NMR Analysis of Helix I from the 5S RNA of *Escherichia coli*†

S. A. White,‡,§ M. Nilges,∥,⊥ A. Huang,§ A. T. Brünger,∥,⊥ and P. B. Moore*,§,∥

Department of Chemistry, Department of Molecular Biophysics and Biochemistry, and The Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06511

Received August 1, 1991; Revised Manuscript Received October 29, 1991

ABSTRACT: The structure of helix I of the 5S rRNA from *Escherichia coli* has been determined using a nucleolytic digest fragment of the intact molecule. The fragment analyzed, which corresponds to bases (−1)–11 and 108–120 of intact 5S rRNA, contains a G–U pair and has unpaired bases at its termini. Its proton resonances were assigned by two-dimensional NMR methods, and both NOE distance and coupling constant information have been used to calculate structural models for it using the full relaxation matrix algorithm of the molecular dynamics program XPLOR. Helix I has A-type helical geometry, as expected. Its most striking departure from regular helical geometry occurs at its G–U, which stacks on the base pair to the 5′ side of its G but not on the base pair to its 3′ side. This stacking pattern maximizes interstrand guanine–guanine interactions and explains why the G–U in question fails to give imino proton NOE's to the base pair to 5′ side of its G. These results are consistent with the crystal structures that have been obtained for wobble base pairs in $tRNA^{Phe}$ [Mizuno, H., & Sundaralingam, M. (1978) *Nucleic Acids Res. 5*, 4451–4461] and A-form DNA [Rabbinovich, D., Haran, T., Eisenstein, M., & Shakked, Z. (1988) *J. Mol. Biol. 200*, 151–161]. The conformations of the terminal residues of helix I, which corresponds to bases (−1)–11 and 108–120 of native 5S RNA, are less well-determined, and their sugar puckers are intermediate between C2′ and C3′-endo, on average.

The structure of 5S RNA has been investigated for decades using every physical and chemical technique imaginable, including nuclear magnetic resonance spectroscopy, which is one of the most powerful. Nonetheless the conformation of 5S RNA[1] is still not fully understood [see Marshall and Wu (1989)]. The reason 5S RNA has not been solved by NMR is its size; it is far larger ($M_r \approx 40000$) than the largest protein whose structure has been determined that way. Indeed, the study of 5S RNA by NMR would be quixotic if were it not possible to divide the molecule into smaller fragments that retain the conformations they have in intact 5S RNA.

Below we show that an RNA fragment that consists of helix I and a few adjoining bases [residues (−1)–11 and 108–120] can be prepared from the 5S RNA of *Escherichia coli* by nucleolytic digestion (Figure 1). The imino proton spectrum of this fragment, which we shall call "helix I", closely resembles that of the helix I region of intact 5S RNA, indicating that their conformations are similar. In addition to its relevance to 5S RNA, helix I is of interest because it includes an example of the most commonly encountered mismatched pair in RNA, a G–U (G9–U111). It also has unpaired bases at both ends, which again is commonplace. Assignments are reported for all of helix I's imino, aromatic, anomeric, 2′, and 3′ proton resonances, as well as for a number of its other proton resonances and several of its phosphorus resonances. A model for helix I is presented that has been derived from nuclear Overhauser effect (NOE) buildup data and coupling constant information, using XPLOR version 2.2 (Brünger, 1990, 1991; Nilges et al., 1991) that minimizes discrepancies between observed and model NMR data.

Helix I, as a whole, is an A-type helix, as we had ample reason to anticipate. G9–U111's placement with respect to its neighbors is strongly influenced by stacking interactions involving G9 and the purines of the base pairs that flank it, G10 and G112. Thus G9 forms an intrastrand stack with G112 and an interstrand stack with G10. The resulting asymmetry of G9–U111's position with respect to its neighbors accounts for the absence of NOE's between G112's imino proton and either of the imino protons belonging to G9–U111, which had been noted earlier (Gewirth et al., 1987). The terminal unpaired bases stack on their neighbors well enough to give the NOE connectivities characteristic of A-type helices, suggesting that the molecule's helical geometry might extend to its termini. However, their NOE's are consistent with a variety of conformations, some of which differ substantially from standard A-type geometry.

Materials and Methods

Deuterated Nucleosides. 5,6-Dideuterocytidine was prepared by dissolving 1 g of cytidine, which had previously been lyophilized from D_2O, in 10 g of dimethyl sulfoxide-d_6 (MSD Isotopes), adding 0.395 g of sodium methoxide in 3 mL of methanol-d_1, and incubating for 24 h at 60 °C (Rabi & Fox, 1973). Guanosine-d_8 was prepared by base-catalyzed exchange. Five grams of guanosine was dissolved in 200 mL of D_2O and 5 mM boric acid, pH 9.5, and incubated at 95 °C

† This work was supported by NIH Grants AI-09167 and GM 41651 (to P.B.M.) and the Howard Hughes Medical Institute (A.T.B. and M.N.), and one of us (S.W.) held an American Cancer Association postdoctoral fellowship.

* To whom correspondence should be addressed.

‡ Present address: Department of Chemistry, Bryn Mawr College, Bryn Mawr, PA 19010.

§ Department of Chemistry.

∥ Department of Molecular Biophysics and Biochemistry.

⊥ The Howard Hughes Medical Institute.

[1] Abbreviations: COSY, correlation spectroscopy; DQF, double-quantum filtered spectroscopy; EDTA, ethylenediaminetetraacetic acid; FID, free induction decay; HPLC, high-performance liquid chromatography; ISPA, isolated spin pair approximation; MES, 2-(*N*-morpholino)ethanesulfonic acid; NMR, nuclear magnetic resonance; NOESY, nuclear Overhauser spectroscopy; NOE, nuclear Overhauser effect; ppm, parts per million.

0006-2960/92/0431-1610$03.00/0

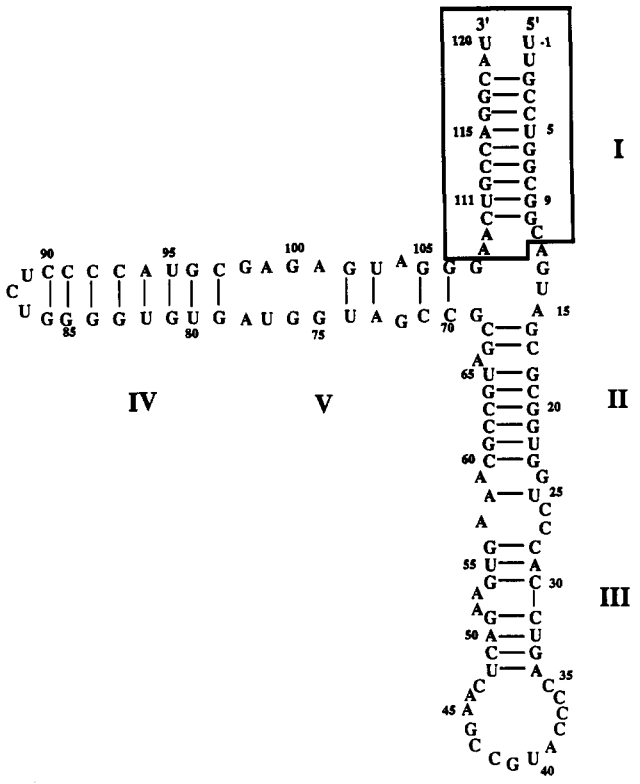

FIGURE 1: 5S rRNA sequence. The 5S ribosomal RNA sequence from the *rrnB* cistron of *E. coli* is shown with helix I boxed.

for 16 h. Uridine-d_5 was prepared following the procedure of Heller (1968). A D_2O solution was prepared that contained 10 mM uridine and 10 mM mercaptoethylamine, pH 9.5, and was incubated at 65 °C for 4 days. 2,8-Dideuteroadenosine was prepared for us by Mr. Gregory Kellogg using a modification of the method of Maeda and Kawazoe (1975). Platinum catalyst was prepared by reducing 1 g of $PtCl_2$ in 5 mL of D_2O with $NaBD_4$ (Calf & Garnett, 1964). The D_2O was then decanted and the solid washed several times with additional D_2O. The catalyst was suspended in 40 mL of argon-purged D_2O, and 5 g of adenosine that had been lyophilized twice from D_2O was added. The mixture was stirred under argon for 48 h at 90–95 °C. The catalyst was removed by filtration and 2,8-dideuteroadenosine was recovered on cooling and partial evaporation.

Synthetic RNA. A blunt-ended derivative of helix I, which we call "blunt helix I", was made from RNA oligomers synthesized at the Yale University School of Medicine Protein and Nucleic Acid Chemistry Facility on an Applied Biosystems 380B synthesizer. Two oligomers were deprotected and purified according to Webster and Spicer (1991). The two strands, whose sequences are 5′-UGCCUGGCGGC-3′ and 5′-GCUGCCAGGCA-3′, were annealed at 50 °C and allowed to cool slowly. Blunt helix I starts at U1–A119 and helix I's A109 has been changed to a G (in bold) to form a second base pair after the G–U pair.

