NMR observation of a novel A-tetrad in a telomeric DNA segment in aqueous solution

Prasanta K. Patel and R. V. Hosur*

Department of Chemical Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Mumbai 400 005, India

NMR studies on the telomeric DNA segment d-AG_3T show that the molecule exists as a mixture of several multistranded structures whose stabilities are dependent on added salt concentrations. There is one major conformation which can be selectively studied at low salt concentration and the NMR data indicate that this is a parallel stranded quadruplex with a novel A-tetrad at the 5′-end. The A-bases adopt the *syn* glycosidic conformation as against the *anti* conformation adopted by the G-bases in the three G-tetrads.

TELOMERES are nucleoprotein complexes at the ends of chromosomes which play important roles in chromosome replication and ageing[1–5]. They contain tandem repeats of short DNA sequences for up to several hundreds of base pairs and a short single stranded overhang at the 3′ end. The length of the telomeric DNA seems to be related to the age of the cell and is dependent on the activity of an enzyme called *telomerase*, which is a reverse transcriptase with the exclusive role of synthesizing and replicating telomeric DNA. In normal cells the telomere length is seen to decrease progressively with every cycle of replication and thus is taken to be a monitor of the ageing process. In contrast, in cancerous cells and immortalized cells the length of the telomeric DNA is maintained by a higher *telomerase* activity. Regulation of the *telomerase* enzyme is brought about by a class of small proteins which contain domains having homologies with the Myb proteins, the oncogene products[6]. Thus these proteins which bind specifically to telomeric DNA have also been implicated to have some roles in oncogene regulations. Telomeric DNA is also believed to be important in the organization of the chromosomes inside the nucleus and in preventing fusion of the chromosomes.

In view of the above important biological roles of the telomeres, they have been a subject of much investigation in recent years from the view-point of understanding their three-dimensional structures[7–23]. The telomeric DNA sequence repeats, which are rich in G-nucleotides, have been shown to form a variety of multistranded structures. T_2AG_3 and T_2G_4, the single repeats of *human* and *tetrahymena* telomeres respectively were seen to form parallel stranded G-quadruplexes with all the nucleotides having *anti* glycosidic torsion angles[16]. The sequence $AG_3(T_2AG_3)_3$, on the other hand, formed an antiparallel quadruplex in which each strand had one parallel and one antiparallel neighbours[17]. The glycosidic torsion angles for the G's altered between *syn* and *anti* in each of the strands and in each tetrad, G(*syn*)–G(*syn*)–G(*anti*)–G(*anti*) pattern was observed. The *Oxytricha* sequence $(T_4G_4)_n$ also formed a similar quadruplex with T4 loops in solution[18,19] but in the crystalline state, the G-tetrad had G(*syn*)–G(*anti*)–G(*syn*)–G(*anti*) alignment and also different loop orientations[20]. A recent study on a truncated *Bombyx mori* sequence TAGG showed the formation of a novel T–A–A triad capping the G-quadruplex with *anti* glycosidic angles[21]. Thus it is evident that there is a great variety in telomeric DNA structures which go well beyond the usual double helical DNA and these along with the other multistranded structures such as triplexes, junctions, i-motifs etc. (see articles in *Methods in Enzymology*, vol. 261, edited by T. L. James, Academic Press, 1995) establish new hitherto not-understood or not known functional roles for the DNA molecule. The functional relevance of tetraplexes has been further supported by the identification of proteins, which recognize and promote the formation of four stranded DNA structures[24–27]. In this context, we report here a novel observation by NMR of the formation of a new structural motif, namely, the A-tetrad in the truncated telomeric DNA segment AG_3T in aqueous solution. The molecule forms a parallel stranded quadruplex, in which the A-tetrad stacks over the adjacent G-tetrad and also exhibits dynamism in its H-bonding patterns. To our knowledge, this is the first observation of an A-tetrad in any of the DNA structures reported so far.

