

in the conventional buffer system both DNA and RNA were degraded (lane c).

In a study to see if the modified buffer protects the DNA against a challenge of purified DNase, the DNA and RNA pellet dissolved in this buffer was treated with 0.4 units of DNase (Sigma) for every 30  $\mu$ l of sample and incubated at 37°C for 30 min. Similar treatment was given to a pellet dissolved in TE. Electrophoretic separation (Figure 2) on agarose gel revealed that DNA remained intact (lane b) and thus protected in the modified buffer, but suffered degradation (lane c) in TE. A possible explanation of inhibition of DNase is removal of otherwise essential free magnesium ions from the reaction by an equilibrium ion exchange phenomenon<sup>7</sup> resulting in formation of a magnesium citrate complex.

It has also been observed in another set of experiments that (i) DNA remains unaffected in the modified SSC buffer even when RNase is added to a final concentration of 200  $\mu$ g/ml, (ii) decreasing the concentration of sodium citrate to 0.04 M does not always protect the DNA against the contaminating DNase.

The DNA in the modified SSC buffer was later extracted with phenol and chloroform and precipitated with ethanol. The pellet dissolved in TE was successfully nick-translated using the BRL kit, which confirmed that the DNA is in good condition. The DNA in the modified SSC buffer was preserved at room temperature for one month without any visible change in electrophoretic pattern.

Our method is efficient, simple (as heating and cooling is dispensed with) and time-saving. It is a promising method for overcoming the problems associated with contaminating DNase in commercial preparations of pancreatic RNase and enables safe isolation and preservation of DNA.

1. Maniatis, T., Fritsch, E. F. and Sambrook, J., *Molecular Cloning*, CSH Laboratory, USA, 1982, pp. 451.
2. Blaschek, R. P. and Valacik, M. A., *Appl. Environ. Microbiol.*, 1984, **48**, 178.
3. Silhavy, T. J., Berman, M. L. and Enquist, L. W., *Experiments with Gene Fusion*, CSH Laboratory, USA, 1984, pp. 199.
4. Mc Carmic, J. J., Larson, L. J. and Masher, V. M., *Biochem. Biophys. Acta*, 1974, **349**, 145.
5. Dagmar, J., M/S Boehringer Mannheim GmbH, West Germany, Personal communication No 20/pb, Dt. 10.10.88.
6. Birnboim, H. C. and Doly, J., *Nucleic Acid Res.*, 1979, **7**, 1513.
7. Schubert, J. and Lindenbom, J., *J. Am. Chem. Soc.*, 1952, **74**, 3529.

