussion

Peak Identification in 5S rRNA. The downfield (9–15 ppm) 500-MHz proton NMR spectrum of *B. megaterium* 5S rRNA (Figure 1) exhibits ~20 resolved peaks corresponding to hydrogen-bonded ring imino proton resonances from Watson–Crick (A·U and G·C) base pairs, "wobble" G·U base pairs (Reid, 1981), and some unpaired U and G imino protons that are shielded from the solvent with a sufficiently slow exchange rate with H_2O (Hare & Reid, 1982). However, ~30 separate imino proton resonances can be identified in the downfield proton NMR spectrum of 5S rRNA from the higher resolution spectra obtained from the three RNA fragments discussed below.

As can be seen from Figure 1, the assignment of particular resonances to specific base-pair protons in helical RNA segments is an underdetermined problem. However, two G·U base pairs (peaks K_1/M and K_2/L) can immediately be identified from their characteristic NOE difference spectra. Starting from the G·U base pair, K_1/M, the short NOE-connected segment G·U-G·C-G·C (K_1/M, E, J) can be established (see Figure 2). From the universal secondary structure model (Figure 1), such a segment occurs only in the terminal stem (helix I) of *B. megaterium* 5S rRNA: either G_4·C_{112}-G_5·C_{111}-U_6·G_{110} or G_8·U_{108}-C_9·G_{107}-G_{10}·C_{106}. If the resonances K_1/M, E, and J are G_8·U_{18}-C_9·G_{107}-G_{10}·C_{106}, then resonance J would be expected to melt quickly at elevated temperature because base pair G_{10}·G_{106} is located at the end of the helix. For example, Gewirth et al. (1987) have reported that base pairs G_{10}·C_{110} and G_9·U_{111} in helix I of *E. coli* 5S rRNA are unstable both in the fragment and in native 5S rRNA. As we can see from the variable-temperature experiment shown in Figure 3, resonances J and E are in fact highly stable and do not melt even at 50 °C. Thus, we exclude the assignments of resonances J and E as G_{10}·C_{106}-C_9·G_{107}, and we assign resonances J, E, and K_1/M as G_4·C_{112}-G_5·C_{111}-U_6·G_{110}: with that assignment, peaks J (G_4·C_{112}) and E (G_5·C_{111}) are expected to be very stable at elevated temperature because they are located in the middle of the most stable helix (helix I).

It is interesting to compare the G_4·C_{112}-G_5·C_{111}-U_6·G_{110} segment (J, E, K_1/M) in *B. megaterium* with the assigned segment G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in *B. subtilis*, another Gram-positive bacterium, as previously reported by Chang and Marshall (1986a). The two bacterial 5S rRNAs exhibit high homology in base-pair sequence, so that the (G_4·C_{112}-G_5·C_{111}-U_6·G_{110}) segments in both 5S rRNAs are located at the same place in the secondary structure, as evidenced by the virtually identical chemical shifts for resonances J, E, and K_1/M (12.10, 13.30, and 11.84/10.84 ppm, respectively) in the spectrum of *B. megaterium* 5S rRNA and resonances O, G, and Q/T (12.12, 13.38, and 11.7/10.9 ppm, respectively) in *B. subtilis* 5S RNA (Chang & Marshall, 1986a). [Resonance K_1 in *B. megaterium* 5S rRNA is shifted slightly (0.14 ppm) from resonance Q in *B. subtilis* 5S rRNA, probably because G_7·U_{109} in *B. megaterium* 5S rRNA is replaced by G_7·C_{109} in *B. subtilis* 5S rRNA.] The virtually identical chemical environment for the base pairs G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in both bacteria offers additional evidence for a common secondary structure for 5S rRNAs from all prokaryotic ribosomes.

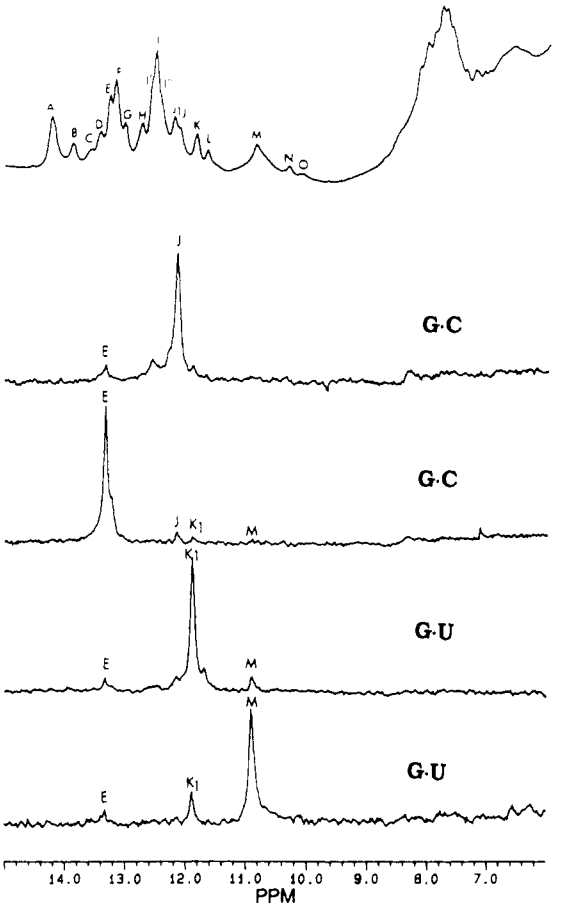

FIGURE 2: 500-MHz proton NOE difference spectra of resonances J, E, K_1, and M of *B. megaterium* 5S rRNA. The NOE difference spectra of resonances K_2/L are not shown. The G·U base pair (resonances K_1/M) is connected to a G·C base pair (E), which is connected to another G·C pair (J). This segment is assigned as G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in helix I (see text).

The assignment of segment G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in helix I was reconfirmed by NOE experiments with fragment C (see below). Although numerous NOE connectivities were observed between the various resonances of intact 5S rRNA, the extensive peak overlap precluded definitive assignment of base-paired segments, and it was necessary to analyze enzyme-cleaved fragments of the 5S rRNA.