The DNA sequences were synthesized on an Applied Biosystems 392 automated DNA-synthesizer on 10 μM scale using solid phase β-cyanoethyl phosphoramidite chemistry and were purified by polyacrylamide gel electrophoresis. The purity of the DNA was checked by both polyacrylamide gel electrophoresis and NMR at high temperatures.

Three NMR samples were prepared by taking approximately 100 $A_{260\,nm}$ units of oligonucleotide in three different 0.6 ml (90% H_2O/10% D_2O) solutions having 0.2 mM potassium phosphate buffer, pH 7.3 and 0 mM, 50 mM and 100 mM KCl. The same were converted into D_2O samples by lyophilization and redissolution for the experiments in D_2O alone.

For each of the samples, phase sensitive NOESY experiments with mixing times in the range from 80 to 350 ms were performed on a Bruker AMX 500 spectrometer. The H_2O signal was suppressed using the 'jump-return' pulse sequence. TOCSY spectra were recorded on the D_2O samples with mixing times of 30 and 80 ms. The time domain data consisted of 2048 comple

*For correspondence. e-mail:hosur@tifrvax.tifr.res.in.

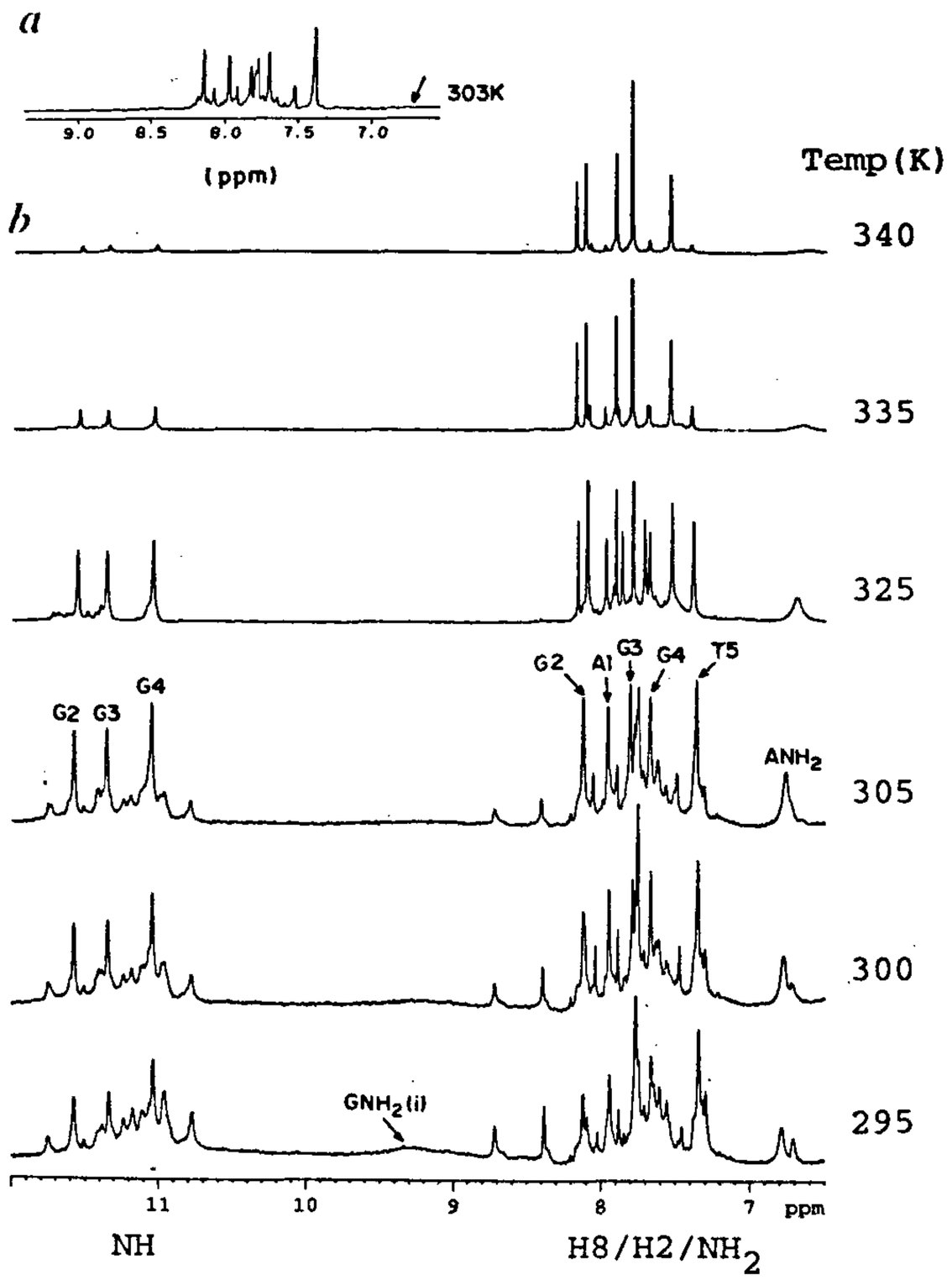
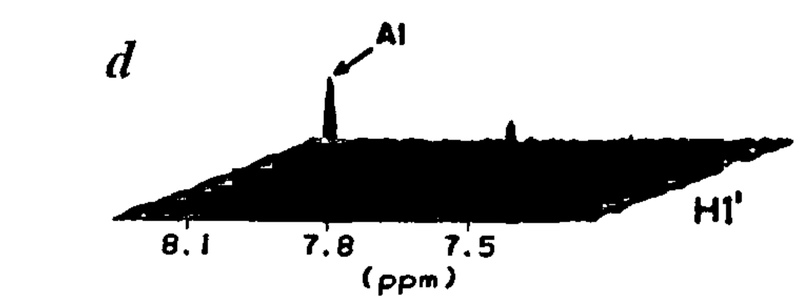