## NMR recipe for sequencing short DNA fragments

R. Ajay Kumar, K. Majumdar and R. V. Hosur

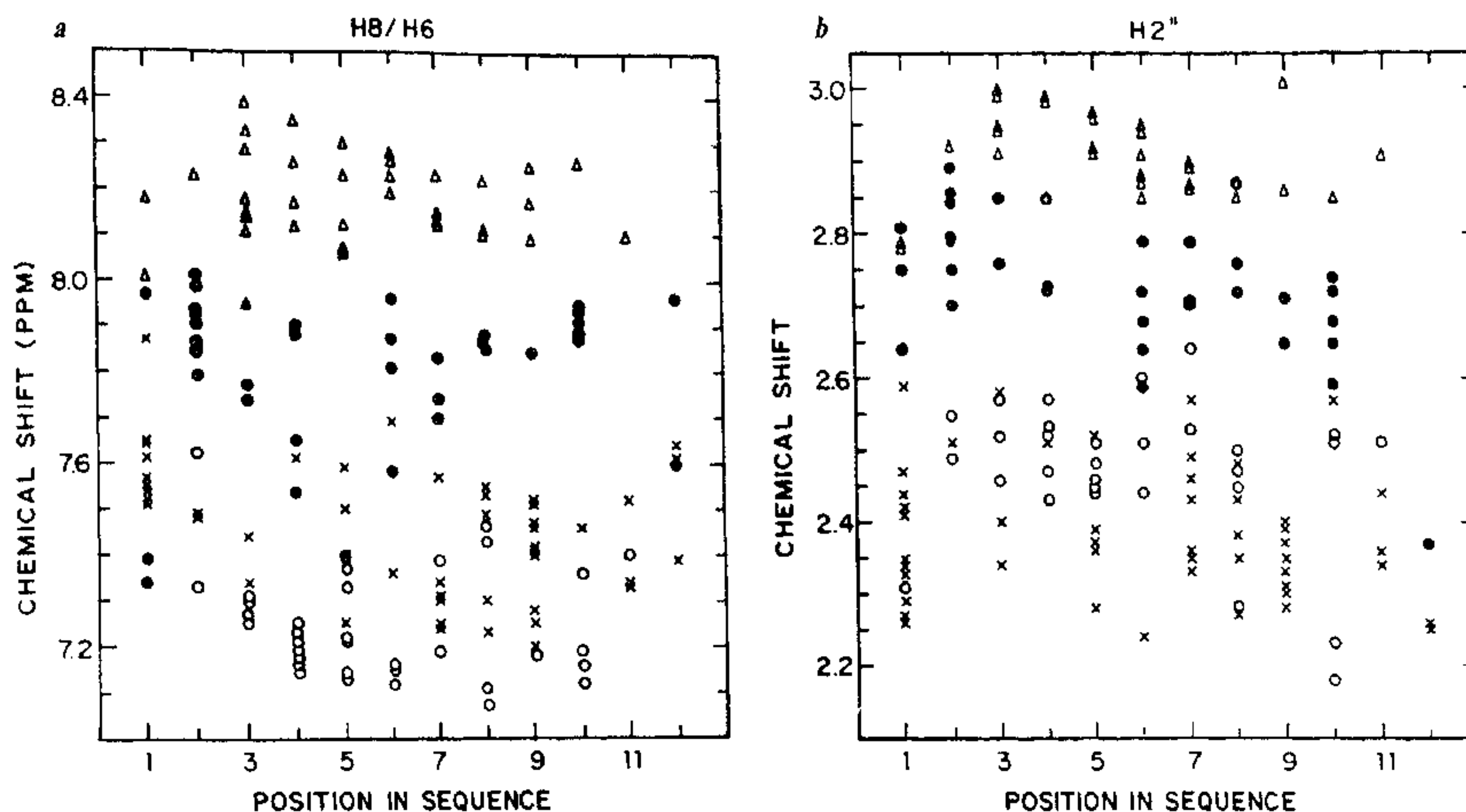
Chemical Physics Group, Tata Institute of Fundamental Research, Bombay 400 005, India

A new recipe has been described for determination of the base sequence in short DNA segments by two-dimensional NMR spectroscopy. The recipe is based on (i) A,T,G,C-distinguishing criteria obtained by analysis of chemical shifts of the non-exchangeable protons and (ii) cross-peak patterns in two-dimensional COSY and NOESY spectra. The base H8 and sugar H2" chemical shifts have been found to be characteristically dependent on the base type to which they belong and the patterns of H8-H2" cross-peaks in NOESY allow determination of sequence of bases in DNA segments.

We describe here a novel application of two-dimensional NMR spectroscopy<sup>1,2</sup>, namely sequencing of DNA segments. The 2D-NMR methods of significance here are J-correlated spectroscopy (COSY)<sup>3,4</sup> and nuclear Overhauser effect correlated spectroscopy (NOESY)<sup>5,6</sup>. The 'cross-peaks' in COSY display J-coupling (through bond) correlations while those in NOESY display dipolar coupling (through space) correlations and carry proximity information (interproton distance less than 5 Å).

There are basically two steps in the sequencing procedure: (i) nucleotide units A, T, G and C must be distinguished in the two-dimensional spectrum, and (ii) adjacent nucleotide units in the sequence must be identified.

The nucleotide units C and T are readily distinguishable from the 2D COSY and NOESY spectra via the characteristic CH6-CH5 and TH6-TCH<sub>3</sub> cross-peaks, which appear in distinct and identical regions in both the spectra. Further, due to the relatively large CH6-CH5 coupling constant ( $\approx$  7 Hz) the cross-peaks originating from CH6 protons often appear as doublets in the NOESY spectrum. As regards A and G, although these units do not display such characteristic features in the 2D spectra, we observed from a statistical analysis of published chemical shift data on various right-handed double-helical DNA segments<sup>7-22</sup>, that they can be distinguished on the basis of chemical shifts of base H8 and sugar H2" protons (Figure 1). In Figure 1,a, the chemical shifts of H8/H6 protons of the bases A, G, T and C in different DNA segments have been plotted against their positions along the sequence; Figure 1,b shows a similar plot for the H2" protons. In both these cases, the A and G nucleotides have markedly distinct chemical shifts, irrespective of position along the sequence or nature of the sequence. A few terminal units deviate from this rule but they can be readily identified (see below).