RNase T1 Cleaved Fragments of B. megaterium 5S rRNA. Three different RNase T1 cleaved fragments of *B. megaterium* 5S rRNA were obtained by the method described under

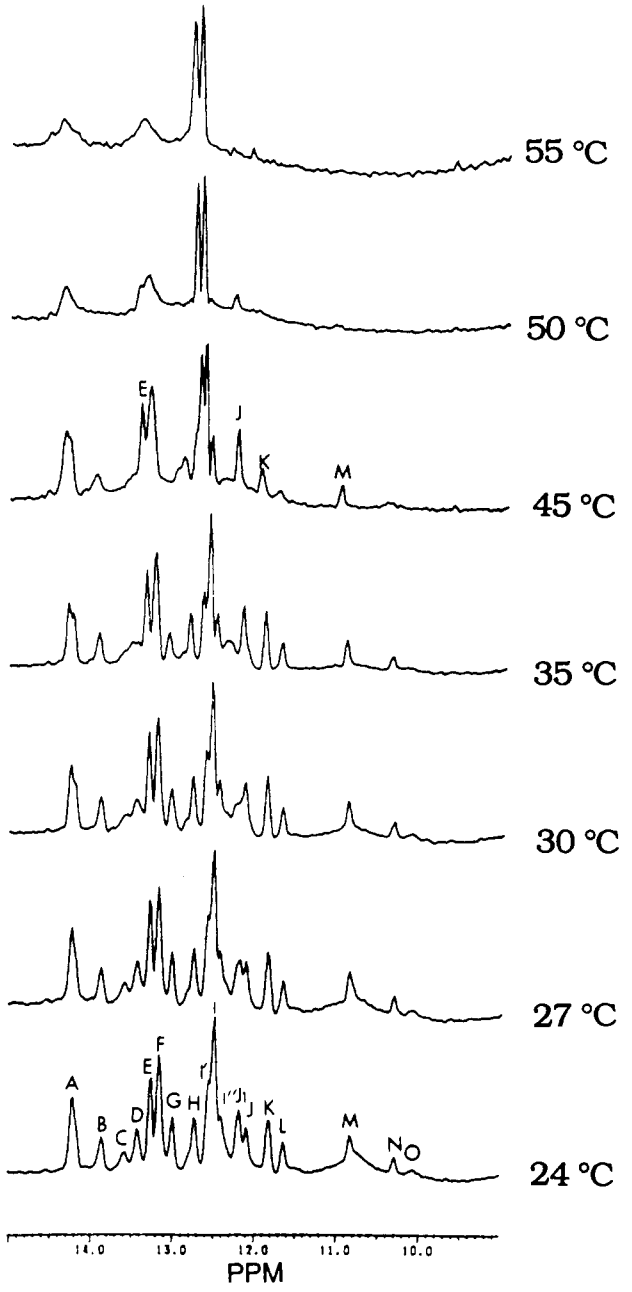

FIGURE 3: 500-MHz downfield proton NMR spectra of *B. megaterium* 5S rRNA in 10 mM cacodylic acid, 0.1 M NaCl, and 1 mM EDTA, pH 7.0, at various temperatures. Peaks J and E are stable to ~50 °C and have about the same melting onset temperature as peaks K_1 and M.

Materials and Methods. Fragments A–C could be separated by Sephadex G-50 gel filtration chromatography. By RNA sequence analysis, RNase T1 cleaved fragment A was confirmed to consist of residues 15–65 of 5S rRNA (Figure 4, top). Fragment B showed a rather complicated pattern. Under denaturing conditions, fragment B exhibited a few fragments of different sizes consisting of bases between residues 66 and 98 but consisted mainly of residues 78–98 (Figure 4, middle). Fragment C is believed to correspond to residues 1–14 and 99–116 (Figure 4, bottom).

Assignment of Base-Pair Sequence in Fragment B (Helix IV). RNase T1 cleaved fragment B contains helix IV (the prokaryotic loop) of *B. megaterium* 5S rRNA (see Figure 4). As can be seen in Figure 4, the imino proton resonances observed in this fragment are located in the most crowded region of the downfield spectrum of 5S rRNA. Hence, fragment B offers an optimal window for study of the secondary structure in helix IV. Helix IV has been reported to be the part of 5S rRNA, whose proton NMR spectrum is most strongly affected by binding of ribosomal protein L25 (Kime & Moore, 1983b). It is therefore believed that the binding site for ribosomal protein L25 on 5S rRNA includes helix IV.

Table I: Identification and Assignment of Base-Pair Imino Proton Resonances in *B. megaterium* 5S rRNA Fragment B by Means of NOE Connectivities

peak[a]	chemical shift (ppm)	base pair	NOE connectivity to
M_1	11.13	unpaired U_{85}	O
O	10.10	unpaired U_{86}	M_1, M_2
M_2	11.28	unpaired U_{87}	O, D_2
D_2	13.39	C_{84}·G_{88}	M_2, A_2
A_2	14.10	A_{83}·U_{89}	D_2, I″
I″	12.44	G_{82}·C_{90}	A_2, spillover
I	12.51	G_{81}·C_{91}	spillover
I'_2	12.60	G_{80}·C_{92}	spillover

[a] See Figure 5.

Eight imino proton resonances were observed in the downfield spectrum of fragment B. Five gave sharp peaks, and the other three upfield peaks were weaker. It appears likely that the five strong resonances (A_2, D_2, I'_2, I, and I″) are due to the base-pair sequence G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88}. The chemical shifts of the resonances can be estimated from ring currents on the basis of an 11-fold RNA A-helix (Arter & Schmidt, 1976), from assumed intrinsic A·U and G·C positions of 14.35 and 13.45 ppm, respectively (Reid et al., 1979). Although ring current estimates of chemical shifts are far from perfect, they were nevertheless quite helpful in establishing a self-consistent set of assignments of all of these base-pair imino proton resonances in $tRNA^{Phe}$ (Johnston & Redfield, 1981). The calculated chemical shift values for G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88} match reasonably well with the experimental values obtained from our proposed assignments of resonances I'_2, I, I″, A_2, and D_2. The other three upfield resonances could be due to the N3-H protons of three unpaired uridines in the loop (U_{85}, U_{86}, and U_{87}). It has been reported that the chemical shifts for N3-H of uridine and N1-H of guanosine dissolved separately in dry Me_2SO-d_6 occur at 11.4 and 10.7 ppm (Hurd & Reid, 1979a,b). Some of the N3-H U and N1-H G protons in RNA may be shielded from exchange with H_2O. Those unpaired imino protons can thus give observably narrow proton NMR signals between 9 and 11 ppm (Hare & Reid, 1982; Chen & Marshall, 1986).