Figure 1. *a*. Base proton region of proton NMR spectrum of AG_3T in D_2O at pH 7.3, 100 mM KCl. This allows distinction of the exchangeable and non-exchangeable proton resonances. The arrow indicates the position of the exchangeable ANH_2 resonance in the H_2O spectra shown in *b*. The pattern of the non-exchangeable proton resonances seen is similar to that in *b* at the appropriate temperature. *b*, Base proton spectra of AG_3T in H_2O, pH 7.3, 100 mM KCl at different temperatures. In addition to the major quadruplex conformer, for which the peak assignments are indicated, other conformers are seen to be present at low temperatures. The quadruplex melts into single strands as the temperature is raised.



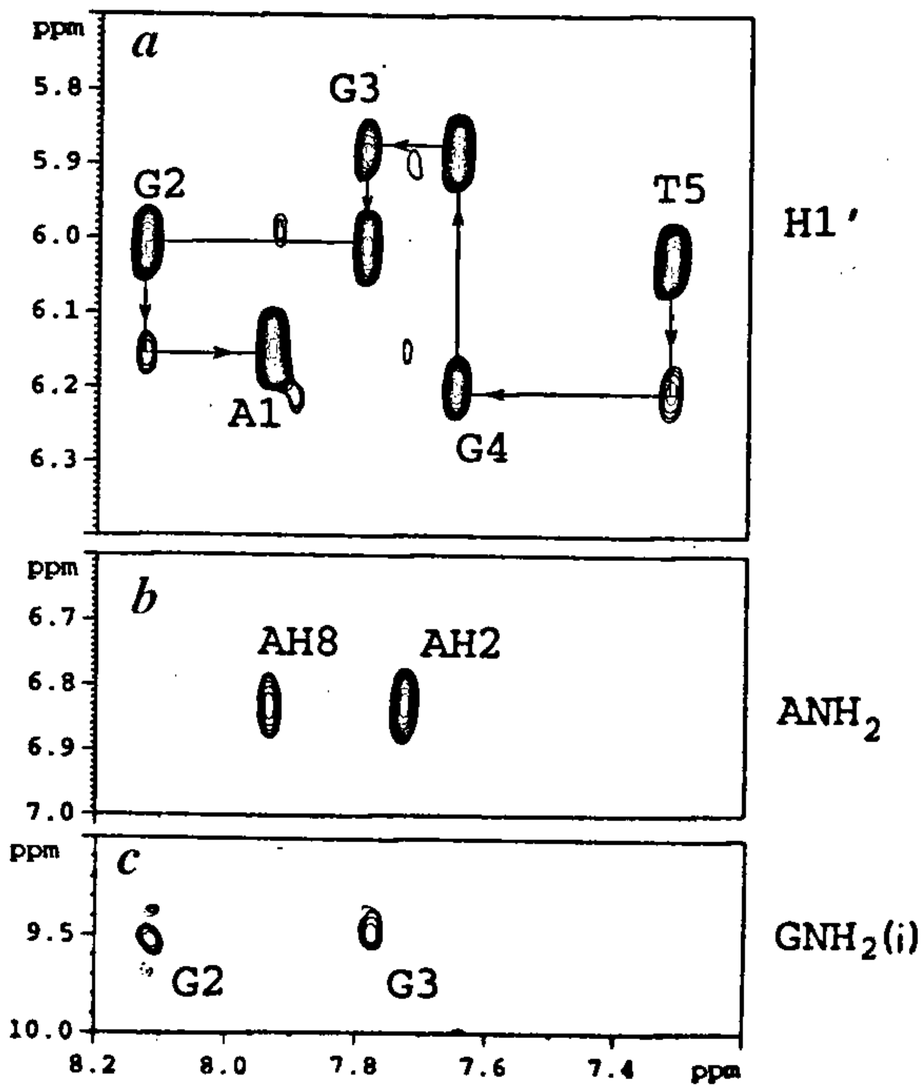


Figure 2. *a*. Selected region of the NOESY spectrum (200 ms) of AG_3T at 285 K, pH 7.3, showing sequential (H8–H1′) connectivity pattern. Peak labels identify self peaks. *b*, Strong NOE cross peaks from ANH_2 to (AH8, AH2), implying A–A pair formation as shown in Figure 4. *c*, GNH2–GH8 NOE cross peaks for the G2 and G3 nucleotide units. G4 amino photons are too broad to show cross peaks. These peaks confirm on the one hand the formation of the G-tetrad, and on the other, the fact that both the imino and the H8 protons have the same nucleotide label indicates that the four strands have parallel orientations. *d*, A stacked plot of the same region as in *a* to emphasize the high intensity of the H8–H1′ cross peak of A, which indicates that the A nucleotide has a *syn* glycosidic conformation.