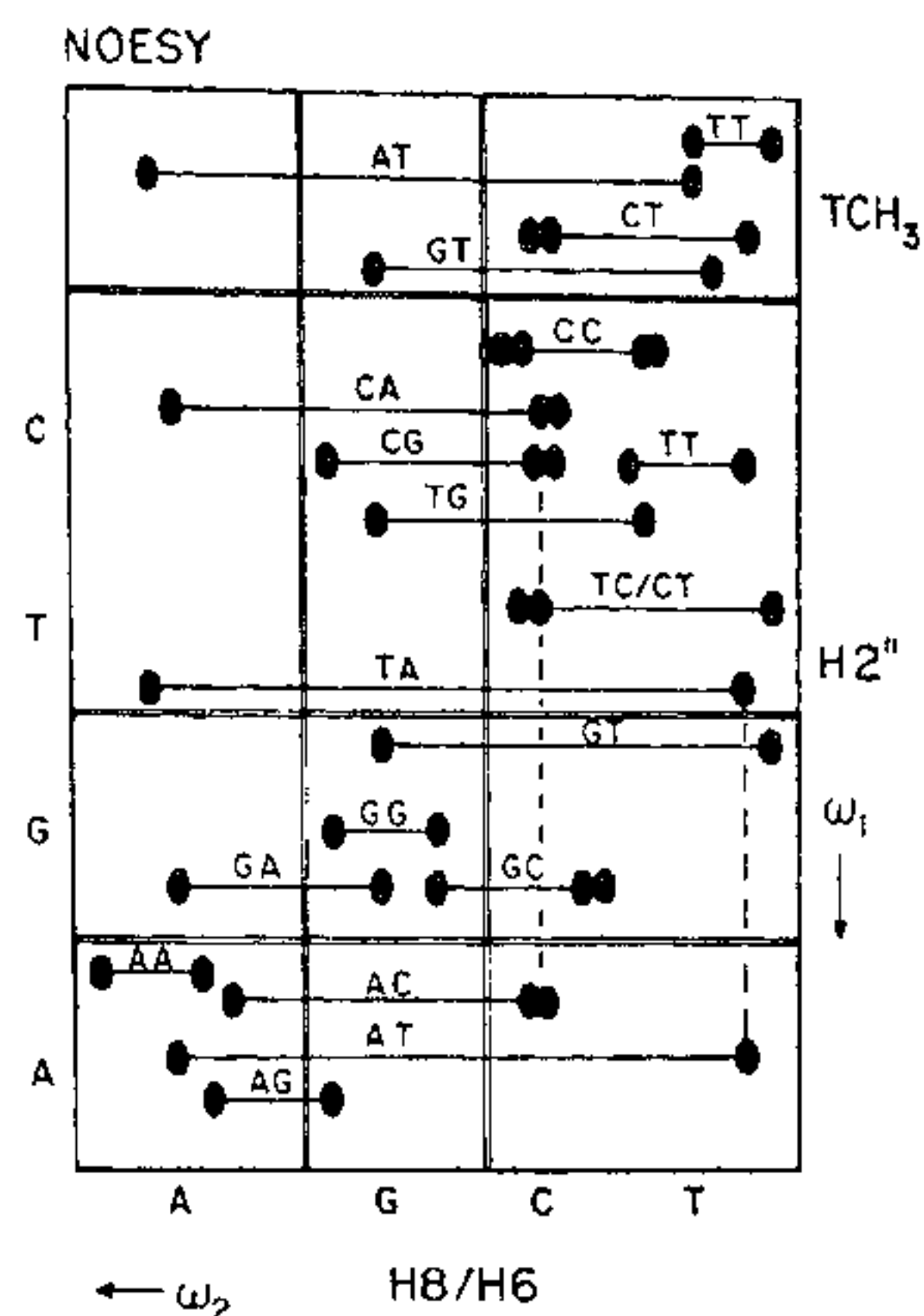
**Figure 1.** *a*, Chemical shifts of H8 or H6 protons of bases A, G, T, C, plotted against the position of the respective nucleotide unit along the sequence in various DNA segments. Secondary structure studies of these DNA segments have been published during the last five years<sup>7-22</sup>. The selected DNA segments varied from 6 to 19 nucleotides in length, and the spectra from which the chemical shift data base was prepared, were recorded under roughly similar experimental conditions of pH ( $\approx 7.0$ ), salt concentration (0.01–0.1 M), temperature (20–35°C, i.e. below melting temperature), etc. All the DNA segments had duplex structure but exhibited localized sequence-specific variations. All of them were right-handed double helices with overall topology ranging from B DNA to midway between A and B DNA structures. The bases are discriminated by different symbols: A ( $\Delta$ ), G ( $\bullet$ ), C ( $\times$ ) and T ( $\circ$ ). *b*, Chemical shifts of H2'' protons, plotted in a similar fashion as in *a*.