5S RNA. 5S RNA was obtained from *E. coli* HB101 transformed with pKK5-1 by overproduction in the presence of chloramphenicol (Brosius et al., 1981; Moore et al., 1988). Selectively deuterated 5S molecules were made by supplying appropriately labeled bases to *E. coli* KL375/pKK5-1, which is auxotrophic for adenine, guanine, and either cytidine or uridine. KL375 was created by Dr. Brooks Low at Yale University from H724 (Nijkamp & De Haan, 1967) by transposition with Tn10 into pyrD34. KL375/pKK5-1 was grown on minimal medium supplemented wth casamino acids, tyrosine, tryptophan, histidine, thiamine, and 0.2 g/L adenosine, guanosine, and cytidine or uridine, deuterated as required (Moore et al., 1988). To minimize the loss of label from guanine-d_8 due to exchange and to circumvent its low solubility, 0.1 g/L was included in the starting medium and the remaining 0.1 g/L portion added with the chloramphenicol. The following partially deuterated 5S RNA's were produced: (1) 5,6-dideuteropyrimidine, 2,8-dideuteroadenine, 8-deuteroguanine 5S RNA ("perdeutero-5S RNA"), (2) pyrimidine-d_5 5S RNA, and (3) 8-deuteroguanine, 5,6-dideuteropyrimidine 5S RNA ("proteo-A 5S RNA").

The level of deuteration achieved in the pyrimidine-d_5 sample, which was particularly important for these studies (see below), was estimated by comparing the H5/H6 cross peak volumes of COSY spectra of protonated helix I with those in COSY spectra of pyrimidine-d_5 helix I, using the H1′/H2′ cross peaks for normalization. By this criterion, the deuteration achieved in these samples was 95% or better at the pyrimidine 5-position.

Helix I. 5S RNA samples were converted into fragment 1 by partial digestion with RNase A as described previously (Moore et al., 1988). Digestion conditions were chosen so that most of the fragment 1 was cleaved at position 87, yielding a complex consisting of three oligonucleotides: strand IV (bases 1–11), strand III (bases 69–87), and strand II (bases 90–120) (Kime et al., 1984). Strand III was removed from the fragment 1 by Sephadex chromatography in EDTA (Moore et al., 1988), leaving a complex of strands II and IV behind.

Strands II and IV can be separated by chromatography on Sephadex in urea-containing buffers (Moore et al., 1988), and the two separate strands can be reannealed once denaturant has been removed (J. Kim and P. Moore, unpublished observations). Following dialysis against 0.1 M KCl, 2 mM EDTA, and 10 mM MES, pH 6, strands II and IV are incubated together at 50 °C at a total RNA concentration of 20 OD_{260nm}/mL and allowed to cool slowly. Two hybrid strand II/IV complexes were prepared this way, one whose strand II side was made from perdeutero-5S RNA and whose strand IV was protonated, and a second in which the labeling pattern was reversed.

Once strand II/IV complexes of the appropriate isotopic composition were obtained, they were digested with T_1 RNase (0.2 units/μL) in 10 mM Tris and 1 mM EDTA at 25 or 37 °C at an RNA concentration of 0.1 $OD_{260nm}/\mu L$ for 45 min. T_1 RNase was removed from the preparation by two rounds of phenol extraction. Helix I was separated from the other T_1 fragments by HPLC on a Nucleogen DEAE 500-7 anion exchange column at room temperature, under nondenaturing conditions. It was eluted from the column using a linear gradient running from 0 to 1 M LiCl in 20 mM NaOAc, pH 5.5, 20% CH_3CN. Peak fractions were pooled and their volumes reduced about 10-fold by rotary evaporation. The RNA was then recovered by ethanol precipitation.

Sequencing. Sequences were verified enzymatically using T_1, U_2, Phy M, and CL_3 RNases (BRL) on a 20% denaturing gel 0.4 mm thick. Helix I was end-labeled with [^{32}P]cytidine 3′,5′-bisphosphate (NEN) using T4 RNA ligase (BRL). The strands of the labeled helix were purified on a 20-cm denaturing gel, and the RNA was extracted using the "freeze and squeeze" method (Kean & Draper, 1985). Unfortunately, helix I is not fully denatured under these gel conditions , and only the longer A108–U120 strand could be isolated and sequenced.

NMR Samples. Ethanolic precipitates containing roughly micromolar amounts of helix I were resuspended in 1–2 mL of 1.05× NMR buffer (1× NMR buffer is 0.1 M KCl, 4 mM $MgCl_2$, 4 mM cacodylic acid, and 0.2 mM EDTA, pH 7). Samples were then dialyzed exhaustively against 1.05× NMR buffer and concentrated down to volumes of about 0.4 mL using Centricon-3 ultrafilters (Amicon). The fully protonated helix I and the pyrimidine d_5 samples were 3 mM and 2 mM, respectively, while the other samples examined were about 1 mM.

Samples were made about 1 mM in dioxane, to provide a chemical shift marker, and 7% D_2O (v/v) was added for the NMR lock signal. The chemical shift of dioxane is taken to be 3.741 ppm at all temperatures. Samples intended for spectroscopy in D_2O were lyophilized and redissolved in D_2O at least three times. 99.996% D_2O (Cambridge Isotopes) was used as the final solvent.

NMR Spectroscopy. NMR data were collected on either a Bruker AME-500 or the Yale Chemical Instrumentation Center 490-MHz spectrometer. Data were processed using FTNMR (Hare Research, Inc.).

Spectroscopy in H_2O. One-dimensional spectra in H_2O were collected routinely at 30 and 23 °C using the twin pulse method (Kime & Moore, 1983a) for water suppression. The offset was either at the downfield or the upfield edge of the imino region, and the sweep width was 10000–15000 Hz. The decoupler power used for one-dimensional NOE's was set so that the intensity of the irradiated peak was reduced about 75%. The relaxation delay was 0.5–2 s. NOE data were collected in an interleaved mode with 16K data points and 200–1000 scans averaged per FID. Difference FID's were obtained by subtracting the off-resonance and on-resonance FID's and then were transformed with modest line broadening.

Imino proton distances were determined using NOE's collected at three mixing times. Initial slopes of the resulting NOE buildup curves were converted into distances using the relationship

$$\frac{R_{ab}}{R_{ij}} = \left(\frac{\mathrm{NOE}_{ij}}{\mathrm{NOE}_{ab}}\right)^{1/6}$$

taking the A115H2 to U5H1 reference distance to be 2.8 Å.

NOESY Spectra. Most two-dimensional NMR experiments had the following setup parameters: 2-s cycle delay, 5000-Hz proton sweep width, offset on the residual HDO peak, and modest presaturation to minimize the HDO peak. The phases of the decoupler and the synthesizer were synchronized using a power splitter (Zagorski, 1990). Experiments were routinely run at 30 °C without spinning. NOESY spectra were measured for mixing times of 50, 100, 200, and 300 ms, and 96–200 FID's were averaged for each t_1 point. Data were collected using States phase cycling (States et al., 1982), and 250–400 t_1 points were measured, each consisting of 4K complex data points. The total time for a typical experiment was 48–60 h. Data sets were processed using a shifted sine bell window function in both dimensions, and the shift was 90° when volume integrals were required. The first row was multiplied by 0.5 to reduce noise ridges (Otting et al., 1986). NOESY experiments done on water-containing samples were set up using the same parameters as the one-dimensional NOE experiments just described except that a twin pulse replaced the final 90° read pulse and mixing times were 200 or 300 ms.

COSY Spectra. Magnitude COSY spectra were obtained according to Aue et al. (1976) and Nagayama et al. (1980), with both the experimental setup and the processing similar to the NOESY experiments just described. Double-quantum filtered COSY spectra (Shaka & Freeman, 1983) were collected according to States (1982). Coupling constants were measured by processing for high resolution and then comparing the maxima and minima of the antiphase peaks with the line width (Bax & Lerner, 1988). A relay COSY experiment was done (Wagner, 1983) in which the offset was placed in an empty spectral region just downfield of HDO, no presaturation was used, and the sweep width was narrowed to 3600 Hz. The delay time was 0.06 s, which is optimal for coupling constants of about 4 Hz, and 450 t_1 points were collected.

T_1 Relaxation Measurements. Both selective and nonselective inversion–recovery experiments were conducted to estimate the T_1 relaxation times of aromatic resonances. Individual resonances were inverted by applying a short pulse through the decoupler channel and then allowing the magnetization to recover for variable amounts of time from 0 to 2 s before acquiring the spectrum. Nonselective experiments were done by applying a hard 180° pulse and then proceeding as above. Although the 5-s cycle delay time used was not greater than five times T_1 in every case, logarithmic plots of intensity versus time were linear, and the T_1 estimates obtained are believed to be acceptably accurate.

^{31}P Spectroscopy. One-dimensional ^{31}P spectra were collected in D_2O with broad-band decoupling using an inverse probe tuned to phosphorus. The spectrometer frequency was 202.45 MHz. ^{31}P chemical shifts were referenced to an external standard of 85% phosphoric acid. The offset was set on the reference frequency and 500–2000 FID's were averaged using a sweep width of 10000 Hz and 32K data points. The receiver gain was set to its maximum value, and the relaxation delay was 2 s. Spectra were processed with a modest amount of resolution enhancement. A proton-detected, phosphorus–proton heteronuclear COSY experiment (Sklenar et al., 1986; Sklenar & Bax, 1987) was done in the inverse mode. The phosphorus offset was the reference frequency, the sweep width was 1600 Hz, and 4K data points were collected. The proton offset was at 9800 Hz and the sweep width was 2000 Hz. A total of 128 FID's were averaged for each t_1 point, and 128 t_1 points were collected.