The assignment of resonances I'_2, I, I″, A_2, D_2, M_1, O, and M_2 to G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88} and U_{85}, U_{86}, and U_{87} can be confirmed more convincingly from the NOE difference spectra of eight irradiated resonances (Figure 5). For example, resonance A_2 can easily be identified as an A·U base pair by its chemical shift (14.1 ppm) and the sharp NOE difference peak from adenine C2-H at ~7–8 ppm. As predicted above, A_2 (A_{83}·U_{89}) is NOE connected to D_2 (C_{84}·G_{88}) and I″ (G_{82}·C_{90}). D_2 is connected to A_2 and M_2 (U_{87}), M_2 to D_2 and O (U_{86}), O to M_1 (U_{85}) and M_2, and M_1 to O. Irradiation of peak I″ gives an NOE difference signal at A_2, but NOE of I (G_{82}·C_{90}) cannot be observed because of spillover. At this stage, the definitive NOE connectivities are sufficient to establish the base-pair sequences G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88} (resonances I'_2, I, I″, A_2, and D_2) and U_{85}, U_{86}, and U_{87} (resonances M_1, O, and M_2) (Table I). [The NOE difference signals corresponding to resonances I'_2 (G_{80}·C_{92}) and I (G_{81}·C_{91}) cannot be observed because any NOE effect is obscured by the pronounced spillover from irradiation.]

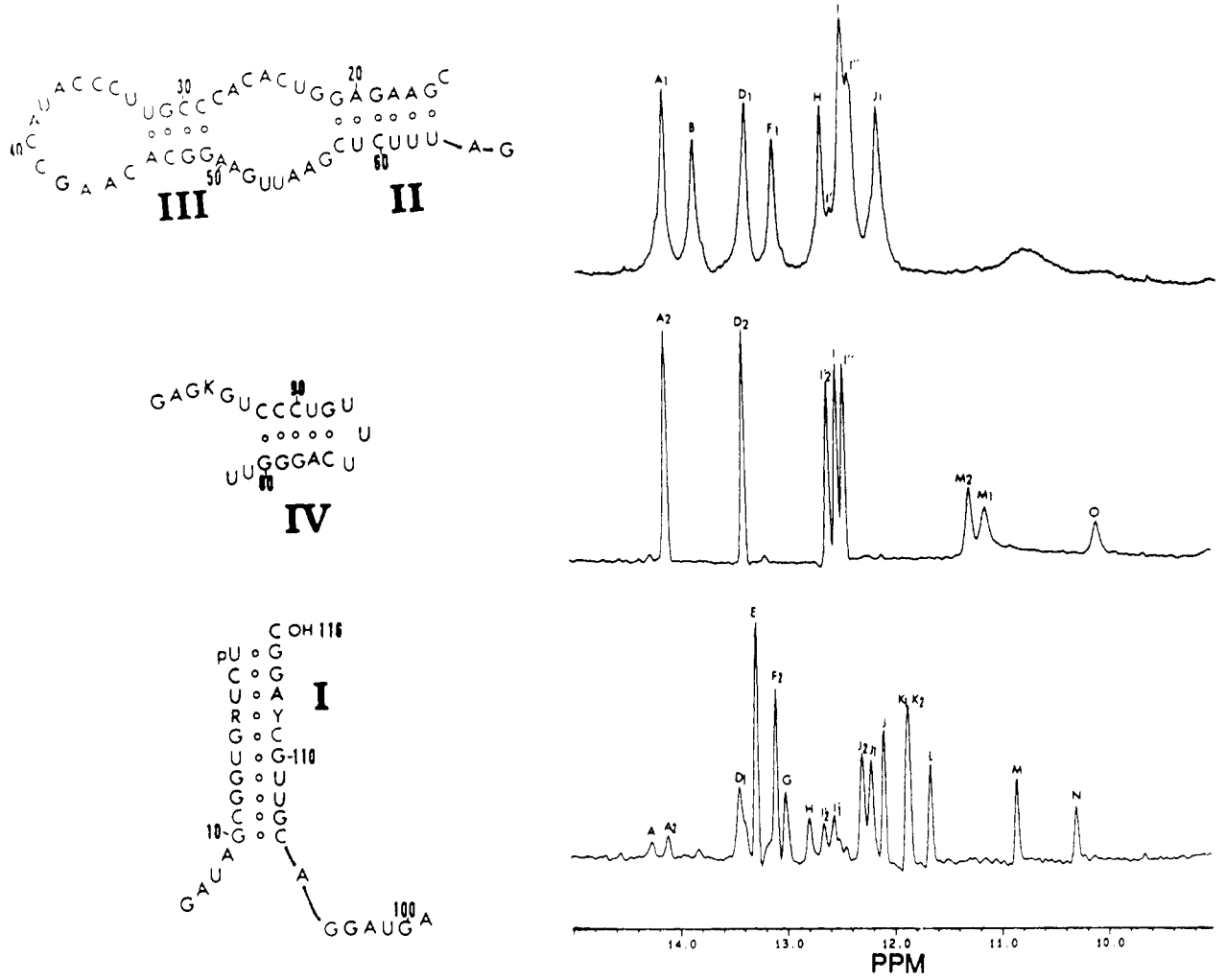

FIGURE 4: Proposed secondary structure segments (adapted from the Fox and Woese model) and the corresponding downfield proton NMR spectra at 23 °C of RNase T1 cleaved fragments A (top), B (middle), and C (bottom) of *B. megaterium* 5S rRNA. Fragment B consists of a few fragments of different sizes ranging between residues 66–98 but mostly residues 78–98. The fragments were dissolved in 10 mM cacodylic acid, 0.1 M NaCl, 1 mM EDTA, pH 7.0, and 95%:5% H_2O/D_2O.