Having thus obtained a discriminating criteria for A, G, T and C, we have devised a strategy based on 2D-NOESY spectra, for identifying adjacent nucleotide units in a given molecule. The useful spectral region covers the chemical shifts of H8/H6 protons along the  $\omega_2$  axis and those of H2'/H2'' and CH<sub>3</sub> protons along the  $\omega_1$  axis. However, peaks originating from H2' protons can be clearly identified<sup>17</sup>, and can be excluded from the analysis. We shall label the above region as the RISD (region of interest for sequence determination). When a NOESY spectrum is recorded with a sufficiently long mixing time ( $\approx 300$ –400 ms), the RISD will contain the cross-peaks (H8/H6)<sub>*i*</sub>  $\rightarrow$  (H2'')<sub>*i*</sub>, (H8/H6)<sub>*i*</sub>  $\rightarrow$  (H2'')<sub>*i-1*</sub>, (TCH<sub>3</sub>)<sub>*i*</sub>  $\rightarrow$  (H6)<sub>*i*</sub> and (TCH<sub>3</sub>)<sub>*i*</sub>  $\rightarrow$  (H8/H6)<sub>*i-1*</sub> (*i* increases from the 5'-end to the 3'-end). From this, the 5'-terminal (H8/H6) proton and the 3'-terminal H2'' proton can be readily identified, since each of these protons produces only one H8/H6–H2'' cross-peak. Every other H2'' proton generates two such peaks, namely (H2'')<sub>*i*</sub>  $\rightarrow$  (H8/H6)<sub>*i*</sub> and (H2'')<sub>*i*</sub>  $\rightarrow$  (H8/H6)<sub>*i+1*</sub> and thus allows identification of adjacent nucleotide units.

Figure 2 shows the approximate positions of H8/H6  $\rightarrow$  H2'' and H8/H6  $\rightarrow$  TCH<sub>3</sub> cross-peaks in the RISD for all possible dinucleotide segments. Clearly, in any given RISD all dinucleotide stretches can be readily identified; CT/TC discrimination which is not possible