Integration of NMR Data Sets. Cross peak volumes were measured using the integration routine in FTNMR (Hare Research), and the results were tabulated using locally written software. For each spectral region, the background integrals for "empty" areas were subtracted from each measurement. To gain confidence in the volumes generated in FTNMR, some measurements were repeated using the interactive integration routine described by Holak et al. (1987). Peaks which overlapped were integrated as a whole, and all possible contributions to their volumes were considered during the XPLOR analysis. Approximately 200 cross peak volumes, evaluated at each of four mixing times, were used in the model building calculations described below. Errors were assigned to cross peak volumes on the basis of comparisons with the symmetry-related peaks, the extent of isolation of the peaks, and the reproducibility of the measurement. Typical error estimates ranged from 5 to 30%.

Model Calculations. Structural models for helix I were refined using a two-stage protocol in which the CHARM, all-hydrogen, nucleic acid force field (Nilsson & Karplus, 1986) was used throughout, and planarity force constants were increased to ensure planarity of the bases. In stage I, the buildup of intensity on each NOE peak was fit to an equation of the form $(at + bt^2)$, where a and b are constants and t is the mixing time. The linear coefficient was used to estimate the intensity that would have been observed for each cross peak

in the absence of spin diffusion at 100 ms. These "spin-diffusion-corrected" intensities were converted into distance restraints using the isolated spin pair approximation (ISPA) and scaled as described below. Care was taken to exclude peaks whose intensities were expected to have important spin diffusion contributions, such as the sequential H1'(n) to (H8 or H6)(n+1) NOE. This NOESY distance information was supplemented by torsion angle restraints obtained from coupling constant measurements. Intermediate models for helix I were then obtained by refining a regular energy-minimized A-type RNA model for helix I using these experimental constraints and conventional distance-restrained molecular dynamics.

A starting model was constructed from A-form crystallographic coordinates. Initial restrained molecular dynamics at 300 K was performed for 0.5 ps to relieve any bad contacts, and we refer to the product as the "initial model". The "spin-diffusion-corrected" data and the torsional angle constraints were then applied. The NOE force constant was increased gradually to a value of 5 kcal mol^{-1} Å^{-2}, while the harmonic restraints, which maintained the atoms in their reference positions, were gradually reduced from 8 to 0 kcal (Bruccoleri & Karplus, 1986). The NOE force constant was intentionally kept low in this first stage so that the models produced would sample a sufficient variety of conformations. Finally, 0.5 ps of restrained molecular dynamics at 300 K was followed by 500 steps of energy minimization.

In the second stage of the refinement protocol, spin diffusion was taken into account by refining back-calculated, full matrix, NOESY intensities directly against the experimental data. The experimental part of the target function, $E_{\text{relaxation}}$, is expressed as a function of the difference between functions of observed and calculated intensities (Nilges et al., 1991).

$$E_{\text{relaxation}} = K_R \sum_{\text{spectra}} \sum_{i=1}^{N_S} \omega_i \times \text{well}(I_i^{\text{cal}}, k_s, I_i^{\text{obs}}, \Delta_i, n) \quad (1)$$

where K_R is the energy constant for the relaxation term, I_i^{cal} and I_i^{obs} are the calculated and observed intensities, respectively, Δ_i is an error estimate for I_i^{obs}, ω_i is a weight factor, k_s is the calibration factor for each spectrum, and N_S is the number of cross peaks in each spectrum. The function well(a,k_s,b,Δ,n) is defined as the absolute value of the difference between the nth powers of a and b, where b has an error estimate Δ:

$$\text{well}(a,b,\Delta,n) \equiv \begin{cases} (b-\Delta)^n - a^n & \text{if } a^n < (b-\Delta)^n \\ 0 & \text{if } (b-\Delta)^n < a^n < (b+\Delta)^n \\ a^n - (b+\Delta)^n & \text{if } a^n > (b+\Delta)^n \end{cases} \quad (2)$$

The individual error estimates, Δ_i, reflect the integration error due to noise and spectral overlap. Several different values for n were tested, and it was found that a value of $^1/_6$ gave the best results (M. Nilges, unpublished). Analytic derivatives with respect to atomic coordinates were obtained by using the results of Yip and Case (1989).

As a measure of the fit of the refined structure to the NOE data, we use a generalized R factor:

$$R_n = \frac{\sum_{\text{spectra}} \sum_{i=1}^{N_S} \omega_i \times \text{well}(I_i^{\text{cal}}, k_s, I_i^{\text{obs}}, \Delta_i, n)}{\sum_{\text{spectra}} \sum_{i=1}^{N_S} \omega_i \times (k_s, I_i^{\text{obs}}, n)} \quad (3)$$

For $n = 1$, $\Delta_i = 0$, and $\omega_i = 1$, this is the standard R factor used in crystallography (Stout & Jensen, 1989); but we also report values with $n = ^1/_6$ as suggested by James (1990), and we call this $R_{1/6\text{-weighted}}$. Note that for the R factor calculation the individual error estimates, Δ_i, were set to zero.

The value of the calibration factor k_s was evaluated simply as follows:

$$k_s = (\sum_{i=1}^{N'_s} I_i^{\text{cal}})/(\sum_{i=1}^{N'_s} I_i^{\text{obs}}) \quad (4)$$

where the sum runs over all cross peaks which are "well determined". For a peak to be included in the calculation of the calibration factor, its error had to be less than 30%. The calibration factor was determined independently for each mixing time after the data had been scaled roughly using the first serial file of each NOESY experiment. The values determined for k_s for the different mixing times did not differ greatly. A single calibration factor could have been used for all spectra. The advantage of using different calibration factors for different mixing times is that it obviates the need to scale the experimental spectra accurately relative to each other in advance. Also, non-NOE relaxation effects [termed "Z-leakage" in Summers (1990)] are dealt with in a simple way, and one does not have to rely upon the calibration of all measurements to a set of reference peaks. The calibration factor is updated at every step in the calculation. All individual weights ω_i were set to 1.

An isotropic correlation time of 3 ns was used, but the results were relatively insensitive to this value. Likewise, preliminary calculations suggested that no improvement would result from using two correlation times which reflect the fact that helix I is twice as long as it is wide.

The annealing schedule used is described as follows: In seven steps of 0.1 ps, the NOE force constant, i.e., the weight given the ISPA data, was reduced from 5 to 0 kcal mol^{-1} Å^{-2} while the weight given the full relaxation matrix data, i.e., its force constant, was increased from 35 to 350. The value of the force constant K_R assigned to the relaxation energy function is somewhat arbitrary. The value used was high enough to ensure that the structure is fit to the data, without distorting the covalent structure. Restrained molecular dynamics at 300 K was performed for an additional 1 ps and followed by 250 steps of energy minimization.

Spectroscopic Results

Preparation of Helix I. Helix I was produced from 5S RNA in two nucleolytic steps. Following the initial digestion of 5S RNA with RNase A to make fragment 1, and the removal of strand III in EDTA, only helix I with a long single-stranded tail remained. This tail was removed by digestion with RNase T_1, and the molecule was purified by HPLC under nondenaturing conditions. The yield for this procedure was about 50% of the theoretical limit starting from 5S RNA.

Several preparations of helix I were made differing in isotopic labeling pattern according to which deuterated bases were fed to the auxotrophic strain. The 5-deuteropyrimidine version of the molecule proved to be most useful. Because it was possible to separate the two strands and recombine them before T_1 digestion, helix I samples could also be prepared in which one strand was isotope-labeled and the other was not. The purity of helix I samples was assessed on 15% nondenaturing gels using either a TBE buffer system without magnesium (0.1 M Tris, 0.1 M boric acid, and 2.5 mM EDTA) or "5S buffer" (5 mM MgOAc, 0.1 M KCl, 50 mM Tris, and 90 mM boric acid). Purified helix I ran as a single band in both gel systems, showing that magnesium is not required for the maintenance of its structure. This conclusion was confirmed by UV melting experiments in which it was shown that helix I melts in one

cooperative transition in the absence of magnesium (data not shown).

Characterization of Samples. Since RNase T_1 cleaves at the 3′ side of single-stranded guanines, it was expected that incubation of strand II/IV complexes with moderate amounts of T_1 RNase would produce resistant molecules that include all of strand IV and the portion of strand II that runs from A108 to U120. The sequence of the strand II side of the helix I samples was determined directly, as described under Materials and Methods. It indeed does start at A108 and ends at U120. Sequencing ladders indicated that some of the molecules in the preparation included G107, but the fraction that did must have been a very small because there was no evidence for this "extra" base in the molecule's NMR spectrum.

It was anticipated that the strand IV side of helix I, which we were unable to sequence directly, might have several different 5′ termini, U(−2), U(−1), and/or U1 because 5S RNA maturation does not go to completion under overproducing conditions (Jordan et al., 1971), but that it would have a unique 3′ terminus, G10 (Kime et al., 1984). The NMR data obtained demonstrate that the majority of the molecules examined started at U(−1) and that effectively all of them terminated with C11. On the basis of integration of one and two-dimensional peaks, there is no reason to believe that any of the terminal bases was present in substoichiometric amounts.

The termination of strand IV at G10 is characteristic of fragment 1 made from *rrnB* 5S RNA, the product encoded by pKK5-1 (Kime et al., 1984). KL375, the strain from which most of the helix I samples analyzed were derived, is *recA*$^+$, however, and pKK5-1 has a high probability of undergoing homologous recombination in KL375 (B. Freeborn, unpublished observations). It is likely that the 5S RNA obtained from KL375/pKK5-1 was wild-type, and strand IV of wildtype fragment 1 terminates at C11 (Douthwaite et al., 1979; Kime et al., 1984).

Finally, there were minor chemical shift differences between the protonated and pyrimidine-d_5 samples we made, the most striking being that the H1′ of C110. There were also small differences in their phosphorus spectra indicative of minor structural differences between the two preparations which are not understood. However, the NOE connectivity patterns were identical in both samples, and there was no evidence of sequence heterogeneity at the termini.