Although prior studies of the base-pairing pattern in the prokaryotic loop of *E. coli* achieved much in the identification and assignment of the base pairs in that region (Kime & Moore, 1983a), ambiguity still remained, and three unpaired imino protons in the loop were never observed. The assignments of $(G{\cdot}C)_3(A{\cdot}U)(G{\cdot}C)$ and three U's presented here provide the first complete proton NMR description of the secondary structure in this helical region of any 5S rRNA and its fragments. In particular, the G·C base pair adjacent to the three unpaired pyrimidines ($G_{86}{\cdot}C_{90}$ in *E. coli* 5S rRNA) which could not be identified in *E. coli* 5S rRNA is observed in *B. megaterium* 5S rRNA ($C_{84}{\cdot}G_{88}$). The putative G·U pair ($U_{78}{\cdot}G_{94}$) is not observed in this fragment, probably because of too fast exchange of N1-H of G_{94} with water. N3-H of U_{85} seems to be spatially farther than N3-H of U_{87} from N1-H of G_{88} in the loop, because NOE connectivities are observed between resonances D_2 (N1-H of G_{88}) and M_2 (N3-H of U_{87}) but not between D_2 and M_1 (N3-H of U_{85}) (Figure 5). The $G_{80}{\cdot}C_{92}$-$G_{81}{\cdot}C_{91}$-$G_{82}{\cdot}C_{90}$-$A_{83}{\cdot}U_{89}$-$C_{84}{\cdot}G_{88}$ segment in helix IV is so stable that even at 40 °C resonances I'_2, I, I″, A_2, and D_2 show strong signals in the proton NMR spectrum (not shown). The resonances melt, however, in the order I'_2 ($G_{80}{\cdot}C_{92}$), I ($G_{81}{\cdot}C_{91}$), I″ ($G_{82}{\cdot}C_{90}$), and D_2 ($C_{84}{\cdot}G_{88}$). This melting pattern is a good indication that the unwinding of helix IV begins with the breakage of base pair $G_{80}{\cdot}G_{92}$ at slightly elevated temperature. As the temperature increases, base pairs melt in succession, ending with the base pair adjacent to the loop. The imino protons of U_{85}, U_{86}, and U_{87} in the loop are also well protected from rapid exchange with water. Overall, the prokaryotic loop is very stable.

Table II: Identification and Assignment of Base-Pair Imino Proton Resonances in *B. megaterium* 5S rRNA Fragment A by Means of NOE Connectivities

peak	chemical shift (ppm)	base pair	NOE connectivity to
A_1	14.18	$A_{20}{\cdot}U_{59}$ and $U_{32}{\cdot}A_{46}$	F_1, I″
B	13.90	$A_{18}{\cdot}U_{61}$	I″
D_1	13.44	unassigned	H
F_1	13.18	$G_{21}{\cdot}C_{58}$	A_1
H	12.73	unassigned	D_1
I	12.52	$C_{29}{\cdot}G_{49}$	J_1
I″	12.44	$G_{19}{\cdot}C_{60}$ and $G_{31}{\cdot}C_{47}$	A_1, B, J_1
J_1	12.19	$C_{30}{\cdot}G_{48}$	I, I″

Base Pairs in Fragment A (Helices II and III)—Tentative Assignment. The RNase T1 cleaved fragment A of *B. megaterium* 5S rRNA contains helices II and III (see Figures 1 and 4). Eight distinct peaks (which contain more than 10 imino proton resonances) are observed in the downfield proton NMR spectrum of fragment A, and many of these resonances in this fragment are well resolved (at the same chemical shifts) in the spectrum of intact 5S rRNA, indicating that the secondary structure in the intact 5S rRNA is preserved in fragment A.

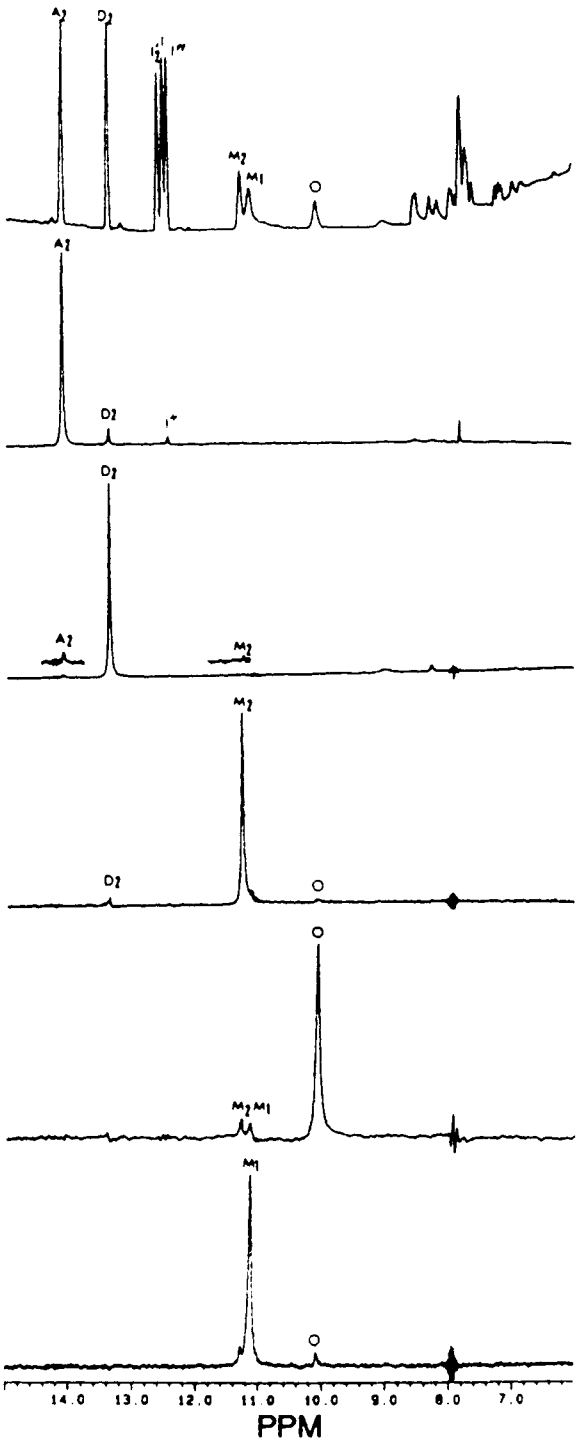

FIGURE 5: NOE difference spectra of resonances A_2, D_2, M_2, O, and M_1 of fragment B. It is evident that D_2 (G·C) is connected to A_2 (A·U) and M_2 (unpaired U) and that M_2 is connected to D_2 and O (unpaired U). Irradiation of A_2 shows NOE connectivity to D_2 and I″ (G·C). From these spectra, the sequence I″ (G·C)-A2 (A·U)-D_2 (G·C)-M_2 (U)-O (U)-M_1 (U) can be confidently assigned.