points in t_2 and 512 increments in t_1 dimension. The relaxation time delay was kept 2 s for all the experiments. The data were apodized by shifted (60°–90°) sine bell functions prior to 2D Fourier transformations.

Figure 1 shows selected regions of the NMR spectra of AG_3T recorded with the sample containing 100 mM KCl in D_2O and H_2O at different temperatures. The multiplication of the resonances at low temperature indicates that more than one conformer coexist, but as the temperature is raised, the more stable conformation survives, which eventually melts into single strands. The same pattern was observed with the other samples containing 0 and 50 mM KCl concentrations except that at lower salt concentrations, the melting temperatures for all the conformers were lower by about 10–15 degrees. But, at low salt or no salt, the proportion of the major conformer was higher in the equilibrium mixture. Therefore, for detailed analysis, we have chosen the sample which did not have any added salt. The assignments of all the exchangeable and nonexchangeable protons for the major conformer could be obtained from intranucleotide TOCSY correlations and sequential NOE correlations expected for right-handed DNA segments[28,29]. Figure 2 *a* shows the illustrative NOE connectivities for nonexchangeable H8...H1′ protons and Figure 3 shows the connectivities from the exchangeable protons. In Figure 3, the G4NH–G4NH_2 cross peaks are not visible at the contour levels plotted, but at lower levels they were seen to be very broad peaks.