Magnesium Effects. In a series of one- and two-dimensional experiments using blunt helix I, chemical shifts were compared in normal NMR buffer, which contains Mg^{2+}, and in a phosphate buffer (0.1 M KCl, 20 mM $KHPO_4$, pH 7, 0.2 mM EDTA), which did not. The imino proton chemical shifts differed by an average of 0.04 ppm, but U111's NH3 moved upfield by 0.17 ppm. The average aromatic and anomeric proton chemical shift change was 0.05 ppm, and these changes occurred throughout the molecule. Although a magnesium-dependent conformation change from A to A′-type helix has been reported for 5S RNA in the past (Müller et al., 1985), we are uncertain about the origin of the chemical shift changes seen here and have not pursued it further.

Exchangeable Protons. The downfield spectrum of helix I is shown in Figure 2, and its resonances are identified according to the scheme worked out for fragment 1 (Kime & Moore, 1983b). The assignments made for helix I imino proton resonances on the basis of one-dimensional NOE experiments are summarized in Table I. Some of the molecule's hydrogen-bonded exocyclic amino protons of cytidines were also assigned using one-dimensional NOE experiments and NOESY experiments (Table I).

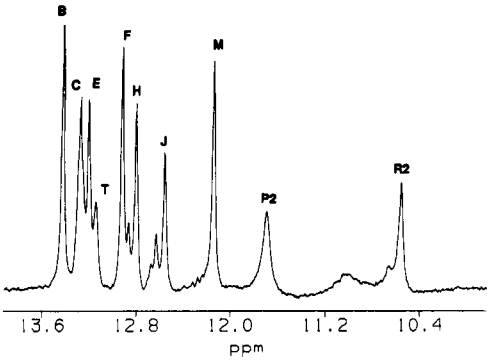

FIGURE 2: Imino proton spectrum of helix I. Data were collected at 30 °C. The buffer was 0.1 M KCl, 5 mM $MgCl_2$, pH 7, and 3 mM cacodylic acid, and water suppression was achieved by the twin pulse method (see Materials and Methods). The sweep width was 11 900 Hz, the offset was at 15 000 Hz, and 2 Hz of line broadening was applied. The RNA concentration was 3 mM.

Table I: Chemical Shifts for Exchangeable Protons at 30 °C[a]

base	N4H[b]	N4H[c]	imino	peak	fragment	Δ
U(−1)						
U1						
G2			12.55	J	12.55	0
C3	8.57	6.80				
C4	8.45	6.82				
U5			13.43	B	13.38	0.05
G6			12.15	M	12.10	0.05
G7			13.13	E	13.13	0
C8	8.40	6.61				
G9			10.56	R2	10.50	0.06
G10			*13.2*	T	13.40	0.2
C11						
A108						
A109						
C110	*8.39*	7.16				
U111			11.71	P2	11.64	0.07
G112			12.80	H	12.73	0.07
C113						
C114	8.57	6.77				
A115						
G116			12.92	F	12.88	0.04
G117			13.27	C	13.24	0.03
C118	8.37	6.76				
A119						
U120						

[a] Italicized values were determined at 21 °C. The columns labeled "N4H" and "imino" refer to helix I, and the column labeled "fragment" refers to fragment I of 5S RNA, which contains helices I, IV, and V. [b] Refers to the hydrogen-bonded proton. [c] Refers to the proton which is not hydrogen bonded.

Imino proton chemical shifts are not identical in fragment 1 and helix 1, but the differences are small, 0.04 ppm, on average, and the largest involve the G10–C110 end of helix I, where the sequence differences between helix I and 5S RNA are most important (Table I). Thus, on chemical shift grounds, there is reason to believe that helix I has a structure similar to that of the helix I portion of fragment 1 and intact 5S RNA.

Helix I resembles fragment 1 and intact 5S RNA in a number of its other spectroscopic properties. First, its U5H3 (B) to G6H1 (M) NOE is much weaker than its U5H3 (B) to G116 (F) NOE (Kime & Moore, 1983b). Second, its G9–U111 imino protons can only be observed below 25 °C (Kime & Moore, 1983c). Third, its G9–U111 imino protons give NOE's to the N1 proton of G10 but not to the N1 proton of G112 (Gewirth et al., 1987).

Blunt helix I differs from helix I in only one important respect. Its G9–U111 imino protons can be observed at 30

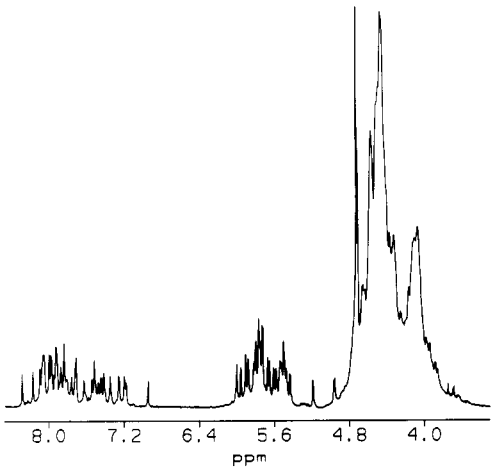

FIGURE 3: Nonexchangeable proton spectrum of helix I. The sweep width was 5000 Hz, the offset was placed on the residual HDO peak, and a small amount of presaturation was applied during the cycle delay. No line broadening was used, and the buffer conditions are as in Figure 2. The solvent was 100% D_2O, and the temperature was 30 °C.

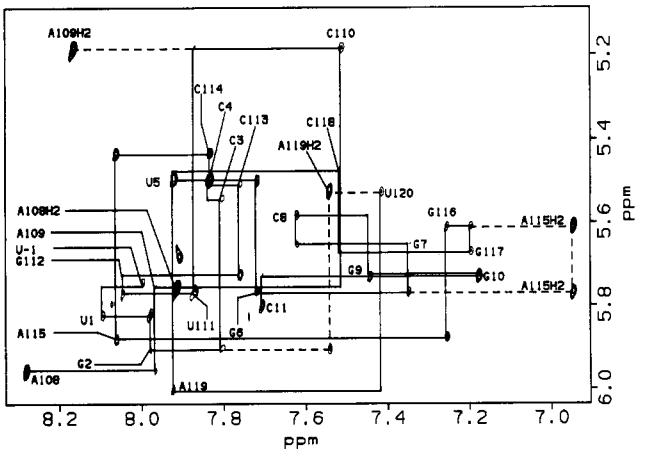

FIGURE 4: Aromatic/anomeric connectivities in helix I. The aromatic/anomeric region of the NOESY spectrum is shown for the pyrimidine-d_5 sample. The spectrum shown was processed for high resolution, and data were collected at 30 °C with a 300-ms mixing time. Labels indicate intrabase aromatic/anomeric cross peaks, and dotted lines represent connectivities to AH2 protons.

°C. Apparently the additional G–C base pair it has beyond G10–C110 is enough to stabilize that base pair. We conclude that interactions that occur in loop A at the inner end of helix I in 5S RNA must contribute little or nothing to helix I's stability. It is unlikely that helix I stacks on either helix II or V.

Nonexchangeable Protons. Figure 3 shows a one-dimensional spectrum of helix I in D_2O. Its aromatic region is reasonably well dispersed, but its ribose region is extremely crowded. Ultimately only about half of helix I's ribose resonances could be assigned. On the basis of the extensive chemical and enzymatic studies that have been done on the 5S RNA from *E. coli*, as well as phylogenetic comparisons, we had every reason to believe at the outset that helix I would be an A-form double helix over most of its length. The NOE's expected between nonexchangeable protons in an RNA double helix are well known [see Wüthrich (1986)] and indeed were identified in this molecule. Nevertheless, the assignment of this molecule's nonexchangeable proton spectrum proved to be a challenge.

Two factors complicated the assignment process. First, initial NOESY spectra taken on fully protonated helix I demonstrated that the strong pyrimidine 5,6 NOE's were obscuring many of the weaker aromatic/anomeric NOE's that were critical for assigning the spectrum. This problem was solved by turning to a helix I sample all of whose pyrimidines were deuterated at the 5-position.

Second, AU base pairs are useful reference points for beginning nonexchangeable proton assignments in a helical nucleic acid. The imino proton of an AU gives a strong NOE to its own AH2, allowing it to be identified unambiguously. In regular A-type helices, AH2 will give an NOE to the H1′ of the residue on the 3′ side of the A and to the H1′ of the residue on the 3′ side of the U [see Wüthrich (1986)]. Thus the H2 of residue A115 (6.94 ppm), which was assigned by imino proton experiments, was the obvious place to begin assigning helix I. Unfortunately, helix I is palindromic around U5–A115; the sequence from G2 to C8 is the same as the sequence from G112 to C118 except for the AU in the middle (see Figure 1). This fact was a major barrier to establishing the imino proton assignments of helix I in the first place (Kime et al., 1984). In this case it made it difficult to be sure which strand was which in the initial stages of the backbone assignment. This problem was resolved, in the end, by examining the COSY spectra given by samples whose bases were protonated on one strand and deuterated on the other. The pyrimidines belonging to each strand could be identified in this manner, ensuring that NOE connectivities had been properly traced.

Also helpful in the initial stages of the assignment process were T_1 measurements done on a helix I sample in which only adenines carried protons. AH2's generally relax much more slowly than AH8's. Three of the eight adenine resonances in helix I were tentatively identified as AH2's in this way, the resonances appearing at 6.94, 7.54, and 8.16 ppm. Three emerged as clear AH8's, the ones at 8.28, 8.06, and 7.92 ppm, while two remained ambiguous.