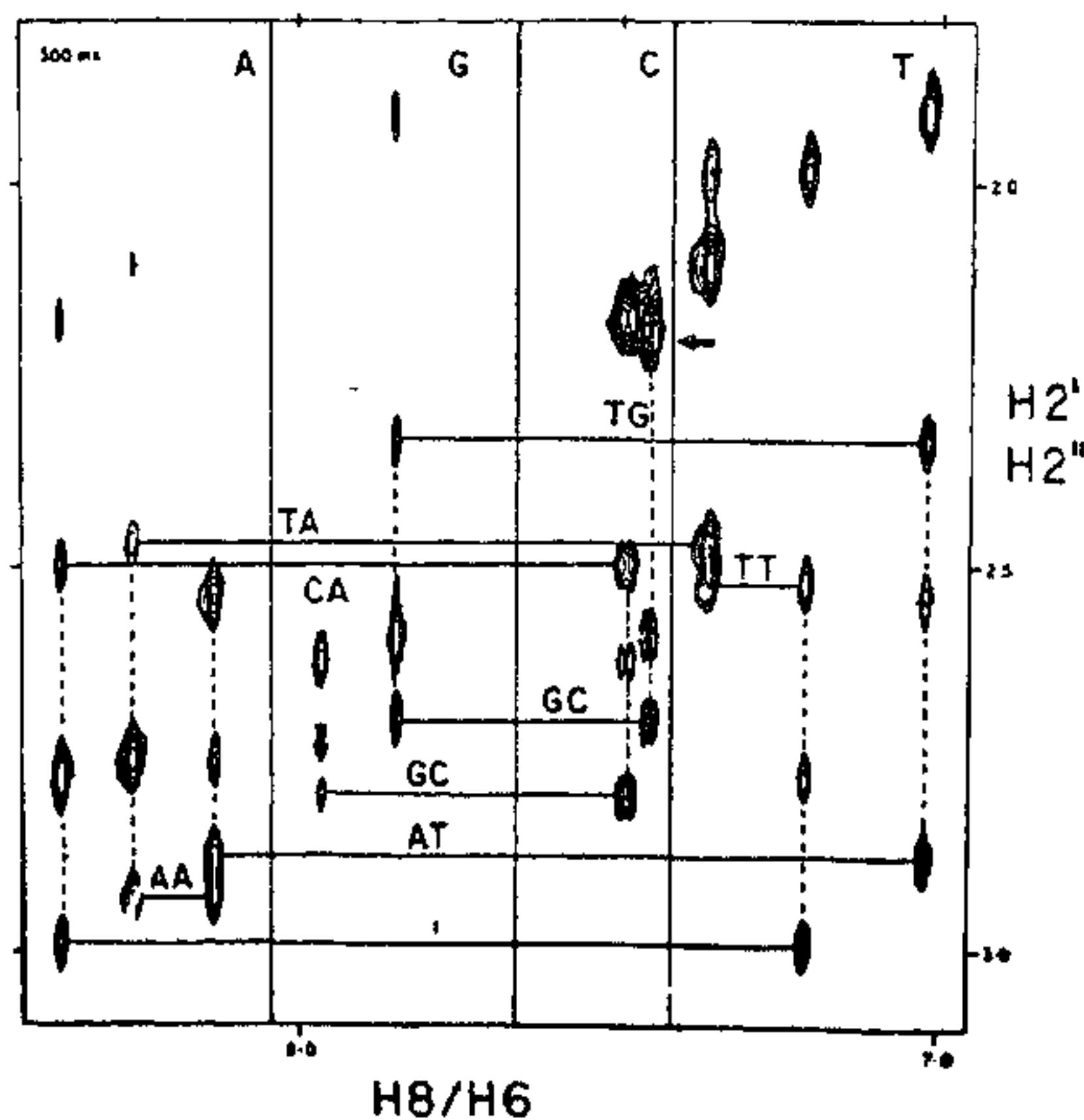
from H8/H6  $\rightarrow$  H2'' peaks is possible from TH6  $\rightarrow$  TCH<sub>3</sub> cross-peaks. Further, the RISD also allows identification of adjacent dinucleotide segments with a common partner; this is illustrated in the Figure 2 by dashed vertical lines joining the peaks belonging to the common units in triplets ATA and ACG. Thus, in a given molecule the complete sequence can be read out from H8/H6  $\rightarrow$  H2'' peaks is possible from TH6  $\rightarrow$  TCH<sub>3</sub> cross-peaks. Further, the RISD also allows identification joining the adjacent pairs with a common partner by vertical lines as indicated above. Such a pattern of horizontal and vertical lines may be abbreviated as SDCP (sequential dinucleotide connectivity pattern). In a given RISD, there will be only one SDCP if the molecule is self-complementary, and two patterns—of the same length—if the two strands of the duplex have different sequences. In the latter case, the sequence derived from the two patterns will be complementary. In both cases, the total number of As equals the total number of Ts and the total number of Gs equals the total number of Cs. These factors provide additional checks for the validity of the derived sequence and any error in A, G discrimination can be immediately identified.

Figure 3 illustrates the above methodology with an experimental spectrum taken from the literature<sup>15</sup>. We



**Figure 2.** Schematic display of H8/H6  $\rightarrow$  H2'' and H8/H6  $\rightarrow$  TCH<sub>3</sub> cross-peaks in the RISD for all possible dinucleotide segments. The spectral region is divided into subregions in accordance with the chemical shift information in Figure 1, pairs of peaks belonging to individual dinucleotide segments are joined by horizontal lines. In reality the RISD also contains peaks originating from H2' proton and these must be clearly distinguished. This discrimination can be easily obtained from NOESY spectra recorded with short mixing time (<100 ms)<sup>19</sup>. In each box, the relative positions of the peaks is not significant (they have been chosen arbitrarily for clarity). For example, the A peak of AT could be above or below the A peak of the AG segment. Accordingly, the positions of the two horizontal lines may get reversed, but will remain in their respective domains. Dashed vertical lines illustrate identification of adjacent dinucleotides with a common partner. Peaks originating from CH6 are shown as doublets.