NOE connectivities obtained by successive irradiation of these resonances are summarized in Table II. Peak I″ (G·C) actually consists of two separate peaks (two imino proton resonances) whose chemical shifts are too close to be irradiated selectively. Peak A_1 (A·U) also consists of two resonances, one of which melts at a slightly elevated temperature (Figure 6). The other resonance, A_1 (A·U), together with F_1 (G·C) and I″ (G·C), is quite stable at high temperature. From the NOE connectivities, the base-pair sequence information can be inferred as follows. Resonance F_1 is NOE connected to A_1 and resonance A_1 to F_1 and I″. Peak I″ (G·C) is connected

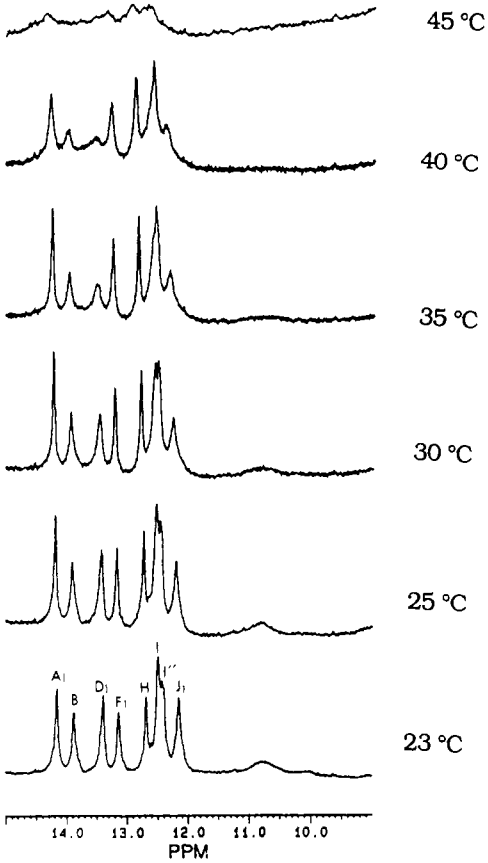

FIGURE 6: Heat-induced melting pattern of the proton 500-MHz NMR spectrum of fragment A in 10 mM cacodylic acid, 0.1 M NaCl, and 1 mM EDTA at pH 7.0. The high thermal stability of peaks F_1-A_1I″ (tentatively assigned as G_{21}·C_{58}-A_{20}·U_{59}-G_{19}·C_{60}) suggests that helix II is more stable than helix III.

to A, B (A·U), and J_1 (G·C) and peak J_1 to I″ and I (see Table II). At this stage, there are several possible combinations of resonances for the establishment of base-pair sequences since peaks A_1 and I″ each contain two resonances. Although unequivocal assignment is not available, it is very likely that F_1-A_1-I″-B is G_{21}·C_{58}-A_{20}·U_{59}-G_{19}·C_{60}-A_{18}·U_{61} and A_1-I″-J_1-I is U_{32}·A_{46}-G_{31}·C_{47}-C_{30}·G_{48}-C_{29}·G_{49}, since those assignments fully account for the observed NOE connectivities manifested in the NOE difference spectra produced by successive irradiation at peaks A_1, B, F_1, I, I″, and J_1. The high stability of one of the A_1 peaks (A_{20}·U_{59}) observed in the variable-temperature experiment (Figure 6) can be explained by its sandwiched location between two G·C pairs (G_{21}·C_{58} and G_{19}·C_{60}) (Table II).

A few secondary structural questions remain. The two G·C resonances (H and D_1), which show strong mutual NOE connectivity, are not assigned. It is interesting to note that a G·C pair which melts at slightly higher temperature (D_1) is spatially adjacent to another G·C pair which is very stable at high temperature (H) (see Figure 6). Another peak, I′, which is weak in intensity and hidden between two intense signals (H and I), cannot be assigned at this juncture. The possible sources of these extra resonances are intraloop base pairs and the base pairs from the region between helices II and III. Intraloop base pairs have been reported to be unlikely in chloroplast 5S rRNA from spinach (Romby et al., 1988). The extra resonances in fragment A of *B. megaterium* 5S rRNA do not seem to correspond to intraloop base pairs in the light of their NOE connectivities; they more likely result from possible base pairs in the region between helices II and III. Imino protons located near the open end of the fragment

appear to exchange rapidly with solvent, since no resonances for A_{17}·U_{62} or G_{16}·U_{63} are observed and peak B (A_{18}·U_{61}) melts at relatively lower temperature than most peaks. A very broad, very weak signal (which melts out completely at 35 °C) between 10.5 and 11 ppm is believed to arise from the resonances of the imino protons of unpaired uridines or guanidines inside the loop, whose exchange rates with water are not slow enough to show detectable signals.

Unlike "fragment 2" of *E. coli* 5S rRNA (Leontis & Moore, 1986), the helix II and III region of *B. megaterium* does not show as many imino resonances—no resonances at 9–12 ppm of fragment A whereas several resonances can be found in this region of "fragment 2"—suggesting that fragment A of *B. megaterium* 5S rRNA may not be so extensively base paired in those helical regions as the corresponding segments of *E. coli* 5S rRNA. No evidence for conformational flexibility has yet been demonstrated. Possible base pairs A_{27}·U_{53}-C_{28}·G_{52} or A_{25}·U_{53}-C_{26}·G_{52} cannot be identified at this stage. Melting of fragment A appears to be quite reversible. Virtually all imino proton resonances disappear by ~45 °C, but the original resonances reappear on a decrease in temperature.

Although our tentative assignments of resonances from fragment A are not complete, they do establish that helices II and III are present in 5S rRNA of Gram-positive bacteria. As base pairs in helices II and III have been already identified and assigned in 5S rRNA from a Gram-negative bacterium, *E. coli* (Leontis & Moore, 1986), and a eukaryote, *Triticum aestivum* (wheat germ) (Li & Marshall, 1986; Li et al., 1987), the helical segments II and III appear to occur universally among prokaryotic (both Gram-positive and Gram-negative bacteria) and eukaryotic 5S rRNAs.