Figure 4 shows the aromatic/anomeric region of a NOESY spectrum obtained using a helix I sample deuterated at the 5-position of all its pyrimidines. Its sequential connectivities are traced out. The NOESY spectrum of helix I at 37 °C and 300-ms mixing time had numerous aromatic/aromatic cross peaks that support the assignment scheme shown graphically in Figure 4 and summarized in Table II. Some H3′ resonances were assigned on the basis of H3′/H6 or H3′/H8 cross peaks. In several cases, notably A108, all ribose protons could be assigned due to spin diffusion from the H8 or H6/H3′ to the H4′, H5′, and H5″. [H5′ and H5″ were not distinguished in these cases.] Similarly, at longer mixing times, spin diffusion cross peaks in the H1′/H2′ region could be assigned to H1′/H3′ correlations.

All the connectivities seen were ones typical of a right-handed A-form helix, and this connectivity pattern continued to the last bases on both strands, suggesting that the unpaired terminal bases in this molecule form a continuation of the helical stack. Even the connectivities from A109H1′ to A108H2 and A119H2 to G2H1′ are strong. The A-form connectivity pattern is maintained across the G9–U111 pair, as expected.

Backbone Torsion Angles. In an RNA double helix, all sugars are in the 3′-endo conformation, and the H1′/H2′ coupling constant is consequently too small to be observed under most circumstances. Nevertheless, both magnitude and DQF COSY spectra of helix I and five cross peaks in the H1′/H2′ region, indicating that five residues have sugars whose pucker has an appreciable 2′-endo character. In the course of the assignment process, it became clear that the sugars in question are those of residues U(−1), U1, C11, A108, and

Table II: Chemical Shifts for Nonexchangeable Protons at 30 °C[a]

base	H6/H8	H2/H5	H1′	H2′	H3′	H4′	phosphorus	$J_{H1'/H2'}$
U-1	7.99	5.84	5.74	4.52	4.34	4.61		2.5–3.5
U1	8.09	5.88	5.82	4.64	4.74			2.5–3.5
G2	7.98		5.91	4.57	4.67		−0.17	0–2
C3	7.80	5.29	5.54	4.39	4.46			0–2
C4	7.83	5.49	5.50	4.36	4.60/4.37			0–2
U5	7.92	5.43	5.50	4.59	4.60			0–2
G6	7.72		5.77	4.50	4.66			0–2
G7	7.34		5.65	4.32	4.56			0–2
C8	7.63	5.18	5.58	4.57	4.46			0–2
G9	7.44		5.73	4.72	4.28		0.12	0–2
G10	7.18		5.73	4.47	4.45		−0.59	0–2
C11	7.70	5.53	5.80	4.17	4.33	4.47	3.13	4–6
C11 T							3.21/3.27	
A108	8.28	7.91	5.95	4.96	4.66	4.44		3.5–5
A109	7.96	8.16	5.75	4.71	4.56			0–2
C110	7.51	5.53	5.18	4.47	4.31			0–2
U111	7.87	5.76	5.76	4.31	4.65		−0.52	0–2
G112	8.04		5.73	4.56	4.67			0–2
C113	7.76	5.27	5.51	4.37	4.45			0–2
C114	7.83	5.51	5.43	4.46	4.59			0–2
A115	8.06	6.94	5.88	4.57	4.76			0–2
G116	7.25		5.61	4.45	4.51		−0.4	0–2
G117	7.20		5.67	4.42	4.42			0–2
C118	7.51	5.23	5.47	4.52	4.39			0–2
A119	7.92	7.54	6.00	4.45	4.37			0–2
U120	7.41	5.29	5.53	4.07	4.34	4.24		2–3
U120T							3.21/3.27	

[a] Chemical shifts were measured as described under Materials and Methods. Two values for observed shifts are given when the assignment is uncertain. Coupling constants are given in Hz. On the basis of the nucleolytic cleavage used to produce helix I, there are terminal phosphates on C11 and U120, called "T", and a 5′-OH precedes A108.

U120. It is interesting that A109, which is opposite C11, has a 3′-endo ribose, as does A119, even though it is opposite U1, which is partially 2′-endo. Table II summarizes the ribose coupling constants derived from a DQF COSY experiment done on a pyrimidine-d_5 helix I sample.

Because there are an appreciable number of H1′/H2′ cross peaks in helix I's COSY spectra, it made sense to do a RELAY COSY experiment to identify the 3′ protons of those sugars. The RELAY COSY spectrum obtained was much simpler than an ordinary COSY spectrum in the sugar region upfield of water, and some additional of H3′ and H4′ assignments resulted.

^{31}P Spectroscopy. A phosphorus–proton heteronuclear COSY experiment was done on helix I, with the results shown in Figure 5. ^{31}P/H3′ couplings are expected to dominate such a spectrum. All the H3′ protons in the molecule have chemical shifts between 4.27 and 4.76 ppm, and all but three of its phosphorus atoms have chemical shifts between −1 and 0.2 ppm with respect to 85% phosphoric acid. The three outlying phosphorus chemical shifts are quite sensitive to buffer conditions and have been assigned to C11 and to the phosphates 3′ to U120 and C11. It is interesting to note that the phosphorus between C8 and G9 also has a slight downfield phosphorus chemical shift. By analogy with the work of Gorenstein on the NMR of G–T base pairs in a DNA helix, it is not implausible that the next most downfield cross peak at (4.67,−0.1), which we are unable to assign unambiguously, belongs to G112 (Gorenstein et al., 1988, Roongta et al., 1990).

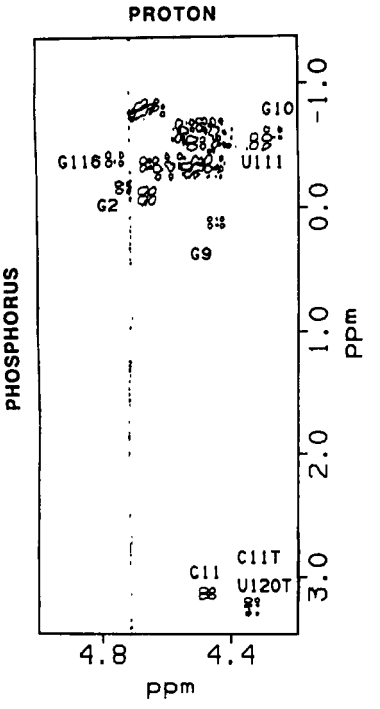

FIGURE 5: Heteronuclear phosphorus-proton COSY spectrum of helix I. The phosphorus sweep width was 1600 Hz, the frequency was 202 MHz, the offset was 8900 Hz, and the 85% phosphoric acid reference was set to 0 ppm. For protons, the sweep width was 2000 Hz and the offset was 9800 Hz. Before zero filling, the resolution in t_2 was 1 Hz/point and in t_1 was 8 Hz/point. "C11T" and "U120T" refer to the phosphates 3′ to the base indicated.

Structure Determination

Model Building. The conformation of a nucleic acid is determined if eight parameters are known for each residue: (1) its sugar pucker, (2) its glycosydic bond angle, and (3–8) its six phosphodiester backbone angles, α–ζ. We were able to measure eight interproton distances and three dihedral angles per residue for most residues in helix I. Thus the number of observations available for model construction was approximately equal to the number of structural parameters to be determined.

It would be incorrect to conclude that the data fully determine the structure, however. The distances measured were those between successive imino protons, between aromatic protons and H1′, H2′, and H3′ protons within the same res-

idue, and between H1′, H2′, and H3′ protons in one residue and the aromatic protons of neighboring residues on the 3′ side. H1′–H2′ dihedral angle estimates were obtained from the corresponding coupling constants, and α and ζ could be constrained to angles between −100° and +100° for all residues whose phosphorus resonances were clustered in the spectral region where gauche$^-$/gauche$^-$ phosphodiester phosphorus resonances normally appear (Gorenstein, 1984). Thus the data overrepresent base–base and ribose–base distances and underrepresent backbone torsion angles. For that reason, the models discussed below should depict the base-stacking geometries more accurately than they do backbone structure.

We began our model building by approximating helix I as a regular A-type helix in which unpaired residues were positioned as though they continued the helix. This model was refined by molecular dynamics using experimental NOE distances obtained by ISPA (see Materials and Methods), as well as experimental torsion angles. Five separate runs were done, and five intermediate molecular models were generated. During this phase of the model building, 2′-OH groups were prohibited from forming hydrogen bonds, as were some of the functional groups of terminal residues.

There was a tendency of the major groove to collapse during the energy minimization phase of the first stage of model building, probably due to the failure to take solvent into account explicitly. (Calculations were performed in vacuo setting the dielectric constant either to 80 or to r_{ij}, the interatomic distance, and scaling the phosphate charges by $^1/_3$). Major groove collapse was prevented by including an ad hoc repulsive distance restraint that kept the phosphates of opposite strands apart. These constraints were set so that phosphates on different strands would not come closer together than 5 Å less than the corresponding distance in regular A-type RNA, except at the ends of the molecule where phosphates were more tightly constrained. None of the final models obtained violated these constraints, the contribution of this component of the target function to the total model energy was zero. Thus, these phosphate restraints appears to have guided the models to the region of the potential energy surface that corresponds to structures with open major grooves.