have purposely taken a spectrum from work belonging to another laboratory to minimize our bias about the molecule. For the reader, the sequence is unknown and therefore, there is no bias about the analysis. The spectrum in Figure 3 represents the RISD without the peaks arising from the TCH<sub>3</sub> protons. Two cytosines can be identified straightaway from their doublet patterns. A total of ten base protons can be easily counted and, in every case, all the expected peaks to H2'/H2'' are clearly distinguishable. The thick vertical arrow identifies the base proton (H8) of the 5'-terminal nucleotide and indicates the beginning of SDCP. From another set of COSY and NOESY spectra it was concluded that the H2'' proton resonates downfield of the H2' proton in all the cases. Using this information, nine dinucleotide pairs originating from H2'' proton connectivities have been identified and joined by horizontal lines in the figure. They have been assigned in accordance with Figure 2 and their assignments are indicated above the horizontal lines. The adjacent dinucleotide pairs with a common partner have then been identified and joined by vertical dashed lines. From these, the sequence can be readily obtained as d-



**Figure 3.** Experimental demonstration of the sequencing procedure. The 2D NOESY spectrum shows NOE cross-peaks between H8/H6 protons and H2'/H2'' protons. Peaks originating from H2'' protons are joined by horizontal lines and their assignment to specific dinucleotide segments are shown above the lines. Adjacent dinucleotide pairs with common partners are joined by dashed vertical lines at the chemical shift position of the common base proton. The thick vertical arrow identifies the H8 proton of 5'-terminal nucleotide and indicates the beginning of SDCP. The thick horizontal arrow identifies the H2'' proton of the 3'-terminal nucleotide and indicates the end of SDCP in the spectrum. The NOESY spectrum has been taken from the literature<sup>17</sup>. Experimental conditions are: mixing time = 300 ms; temperature = 28°C, pH = 7.0.

GCATTAATGC. This is indeed the true sequence of the molecule<sup>15</sup>.

The above approach has also been successfully tested on a few DNA segments. An important requirement for the successful application of the proposed procedure is that the cross-peaks should be well separated in the NOESY spectrum. In this context use of the modern NMR techniques such as selective pulse techniques<sup>23</sup> and three-dimensional techniques<sup>24,25</sup> may be envisaged. Experimental conditions must also be suitably adjusted to ensure right-handed duplex structure for the molecule; other physical techniques such as circular dichroism will be helpful in this regard. Observation of imino proton resonances in H<sub>2</sub>O spectra will help confirm the duplex state of the DNA segment.

The NMR method of sequencing DNA segments also has some limitations. (i) At present the size of DNA segment has to be less than 20 base pairs; this arises due to large line width (short T<sub>2</sub>) problems and insufficient resolution in the spectra. (ii) Repetitive sequences such as AAAAATTTTT or GGGGGCCCCC are difficult to handle. (iii) Abnormal synthetic sequences such as hair pins or loop structures are not

amenable to analysis because of their modified chemical shift patterns.