Base Pairs in Fragment C (Helix I). The downfield proton NMR spectrum of RNase T1 cleaved fragment C (mainly the helix I region) of *B. megaterium* 5S rRNA shows more than a dozen peaks arising from many imino proton resonances in the terminal stem (Figure 4). The sequence of base pairs G_4·C_{112}-G_5·C_{111}-U_6·G_{110}, previously inferred from NOE difference spectra of intact 5S rRNA, can now be confirmed with greater confidence from NOE difference spectra obtained by irradiating resonances J, E, K_1, and M in the more highly resolved spectrum of fragment C, which gave the same NOE connectivities as observed in 5S rRNA. Peaks K_1 and K_2, which could not be resolved in the spectrum of intact 5S rRNA, gave two separate signals (very close in chemical shift) in a spectrum of fragment C computed from a 32K time domain data set. Hence, K_1 can be assigned as the N3-H resonance of the U_6 paired with G_{110}. Melting of fragment C, like that of intact 5S rRNA, was reversible, suggesting thermodynamic stability of the terminal helix.

Conformations of Enzyme-Cleaved Fragments and the Corresponding Segments of Intact 5S rRNA. Comparison of the proton NMR spectra of the enzyme-cleaved 5S rRNA fragments with the proton NMR spectrum of intact 5S rRNA serves to establish whether or not the fragments retain the original conformation of the intact 5S rRNA molecule. Since virtually all of the imino proton resonances observed in the spectra of fragments A–C (Figure 4) can be found at the same chemical shifts in intact 5S rRNA, the base-paired regions of the fragments are likely to be highly homologous in secondary structure to those of intact 5S rRNA. Moreover, demonstration of the same NOE connectivity for an imino proton resonance of a fragment and the corresponding resonance in intact 5S rRNA further corroborates the similarity of conformations between them (Li & Marshall, 1986). For example, the NOE difference spectra obtained by irradiating

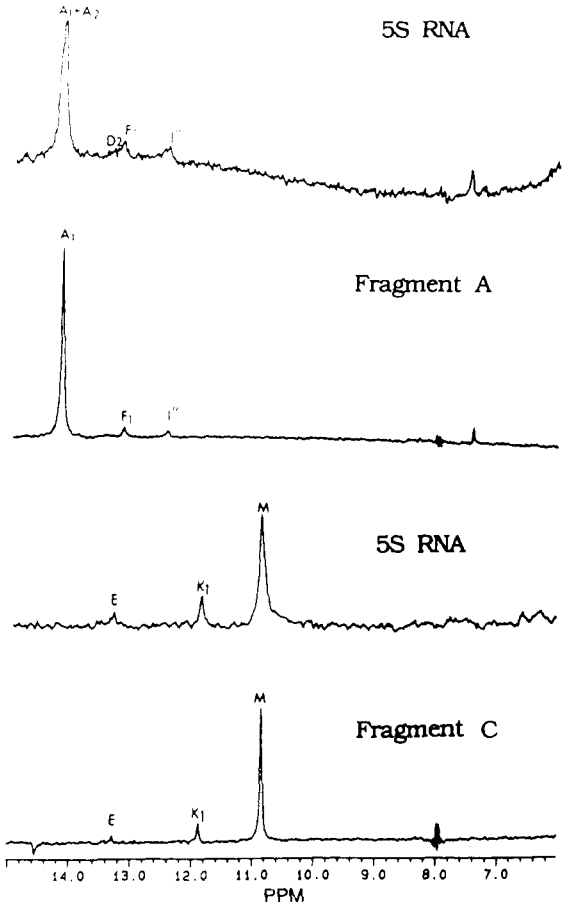

FIGURE 7: Comparison of the NOE connectivities of intact 5S rRNA with those of the corresponding enzyme-cleaved fragments. This figure shows that irradiation of peak A in intact 5S RNA (top spectrum, in which peaks A_1 and A_2 are irradiated simultaneously) gives several NOE difference peaks, including D_2, F_1, and I″, which are also seen by irradiation of A_1 of fragment A (NOE difference peaks at F_1 and I″ in this figure) and by irradiation of A_2 of fragment B (NOE difference peaks at D_2 and II″ in Figure 5). Retention of the original structure of intact 5S rRNA in fragment C is shown by comparing the NOE connectivities of peak M in 5S rRNA and fragment C (two lower-most spectra). These homologous NOE connectivities in intact 5S rRNA and its enzyme-cleaved fragments strongly support the retention of secondary structure on enzymatic cleavage.

resonances K_1/M, E, and J of fragment C and those of intact 5S rRNA are nearly identical (two lower-most spectra of Figure 7). Furthermore, irradiation of peak A of intact 5S rRNA shows NOE connectivities to resonances I″, F_1, and D_2, which are observed separately by irradiation of peak A_1 in fragment A (NOE connectivity to F_1 and I″) and peak A_2 in fragment B (NOE connectivity to D_2 and I″), as seen in Figures 5 and 7. These near-identical NOE patterns for the fragments and intact 5S rRNA strongly support the retention of conformation on cleavage of 5S rRNA into fragments A–C.

Base Pairs in Other Structural Segments. The existence of helix V in 5S rRNA has been demonstrated from assignment of base pairs in that region for both prokaryotic (Leontis et al., 1986; Chang & Marshall, 1986a; Zhang & Moore, 1989) and eukaryotic (Li & Marshall, 1986; Chen & Marshall, 1986) 5S rRNAs. Unfortunately, we are unable to identify helix V base pairs from *B. megaterium* 5S rRNA and its fragments in this paper. Helix V is shown to be the least stable among all helical regions in 5S rRNA, whereas helices I and IV are the most stable (Chang & Marshall, 1986b). The existence of loop E (the internal loop between helices IV and V) has been confirmed in chloroplast 5S rRNA from spinach (Romby et al., 1988) and "fragment 1" of *E. coli* 5S rRNA