In the first stage of refinement, it was important to eliminate data from consideration which were either erroneously evaluated or assigned or strongly affected by spin diffusion. Otherwise, severely distorted models resulted. As the calculations progressed, back-calculated spectra were compared to the experimental data to further eliminate cross peaks whose intensities were greatly affected by spin diffusion. In several instances, the calculated spectrum indicated that a cross peak should exist which the experimental spectra did not support. In such cases an "un-NOE" restraint was added to keep the protons in question apart.

In the second stage of the model building, each first stage models was subjected to full relaxation matrix refinement using the totality of the experimental data available (see Materials and Methods). In the first calculations of this kind we did, a standard residual factor (R factor) was used as the target function (see eq 1 with $n = 1$ and $\omega_i = 1$ in Materials and Methods). The models that resulted fit the strong cross peaks quite well but represented the weak ones poorly, an unsurprising result given that cross peaks varied 1000-fold in intensity. For that reason, we switched to a target function that involves differences between the sixth root of intensities, which effectively reduced the dynamic range to be considered to less than an order of magnitude and led to a more uniform fitting of the data (see eq 1 with $n = ^1/_6$). The five final models, on

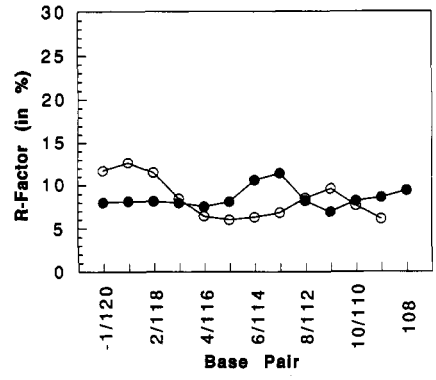

FIGURE 6: Plot of $^1/_6$-weighted R factor against sequence position. The contribution to the total R factor made by each residue in the sequence is plotted by position. Open circles represent bases (−1)–11 and closed circles bases 108–120.

Table III: R Factors and Energies[a]

	$R_{1/6\text{-weighted}}$	R_{standard}	E_{NOE} (kcal)
A-form helix	0.13	0.57	132
distance refined from A	0.12	0.56	48
relaxation refined from A	0.086	0.35	
B-form helix	0.26	1.17	1620
distance refined from B	0.12	0.58	75
relaxation refined from B	0.088	0.38	

[a] R factors, as defined in eq 3 of Materials and Methods, obtained at different stages in the model building are summarized. R factors from several runs are averaged. A- or B-form helix refers to the energy-minimized starting structure, and "distance refined" refers to the ISPA-refined intermediate models while "relaxation refined" refers to final models produced by the full matrix relaxation refinement.

which the rest of our analysis depends, were all calculated using this "$^1/_6$-weighted" target function, but both standard and $^1/_6$-weighted R factors are reported for these models in Table III. Figure 6 shows the $^1/_6$-weighted R factors calculated by base. The largest values are at the ends, where the conformation may be poorly determined and where the molecule may be flexible, and at bases where cross peak overlap made peak integration difficult namely C3, C4, C113, and C114. The scale factor K_R given to the NMR data in the second stage of refinement was adjusted empirically so that the final model fit the data without distortion of the covalent bonds. The average of the RMS deviations from ideal values for bond lengths was 0.011 Å, for bond angles 3.4°, and 0.31° for base planarity.

In Figure 7 we present a series of scatterplots that compare the back-calculated intensities with observed intensities. All of the intensities have been normalized by $1/I^{(1/6)}$ to make them roughly equal to the interproton distances to which they correspond. The deviation is the difference between calculated and observed "distances" and is plotted against the observed "distance" derived from the NOE peak intensities. If the model fit the data perfectly, all points would lie on the line $x = 0$. The experimental data were compared to two possible starting structures, the ideal A- and B-form helices. It is clear from Figure 7 that helix I is much closer to an A-form helix and that refinement decreased the deviations between experimental and back-calculated data significantly. The final $^1/_6$-weighted R factor is 8.6%. It will be noted that a disproportionate fraction of the large distance points have negative deviations. This may reflect difficulties in integrating small NOE cross peaks, or it could be due to non-NOE losses of magnetization.

As a further test of this model building method, a calculation was started from the ideal B-form geometry. The model that

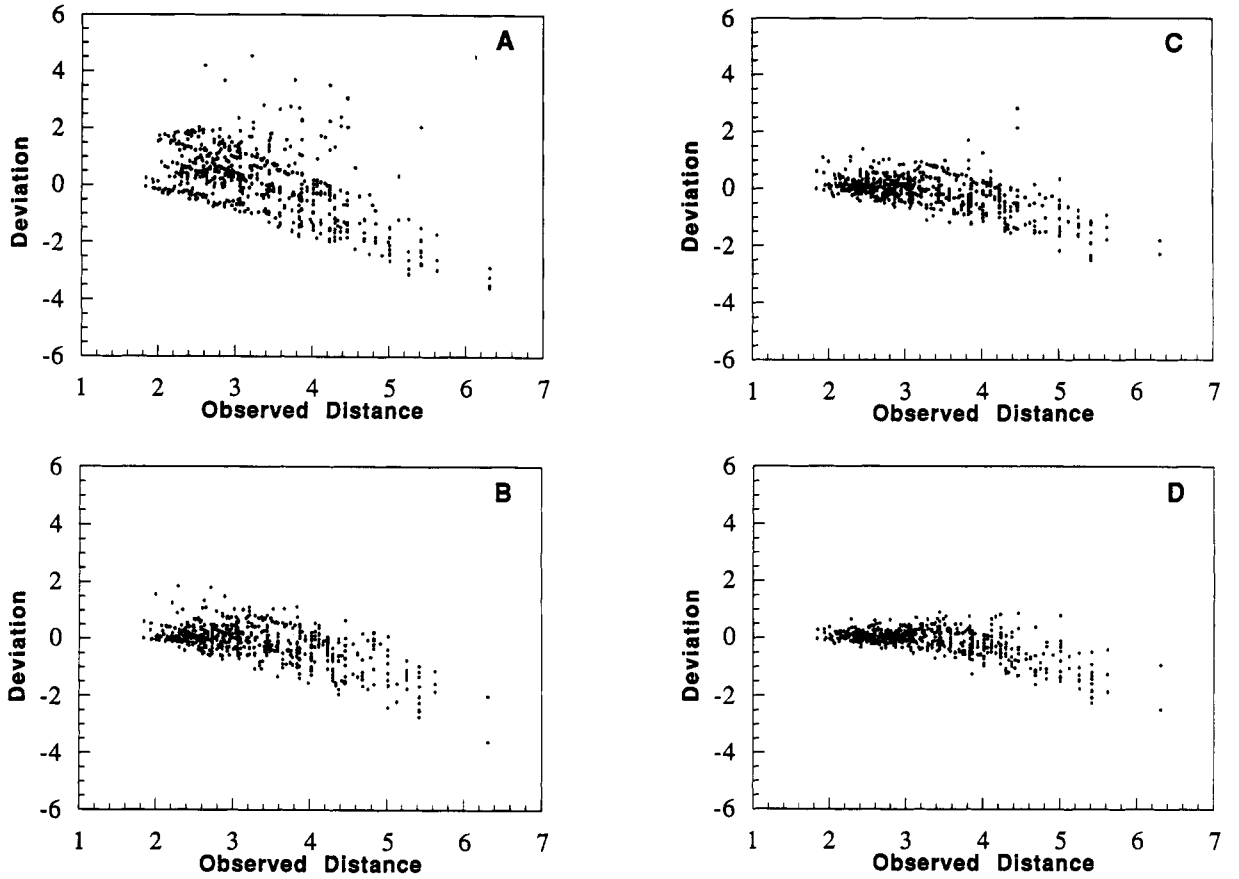

FIGURE 7: Observed interproton "distance" versus "distance deviation" for four different models of helix I. Distance have been calculated from cross peak intensities using $6.32/(I^{1/6})$ and correspond roughly to actual interproton distances in angstroms where I is an NOE cross peak intensity. "Deviations" are calculated as model distances minus observed distance, and all four mixing times are included. (A) B-form helix; (B) A-form helix; (C) intermediate model after ISPA refinement; and (D) final model after full matrix relaxation.

resulted resembles those obtained by starting from ideal A-form helix, but its energy is slightly higher. For reasons we don't understand, even prolonged ISPA calculations produced a helix I-like intermediate model that is still slightly higher in energy than expected. When this intermediate model was subjected to full matrix refinement, again the energy was slightly higher although the model that resulted was very similar to those obtained starting from the A-form helix. At each stage, the R factors for the B-helix calculations are larger than those for models which originated from A-form helix. They are summarized in Table III.

Structural Conclusions. By all measures, helix I is unquestionably an A-form helix. In Figure 4 the conventional "NOE walk" for each strand is traced from one end to the other including the unpaired terminal bases and the G–U pair. The connectivity pattern is exactly what is expected for an A-form helix where the sequential aromatic to H1′ cross peaks are much smaller than the corresponding interbase cross peaks. Similar connectivity pathways can be traced for both the H2′ and H3′ protons (data not shown). None of the chemical shifts of the nonexchangeable protons in Table II falls outside of the range typical for each proton type (Table II). There is very little chemical shift dispersion in the phosphorus spectrum with the exception of the terminal residues, as shown in Figure 5. Thus, all of the nonterminal phosphorus atoms are in similar environments. Furthermore, the scatterplots confirm that helix I is more like A- than B-type helix.