1. Ernst, R. R., Bodenhausen, G. and Wokaun, A., *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*, Clarendon Press, Oxford, 1987.
2. Wüthrich, K., *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
3. Jeener, J., *Ampere Summer School on Magnetic Resonance and Related Phenomena*, Basko Polje, Yugoslavia, 1971.
4. Aue, W. P., Bartholdi, E. and Ernst, R. R., *J. Chem. Phys.*, 1976, **64**, 2229.
5. Jeener, J., Meier, B. H., Bachmann, P. and Ernst, R. R., *J. Chem. Phys.*, 1979, **71**, 4546.
6. Anil Kumar, Wüthrich, K. and Ernst, R. R., *Biochem. Biophys. Res. Commun.*, 1980, **95**, 1.
7. Chary, K. V. R., Hosur, R. V., Govil, G., Tan, Z. and Miles, H. T., *Biochemistry*, 1987, **26**, 1315.
8. Chary, K. V. R., Hosur, R. V., Govil, G., Tan, Z. and Miles, H. T., *Biochemistry*, 1988, **27**, 3858.
9. Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G. and Reid, B. R., *J. Mol. Biol.*, 1983, **171**, 319.
10. Ravikumar, M., Hosur, R. V., Roy, K. B., Miles, H. T. and Govil, G., *Biochemistry*, 1985, **24**, 7703.
11. Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z. and Miles, H. T., *Biochem. Biophys. Res. Commun.*, 1987, **144**, 26.
12. Sheth, A., Ravikumar, M., Hosur, R. V., Govil, G., Tan, Z., Roy, K. B. and Miles, H. T., *Biopolymers*, 1987, **26**, 1301.
13. Chary, K. V. R., Modi, S., Hosur, R. V., Govil, G., Chen, C. Q. and Miles, H. T., *Biochemistry*, 1989, **28**, 5240.
14. Sheth, A., Hosur, R. V., Govil, G., Hosur, M. V., Kannan, K. K., Tan, Z. and Miles, H. T., *Biochemistry*, 1989, **28**, 7275.
15. Chazin, W. J., Wüthrich, K., Hyberts, S., Rance, M., Denny, W. A. and Leupin, W., *J. Mol. Biol.*, 1986, **190**, 439.
16. Lown, J. W., Hanstock, C. C., Bleackley, R. C., Imbach, J. L., Rayner, B. and Vasseur, J. J., *Nucl. Acids. Res.*, 1984, **12**, 2519.
17. Hosur, R. V., Govil, G. and Miles, H. T., *Magn. Reson. Chem.*, 1988, **26**, 927.
18. Rinkel, L. J., Sanderson, M. R., Van der Marel, G. A., Van Boom, J. H. and Altona, C., *Eur. J. Biochem.*, 1986, **159**, 85.
19. Scheek, R. M., Boelens, R., Russo, N., Van Boom, J. H. and Kaptein, R., *Biochemistry*, 1984, **23**, 1371.
20. Wemmer, D. E., Chou, S. H., Hare, D. R. and Reid, B. R., *Biochemistry*, 1984, **23**, 2262.
21. Feigon, J., Leupin, W., Denny, W. A. and Kearns, D. R., *Biochemistry*, 1983, **22**, 5943.
22. Shindo, H., Okhubo, S., Matsumoto, U., Giessner-Prettre, C. and Zon, G., *J. Biomol. Struct. Dyn.*, 1988, **5**, 913.
23. Bermel, W., Kessler, H., Griesinger, C. and Oschkinat, H., *BRUKER Report*, 1986, **1**.
24. Griesinger, C., Sorensen, O. W. and Ernst, R. R., *J. Magn. Reson.*, 1987, **73**, 574.
25. Griesinger, C., Sorensen, O. W. and Ernst, R. R., *J. Magn. Reson.*, 1989, **84**, 14.

ACKNOWLEDGEMENT. We thank Prof. G. Govil for useful suggestions.

9 April 1990; revised 15 June 1990

#### National Symposium on Mushrooms

Place: Thiruvananthapuram  
Date: 22-24 January 1991  
Contact: Dr M. C. Nair  
Department of Plant Pathology  
College of Agriculture  
Vellayani 695 022

#### International Conference on Neutron Scattering

Place: Bombay, India  
Date: 21-25 January 1991  
Contact: Dr K. R. Rao  
Organizing Committee-NS '91  
Nuclear Physics Division  
Bhabha Atomic Research Centre  
Bombay 400 085

Topics to be covered include Interferometry; Neutron optics; High-temperature superconductors; Magnetic structures; Elementary excitations; Phase transitions; Molecular spectroscopy; Low-dimensional systems; Incommensurate and quasiperiodic systems; Liquids and amorphous systems; Surfaces and interfaces; Micelles, microemulsions and membranes; Polymers and macromolecules; Structures of biological interest; Emerging trends in instruments and techniques; Pulsed and steady-state sources.

#### International Symposium on Oceanography of the Indian Ocean

Place: Goa, India  
Date: January 1991

The symposium is aimed to provide a forum to present and discuss the advances in Oceanography of the Indian Ocean and its adjacent seas, especially in the post International Indian Ocean Expedition years.

Contact: Dr A. H. Parulekar  
Convener, Organizing Committee  
ISOIO, National Institute of Oceanography  
Goa 403 004

#### International Conference on Brazil Gold '91

Place: Belo Horizonte, Minas Gerais, Brazil  
Date: 7-22 May 1991  
Contact: Prof. V. K. Nayak  
Department of Applied Geology  
Indian School of Mines  
Dhanbad 826 004

The conference aims to provide a forum to present and discuss matters related to various aspects of geology, exploration and genesis of gold ores.