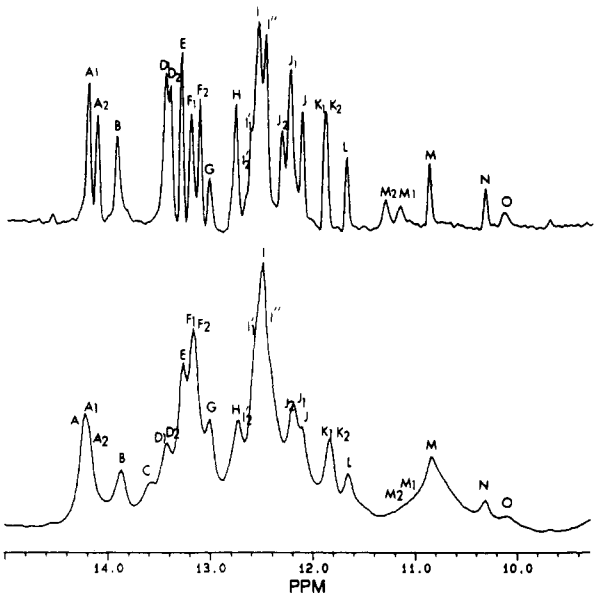

FIGURE 8: 500-MHz ^{1}H NMR spectrum of intact 5S rRNA (bottom) and composite spectrum obtained by the addition of the spectra of fragments A–C (top). The close match in chemical shifts between the two spectra (taking into account the broader lines for the larger intact 5S rRNA molecule) offers further evidence for high conformational similarity between intact 5S rRNA and fragments. Some peaks present in intact 5S rRNA are absent in the combined spectrum (e.g., peak C). The missing peaks in the combined spectrum are believed to result from secondary base pairs in helix V and from tertiary base pairs.

(Zhang & Moore, 1989). Zhang and Moore have reported that an extensively base-paired model for loop E is incompatible with the NMR data obtained, suggesting that many of the bases in the loop are only slightly protected from solvent exchange and that loop E may well not contain any conventional base pairs, thereby resulting in a shortened helix V segment. In *B. megaterium* 5S rRNA, helix V not only is relatively short but also contains a noncanonical base pair (A_{102}·R_{70}) in the middle of the helix in addition to the weak G_{100}·U_{72} base pair (see Figure 1), resulting in a relatively unstable helix V that evidently does not survive the enzymatic cleavage process. Comparison of the spectrum of intact 5S rRNA with the spectrum obtained by the addition of the spectra of fragments A–C (Figure 8) reveals some peaks that are present in intact 5S rRNA but not in those fragments. For example, peak C is clearly absent in the fragments. The missing resonances not observed in the fragments probably arise from base-paired imino protons in helix V and/or tertiary base pairs (Pieler & Erdmann, 1982; Chang & Marshall, 1986b). As one might expect for helix V or tertiary base pairs, these resonances disappear quickly as the temperature increases (see Figure 3), leaving the more stable base pairs assigned above.

CONCLUSIONS

In this paper, the complete secondary structure of the prokaryotic loop (helix IV) of *B. megaterium* ribosomal 5S rRNA, G_{80}·C_{92}-G_{81}·C_{91}-G_{82}·C_{90}-A_{83}·U_{89}-C_{84}·G_{88}, U_{85}, U_{86}, and U_{87}, is described. Also, U_{32}·A_{46}-G_{31}·C_{47}-C_{30}·G_{48}-C_{29}·G_{49} in helix III, G_{21}·C_{58}-A_{20}·U_{59}-G_{19}·C_{60}-A_{18}·U_{61} in helix II, and G_4·C_{112}-G_5·C_{111}-U_6·G_{110} in helix I are assigned. Although the assignments of base pairs in helices I–III are not complete, the present NOE-based conclusions strongly support the *B. megaterium* ribosomal 5S rRNA secondary structure adapted from the Fox and Woese model (Figure 1), offering additional evidence for the universality of the secondary structure of all prokaryotic and eukaryotic 5S rRNAs.

ACKNOWLEDGMENTS

We thank C. E. Cottrell for advice on NMR experiments and the Fermentation Lab of The Ohio State University for providing *B. megaterium* cells. We also thank J. Wu and H.-W. Huang for helpful discussions.

Registry No. G, 73-40-5; C, 71-30-7; A, 73-24-5; U, 66-22-8.

REFERENCES

Arter, D. B., & Schmidt, P. G. (1976) *Nucleic Acids Res. 3*, 1437–1447.

Chang, L.-H., & Marshall, A. G. (1986a) *Biochemistry 25*, 3056–3063.

Chang, L.-H., & Marshall, A. G. (1986b) *Biopolymers 25*, 1299–1313.

Chang, L.-H., Burkey, K. O., Alben, J. O., & Marshall, A. G. (1984) *Biochemistry 23*, 3659–3662.

Chen, S.-M., & Marshall, A. G. (1986) *Biochemistry 25*, 5117–5125.

Delihas, N., & Anderson, J. (1982) *Nucleic Acids Res. 10*, 7323–7344.

De Wachter, R., Chen, M.-W., & Vandenberghe, A. (1982) *Biochemie 64*, 311–329.

Erdmann, V. A., & Wolters, J. (1986) *Nucleic Acids Res. 14*, r1–r55.

Fox, G. E., & Woese, C. R. (1975) *Nature (London) 256*, 505–507.

Gewirth, D. T., Abo, S. R., Leontis, N. B., & Moore, P. B. (1987) *Biochemistry 26*, 5213–5220.

Hare, D. R., & Reid, B. R. (1982) *Biochemistry 21*, 5129–5131.

Heerschap, A., Haasnoot, C. A. G., & Hilbers, C. W. (1983a) *Nucleic Acids Res. 11*, 4483–4499.

Heerschap, A., Haasnoot, C. A. G., & Hilbers, C. W. (1983b) *Nucleic Acids Res. 11*, 4501–4520.

Hore, J. (1983) *J. Magn. Reson. 55*, 283–300.

Hurd, R. E., & Reid, B. R. (1979a) *Biochemistry 18*, 4005–4011.

Hurd, R. E., & Reid, B. R. (1979b) *Biochemistry 18*, 4017–4024.

Johnston, P. D., & Redfield, A. G. (1978) *Nucleic Acids Res. 5*, 3913–3927.

Johnston, P. D., & Redfield, A. G. (1981) *Biochemistry 20*, 1147–1156.

Kearns, D. R., & Wong, Y. P. (1974) *J. Mol. Biol. 87*, 755–774.

Kim, S.-H. (1976) *Prog. Nucleic Acid Res. Mol. Biol. 17*, 182–216.

Kime, M. J., & Moore, P. B. (1983a) *Biochemistry 22*, 2615–2622.

Kime, M. J., & Moore, P. B. (1983b) *Biochemistry 22*, 2622–2629.

Kime, M. J., Gewirth, D. T., & Moore, P. B. (1984) *Biochemistry 23*, 3559–3568.

Lee, K. M., & Marshall, A. G. (1987) *Biochemistry 26*, 5534–5540.