The NMR data described above are clearly indicative of formation of stable secondary structures. The association patterns of such G-rich sequences have been ex-

tensively documented in the literature[12–19] and it is observed that continuous G stretches invariably to form quadruplexes. Both parallel and antiparallel strand orientations are possible and these have distinctly different NMR spectral features. First of all the antiparallel structures have an asymmetry which arises because of the presence of both *syn*, *anti* glycosidic torsion angles in each G-tetrad and this results in multiple sets of resonances. The sequential NOE connectivities involving G-imino protons are also very different for the two types of structures because of the different patterns of GNH–GNH, and GNH–GH8 intrastrand and interstrand proton–proton distances. We do not wish to list all these differences here which have been very well documented[12–19]. Suffice it to point out the significant features of the NMR spectra are presented in Figures 1–3: (1) The G-imino protons resonate in the range of 10.7–11.8 ppm and exhibit high stability, both of which are characteristic of quadruplex structures. (2) $GNH_2(i)$–GH8 NOE cross peaks are seen for G2 and G3 units (Figure 2 *c*) and these confirm G-tetrad formation. (3) Continuous sequential connectivity pattern, $(GH8)_i$–$(GH1')_{i,\,i-1}$ is observed (Figure 2 *a*) which is characteristic of *anti* glycosidic conformation for all the G's in a right handed helix. (4) Continuous sequential $(GNH)_i$–$(GH8)_{i,\,i-1}$ connectivities are observable (Figure 3) which is not possible if there is an alternation between *syn* and *anti* conformations in any strand. (5) A single resonance is seen for each of the protons which implies equivalence of the four strands.

Taken together, all these indicate that the three G-nucleotide stretch in the molecule forms a symmetrical parallel stranded G-quadruplex structure as shown in Figure 4 *a*; the H-bonding pattern in the G-tetrad is also shown for ready reference.

In addition to the G-quadruplex features described above, the NMR spectra shown in Figures 1 and 2 have some distinctive features. Firstly, the ANH_2 protons are highly stable and melt synchronously with the GNH protons (Figure 1 *b*). This stability of the ANH_2 protons is in sharp contrast to that of the H-bonded G-NH_2 protons which disappear at much lower temperatures. Second, a strong AH8–AH1′ NOE cross peak is seen (Figure 2 *d*) which indicates that the A-nucleotide exhibits a *syn* glycosidic conformation. Both these observations are surprising since the A nucleotide is at the 5′ terminal and one would not expect such a stable exchangeable proton resonance or such a strong deviation from the standard *anti* glycosidic conformation. Further, the NOESY spectra show ANH_2–(AH2, AH8) cross peaks (Figure 2 *b*) and their intensities compared to the self GH8–GH1′ cross peaks for G4 and G3 nucleotide units in Figure 2 *a* permit putting an upper bound of 3.9 Å for the ANH_2–AH2 and ANH_2–AH8 distances (the intra-nucleotide H8–H1′ distance can have a maximum value of 3.9 Å). The actual distance could in fact be much less, since the exchange contributions, if any, would only lower the NOE intensities and thus cause an overestimation of the distances. Thus it is clear that the ANH_2–(AH2, AH8) cross peaks are inter-nucleotide in nature and the A-amino protons are imvolved in A–A pair formation via H-bonds. Although the peaks in Figure 2 *b* do not show any chemical shift difference between the two A-amino protons, we observed in some other spectra recorded under different conditions that they may have slightly different chemical shifts. This means the amino protons giving rise to the cross peaks to AH8 and AH2 are different. These observations in conjunction with the requirement of equivalence of the four A-bases, indicate that a symmetrical A-tetrad is formed in which the adjacent A-bases are held together by H-bonds via ANH_2 protons. The spectra also show several NOE correlations between the adjacent A1 and G2 nucleotide units: G2NH–($A1NH_2$, A1H8, A1H2 and A1H2′/2″) (Figure 3), indicating a good stacking of the two bases. As a consequence, the G2-amino is better