Figure 8 shows a stereoview of a superposition of five final structures oriented so that the G–U is near the bottom front. These structures superimpose well through their central regions and diverge near their ends. In addition, the central bases have conformations quite close to the ones they had in the starting model. The pairwise RMS differences between models is slightly over 1 Å overall but is twice that for the terminal bases. It can be seen that the width of the major groove depends critically on the positions of the terminal bases' backbone. In order to determine the conformation here with more confidence, we would have to be able to measure more backbone torsion angles than we did and to include both water and more realistic electrostatic potentials in our calculations.

The G–U base pair is accommodated in the helix I in a way that accentuates stacking between purines as Figure 9 illustrates. In all of our models, G9 and G112 form an intrastrand stack and G9–U111 stacks neatly on G10–C110, while stacking between C8 and G9 is nonexistent. This pattern of cross-chain purine stacking is seen to a much lesser extent at the U5–U6 step, the only other 5′ pyrimidine/3′ purine step in helix I, where the six-membered rings of A115 and G6 just barely overlap. Imino NOE's were seen between neighboring imino protons in helix I in all cases except between G112 and G9 or U111. That failure is due to the translation of G9 toward G112. That the distance between C8 and G9 is indeed larger than average is supported by the observation that the sequential C8H1′–G9H8 and C8H3′–G9H8 cross peaks are very small. U111 and C8 both protrude toward the helical axis into the major groove.

Analysis of the helical parameters using CURVES (Lavery & Sklenar, 1988) indicates that in many respects the final model resembles the initial A-form model following energy minimization, but trends seen at the G-U base pair in the starting model are accentuated in the final models. Comparison of the helical twist values for the initial and final models shows a low twist angle for the G112–U111/C8–G9 step. The low twist at the G9–U111 base pair is deceptive,

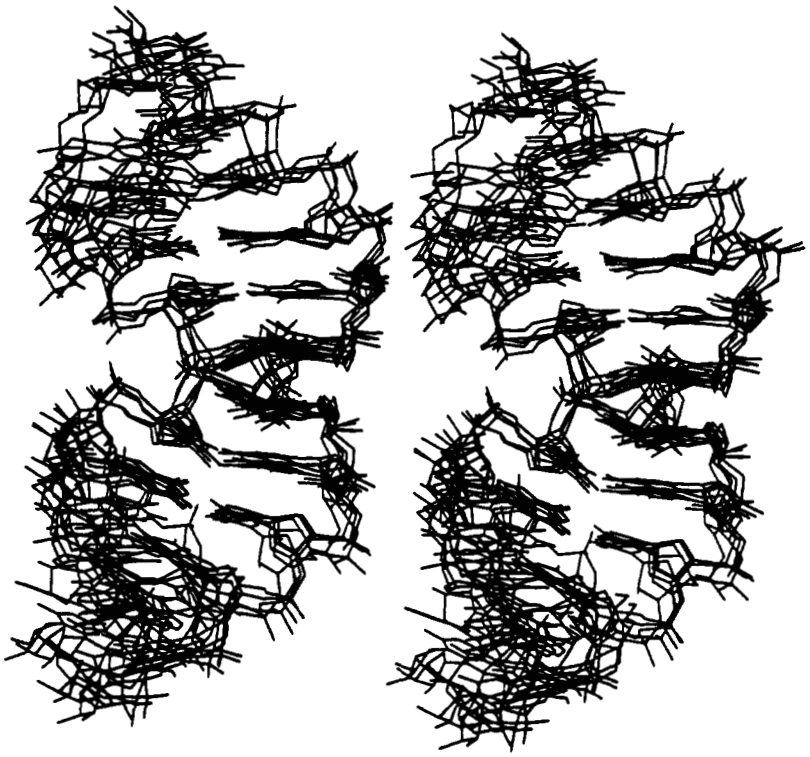

FIGURE 8: Stereopairs of five final structures derived from separate ISPA-refined models. The non-hydrogen atoms of the central base pairs, C3/G117–C8/G112, were superimposed using INSIGHT (BioSym Technologies, Inc.).

Table IV: Helical Parameters for the Starting Model and the Final Models at G9–U111[a]

	twist (deg)		α (deg)		ζ (deg)		X displacement (Å)	
	helix I	initial	helix I	initial	helix I	initial	helix I	initial
C8							−2.4	−2.6
p	26	22	−72	−70	−77	−71		
G9							−3.2	−4.0
p	41	42	−60	−73	−57	−64		
G10							−1.8	−2.6
C110							−2.1	−2.2
p	42	50	−89	−76	−68	−61		
U111							−1.1	−1.6
p	22	23	−70	−74	−88	−63		
G112							−3.3	−3.2
average	35	33	−72	−73	−72	−64	−3.4	−3.1

[a] "Helix I" refers to the final, full relaxation matrix refined model while "initial" refers to the energy-minimized A-form starting model. Twist, α, ζ, and X disp are defined in Lavery and Sklenar (1988). The helical parameters for all bases except 108 were averaged to obtain "average" values.

however. C8 and G9 are definitely unstacked and appear to be overwound, but the large separation between these two bases is due to translation, not twist. The displacement of each base parallel to the base pair axis away from the helix axis is called X displacement and is summarized in Table IV. Again, the final structures follow the pattern of the initial starting structure. C8 is displaced slightly into the major groove as observed in model building, while G112 is pushed slightly toward the exterior and U111 is more than 2 Å closer to the center of the helix. Table IV also shows that our average final model's backbone torsional angles follow the trend of the starting model but that angle ζ is exaggerated at bases at the C8–G112/G9–U111 step and is smaller than average at G9 and G112. This pattern of compensating changes, such as small twist preceded by a large one, is often seen in nucleic acids. While it was not possible to assign all of the phosphorus resonances, larger ζ angles seem to correlate with the downfield phosphorus chemical shifts in the same fashion as in DNA oligomers. In summary, the stacking pattern of the bases in the G–U pair is governed by the wobble geometry, but in helix I the values of the twist, displacement from the axis, and torsional angle ζ are more extreme than the corresponding values in the starting model.

Not surprisingly, the structures of the unpaired ends of helix I are not well determined. The fact that we see some C2′-endo character in terminal residues is probably indicative of a dynamic equilibrium between a C2′- and C3′-endo sugar pucker. The coupling constants of 2–4 Hz mean that 20–40% of the time the riboses of the termini are in the C2′-endo conformation (Altona & Sundaralingam, 1973). The R factor for bases U1 and U(−1) is also larger than average. Thus, a detailed structural analysis of the unpaired terminal bases would not be justified. It should be noted, however, that many of our final models suggest that bases C11, A109, and A108 can form an interdigitated structure where C11 hydrogen-bonds weakly to both A109 and A108. This "staggered" structure leads to a more extended helix, and relatively small changes in the conformations of the terminal bases such as these produce marked alterations in the shape of the major and minor grooves.

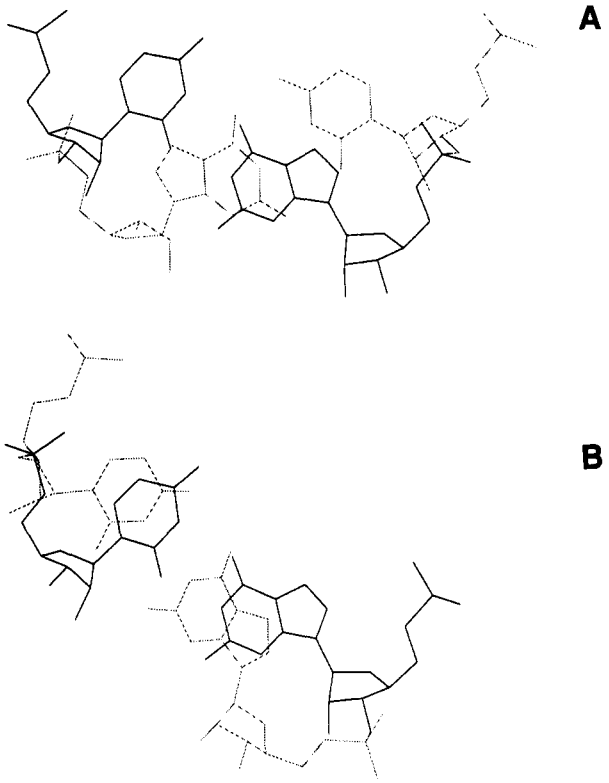

FIGURE 9: Local environment of G9–U111. Views of G9–U111 base pair (solid lines) down the helical axis: (A) G9–U111 and C8–G112 (dotted lines) and (B) G9–U111 and G10–C110 (dotted lines).

Discussion

Rabbinovich et al. (1988) have solved the crystal structure of a symmetric A-form DNA helix having central G–T mismatches. They also see that the apparently large twist and slide associated with 5′ pyrimidine/3′ purine steps is accenuated by G–T wobble base pairs. This twist is a consequence of the rotation of the G C1′–U C1′ vector necessitated by wobble pairing. Thus, helix I's G–U resembles G–T's in A-form DNA's.

On the basis of their $tRNA^{Phe}$ crystal structure, Mizuno and Sundaralingam (1978) predicted that base pairs to the 5′ side of the U of a G–U pair and the 3′ side of the G will stack favorably with the G–U, while base pairs on the opposite side should stack very poorly on the G–U. This is exactly what is seen in helix I where the C8–G112 does not stack with G9–U111, but G10–C110 stacks very well with the G–U pair. Van Knippenberg et al. (1990) have pointed out that most of the G–U pairs found at the ends of helices are oriented so that they will stack on their helices. Helix I is exceptional in this regard, since the G–U is stacked with the last base pair before the complicated junction region where helices I, V, and II come together, and thus G–U stabilizes the junction region at the expense of the stem.