Leontis, N. B., & Moore, P. B. (1986) *Biochemistry 25*, 3916–3925.

Leontis, N. B., Ghosh, P., & Moore, P. B. (1986) in *Biomolecular Stereodynamics, Proceedings of the Fourth Conversation in the Discipline of Biomolecular Stereodynamics*, State University of New York, Albany, NY, June 4–9, 1985 (Sarma, R. H., & Sarma, M. H., Eds.) pp 287–306, Adenine, Guilderland, NY.

Li, S.-J., & Marshall, A. G. (1986) *Biochemistry 25*, 3673-3682.
Li, S.-J., Chang, L.-H., Chen, S.-M., & Marshall, A. G. (1984) *Anal. Biochem. 138*, 465-471.
Li, S.-J., Wu, J., & Marshall, A. G. (1987) *Biochemistry 26*, 1578-1585.
Luehrsen, K. R., & Fox, G. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2150-2154.
Luoma, G. A., & Marshall, A. G. (1978a) *Proc. Natl. Acad. Sci. U.S.A. 75*, 4901-4905.
Luoma, G. A., & Marshall, A. G. (1978b) *J. Mol. Biol. 125*, 95-105.
Marshall, A. G., & Wu, J. (1989) *Biol. Magn. Reson.* (in press).
Nishikawa, K., & Takemura, S. (1974) *J. Biochem. (Tokyo) 76*, 935-947.
Osterberg, R., Sjoberg, B., & Garrett, R. A. (1976) *Eur. J. Biochem. 68*, 481-487.
Pieler, T., & Erdmann, V. A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4599-4603.
Pieler, T., Digweed, M., & Erdmann, V. A. (1984) in *Gene Expression* (Clark, B. F. C., & Petersen, H. U., Eds.) pp 353-376, Munksgaard, Copenhagen.
Redfield, A. G., Kunz, S. D., & Ralph, E. K. (1975) *J. Magn. Reson. 19*, 114-117.
Reid, B. R. (1981) *Annu. Rev. Biochem. 50*, 969-996.
Reid, B. R., McCollum, L., Ribeiro, N. S., Abbate, J., & Hurd, R. E. (1979) *Biochemistry 18*, 3996-4005.
Rich, A. (1977) *Acc. Chem. Res. 10*, 388-395.
Romby, P., Westhof, E., Toukifimpa, R., Mache, R., Ebel, J.-P., Ehresmann, C., & Ehresmann, B. (1988) *Biochemistry 27*, 4721-4730.
Roy, S., & Redfield, A. G. (1983) *Biochemistry 22*, 1386-1390.
Schimmel, P. R., & Redfield, A. G. (1980) *Annu. Rev. Biophys. Bioeng. 9*, 181-221.
Studnicka, G. M., Eiserling, F. A., & Lake, J. A. (1981) *Nucleic Acids Res. 9*, 1885-1904.
Zhang, P., & Moore, P. B. (1989) *Biochemistry 28*, 4607-4615.

Apolipoprotein(a) Size Heterogeneity Is Related to Variable Number of Repeat Sequences in Its mRNA†

Marlys L. Koschinsky,‡ Ulrike Beisiegel,§ Doris Henne-Bruns,§ Dan L. Eaton,‡ and Richard M. Lawn*,‡

Department of Cardiovascular Research, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, and Department of Medicine and Surgery, University of Hamburg, Martinistrasse 52, 2000 Hamburg 20, FRG

Received June 30, 1989; Revised Manuscript Received August 25, 1989

ABSTRACT: Plasma apolipoprotein(a) [apo(a)] shows considerable size heterogeneity, existing as discrete glycoprotein isoform variants that range in apparent molecular mass from approximately 400 to 800 kDa. To study the molecular basis of protein size variability, we have isolated liver RNA from individuals with different apo(a) isoforms, and identified apo(a)-specific transcripts using Northern blot analysis. Transcript sizes were shown to be variable (8.0-12 kb) and in all cases were closely correlated with protein masses (590-850 kDa) as determined from immunoblots. Thus, it is almost certain that apo(a) isoform size variation is due to allelic differences in the number of its tandemly repeated sequences of 114 amino acids that resemble kringle four of plasminogen. The high carbohydrate content of apo(a) makes true molecular weight estimations in SDS-PAGE gels difficult. However, a recombinant form of apo(a) containing 17 kringle repeats (calculated molecular mass of 250 kDa) migrates on SDS-PAGE gels only slightly below apoB-100, with an apparent molecular mass of approximately 500 kDa. Since smaller protein isoforms have been observed in the population, this suggests that plasma apo(a) isoforms contain from less than 17 to greater than 30 tandemly repeated kringle units.

Elevated levels of serum Lp(a)[1] are strongly correlated with atherosclerosis in human populations and Lp(a) has been assessed as an independent risk factor in the development of cardiovascular disease (Rhoads et al., 1986; Dahlen et al., 1986; Durrington et al., 1988). Lp(a) closely resembles low-density lipoprotein (LDL) in both lipid composition and the presence of apolipoprotein B-100 (apoB-100). Unlike LDL, Lp(a) contains an additional protein designated apolipoprotein(a) [apo(a)] which is bound to apoB-100 by a disulfide linkage (Gaubatz et al., 1983; Utermann & Weber, 1983; Armstrong et al., 1985; Fless et al., 1985). Apo(a) is a large glycoprotein that is synthesized primarily in the liver (Tomlinson et al., 1989; Kraft et al., 1989) and exhibits considerable size heterogeneity in the human population (Fless et al., 1984; Gaubatz et al., 1987; Utermann et al., 1987, 1988a,b). Marked variability has also been observed with respect to plasma Lp(a) levels, which range from less than 1 over 100 mg/dL in the population. An inverse relationship between isoform size and mean plasma concentration in the population has been observed (Utermann et al., 1987, 1988a). The fact

† This work was supported by Genentech, Inc., and the Deutsche Forschungsgemeinschaft. M.L.K. is partially funded by a postdoctoral fellowship from the Medical Research Council of Canada.

* To whom correspondence should be addressed.
‡ Genentech, Inc.
§ University of Hamburg.

[1] Abbreviations: apo(a), apolipoprotein(a); apoB-100, apolipoprotein B-100; Lp(a), lipoprotein(a); LDL low-density lipoprotein; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis.

0006-2960/90/0429-640$02.50/0