Helix I is significantly larger than the other RNAs whose structures have been determined by two-dimensional NMR methods to date [e.g., Varani et al. (1991)]. As other work shows, when all of the cross peaks of an RNA's COSY and NOESY spectrum are well-resolved and nearly all of coupling constants can be measured, RNA structures can be overdetermined by NMR data, and high-resolution models may be built with confidence. This was not the case here, because the relatively large size of helix I made it impossible to obtain complete ribose assignments. It was possible to build a less precise "medium-resolution" model nevertheless. Because NMR relies on short distance measurements to determine molecular structures, it is not an accurate method for determining the conformation of long linear molecules such as this one (van de Ven & Hilbers, 1988). However, the geometries between neighboring bases can be determined with some confidence by this means, and local departures from regular helical geometry can be detected, as we have shown.

Acknowledgments

We thank Dr. K. Brooks Low of Yale University for creating strain KL375 for us. We are grateful to Betty Freeborn who produced and purified large quantities of 5S RNA and fragment I. We wish to acknowledge the assistance of Gregory Kellogg and Dr. Robert Rycyna who worked out the methods used to make deuterated nucleosides. Mr. Kellogg gave us the 2,8-dideuteroadenosine and introduced us to the heteronuclear COSY experiment we employed. We are particularly grateful to Joseph Kim, who suggested that it might be possible to make helix I, and to Marc Jacobs, who pioneered its production. Dr. Paul Jones wrote the interactive integration routine we used. We also thank David Schweisguth and Katherine Long for calculating some of the initial models we used. Finally, we are grateful for the assistance of Mr. Peter Demou of the Yale Chemical Instrumentation Center for his help in the collection of NMR data.

References

Altona, C., & Sundaralingam (1973) *J. Am. Chem. Soc. 95*, 2333–2344.

Aue, W. P., Bartholdi, E., & Ernst, R. R. (1976) *J. Chem. Phys. 64*, 2229–2246.

Banks, K. M., Hare, D. R., & Reid, B. R. (1989) *Biochemistry 28*, 6996–7010.

Bax, A., & Lerner, L. (1988) *J. Magn. Reson. 79*, 429–438.

Bruccoleri, R. E., & Karplus, M. (1986) *J. Comput. Chem. 7*, 175–185.

Brünger, A. T. (1990) Refinement of three-dimensional structures of proteins and nucleic acids, in *Topics in Molecular Biology* (Goodfellow, J., Ed.) pp 137–178, Macmillan, London.

Brünger, A. T. (1991) XPLOR *Version 2.2 User Manual*, Yale University, New Haven, CT.

Calf, G. E., & Garnett, J. L. (1964) *J. Phys. Chem. 68*, 3887.

Douthwaite, S., Garrett, R. A., Wagner, R., & Feunteun, J. (1979) *Nucleic Acids Res. 6*, 2453–2470.

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

Gorenstein, D. G., Ed. (1984) *Phosphorus-31 NMR. Principles and Applications*, Academic Press, Inc., New York.

Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V., & Jones, C. R. (1988) *Biochemistry 27*, 7223–7237.

Heller, S. R. (1968) *Biochem. Biophys. Res. Commun. 32*, 998–1001.

Holak, T. A., Scarsdale, J. N., & Prestegard, J. H. (1987) *J. Magn. Reson. 74*, 546–549.

James, T. L., Gochin, M., Kerwood, D. J., Schmitz, U., & Thomas, P. D. (1991) Refinement of Three-Dimensional Protein and DNA Structures in Solution from NMR Data, in *Computational Aspects of the Study of Biological Macromolecules by NMR* (Hoch, J. C., Ed.), Plenum Press, New York.

Jordan, B. R., Forget, B. G., & Monier, R. (1971) *J. Mol. Biol. 55*, 407–421.

Kean, J. M., & Draper, D. E. (1985) *Biochemistry 24*, 5052–5061.

Kime, M. J., & Moore, P. B. (1983a) *FEBS Lett. 153*, 199-203.
Kime, M. J., & Moore, P. B. (1983b) *Biochemistry 22*, 2615-2622.
Kime, M. J., & Moore, P. B. (1983c) *Biochemistry 22*, 2622-2629.
Kime, M. J., Gewirth, D. T., & Moore, P. B. (1984) *Biochemistry 23*, 3559-3568.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63-91.
Maeda, M., & Kawazoe, Y. (1975) *Tetrahedron Lett. 16*, 1643.
Marshall, A. G., & Wu, J. (1989) *Biol. Magn. Reson. 9*, 55-118.
Mizuno, H., & Sundaralingam, M. (1978) *Nucleic Acids Res. 5*, 4451-4461.
Moore, P. B., Abo, S., Freeborn, B., Gewirth, D. T., Leontis, N. B., & Sun, G. (1988) *Methods Enzymol. 164*, 158-174.
Müller, J. J., Misselwitz, R., Zirwer, D., Damaschun, G., & Welfle, H. (1985) *Eur. J. Biochem. 148*, 89-95.
Nagayama, K., Jumar, A., Wüthrich, K., & Ernst, R. R. (1980) *J. Magn. Reson. 40*, 321-334.
Nijkamp, H. J. J., & De Haan, P. G. (1967) *Biochim. Biophys. Acta 145*, 31.
Nilges, M., Habazettl, J., Brünger, A. T., & Holak, T. A. (1991) *J. Mol. Biol. 219*, 499-510.
Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591-616.
Otting, G., Widner, H., Wagner, G., & Wüthrich, K. (1986) *J. Magn. Reson. 66*, 187-193.
Rabbinovich, D., Haran, T., Eisenstein, M., & Shakked, Z. (1988) *J. Mol. Biol. 200*, 151-161.
Rabi, J. A., & Fox, J. J. (1973) *J. Am. Chem. Soc. 95*, 1628-1632.
Roongta, V. A., Jones, C. R., & Gorenstein, D. G. (1990) *Biochemistry 29*, 5245-5258.
Shaka, A. J., & Freeman, R. (1983) *J. Magn. Reson. 51*, 169-173.
Sklenar, V., & Bax, A. (1987) *J. Am. Chem. Soc. 109*, 7525-7526.
Sklenar, V., Miyashiro, H., Zon, G., Miles, H. T., & Bax, A. (1986) *FEBS Lett. 208*, 94-98.
States, D. J., Haberkorn, R. A., & Reuben, D. J. (1982) *J. Magn. Reson. 48*, 286.
Stout, G. H., & Jensen, L. H. (1989) *X-ray Structure Determination*, John Wiley and Sons, New York.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329-340.
van de Ven, F. J. M., & Hilbers, C. W. (1988) *Eur. J. Biochem. 178*, 1-38.
van Knippenberg, P. H., Formenoy, L. J., & Heus, H. A. (1990) *Biochim. Biophys. Acta* 1050, 14-17.
Varani, G., Cheong, J.-C., & Tinoco, I. (1991) *Biochemistry 30*, 3280-3289.
Wagner, G. (1983) *J. Magn. Reson. 55*, 151-156.
Webster, K. R., & Spicer, E. K. (1991) *J. Biol. Chem.* (in press).
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley and Sons, New York.
Yip, P., & Case, D. A. (1989) *J. Magn. Reson. 83*, 643-8.
Zagorski, M. (1990) *J. Magn. Reson. 80*, 400-405.
Zhang, P., Rycyna, R., & Moore, P. B. (1989) *Nucleic Acids Res. 17*, 7295-7301.

A Fully Active Variant of Dihydrofolate Reductase with a Circularly Permuted Sequence†

Anne Buchwalder,‡ Halina Szadkowski, and Kasper Kirschner*
Abteilung Biophysikalische Chemie, Biozentrum der Universität, Klingelbergstrasse 70, CH-4056 Basel, Switzerland
Received August 15, 1991; Revised Manuscript Received November 5, 1991

ABSTRACT: The amino acid sequence of mouse dihydrofolate reductase was permuted circularly at the level of the gene. By transposing the 3′-terminal half of the coding sequence to its 5′ terminus, the naturally adjacent amino and carboxyl termini of the native protein were fused, and one of the flexible peptide loops at the protein surface was cleaved. The steady-state kinetic constants, the dissociation constants of folate analogues, and the degree of activation by both mercurials and salt as well as the resistance toward digestion by trypsin were almost indistinguishable from those of a recombinant wild-type protein. Judged by these criteria, the circularly permuted variant has the same active site and overall structure as the wild-type enzyme. The only significant difference was the lower stability toward guanidinium chloride and the lower solubility of the circularly permuted variant. This behavior may be due to moving a mononucleotide binding fold from the interior of the sequence to the carboxyl terminus. Thus, dihydrofolate reductase requires neither the natural termini nor the cleaved loop for stability, for the conformational changes that accompany catalysis as well as the binding of inhibitors, and for the folding process.

The specific exchange of single amino acid residues by recombinant DNA technology is a powerful approach to identify residues that are crucial for the folding mechanism and stability of proteins (Matthews, 1991a,b). Circular permutation of the protein sequence is a more radical approach (Goldenberg & Creighton, 1983). It consists of connecting the natural amino and carboxyl termini covalently with a new peptide linker and cleaving the circular molecule at one of the surface loops to generate new termini. A given sequence can be permuted without compromising the interaction between secondary structural elements in the protein core in as many

† This work was supported by Grant 3100-25711.88 from the Swiss National Science Foundation.
* To whom correspondence should be addressed.
‡ Present address: Beckman Research Institute of the City of Hope, Division of Immunology, 1450 E. Duarte Rd., Duarte, CA 91010.

0006-2960/92/0431-1621$03.00/0