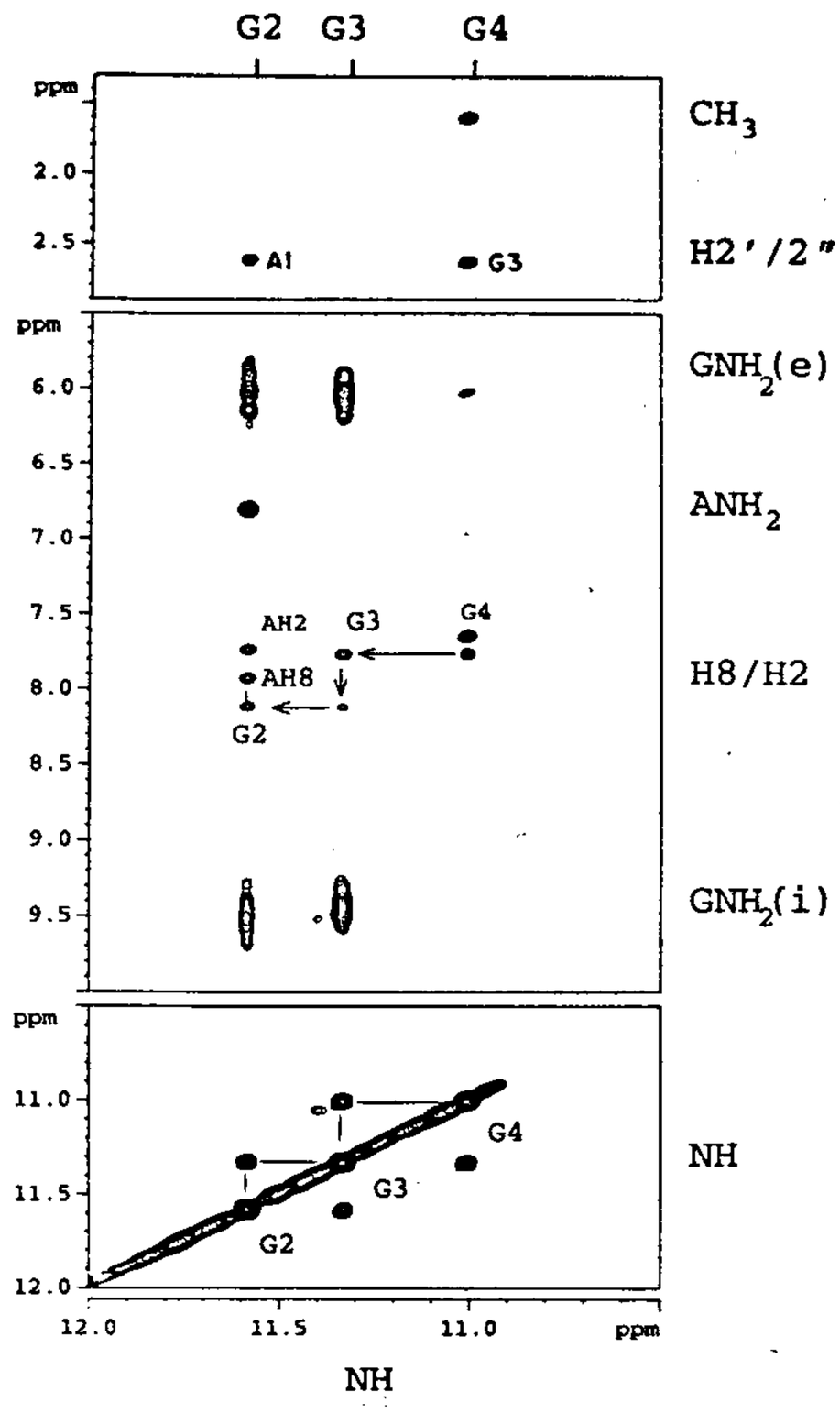


Figure 3. NOE cross peaks from the same NOESY spectrum as in Figure 2, arising from the G-imino protons in the quadruplex. The various sequential connections and peak identifications are given. The self cross-peaks are identified by sequence specific labels near the peaks.

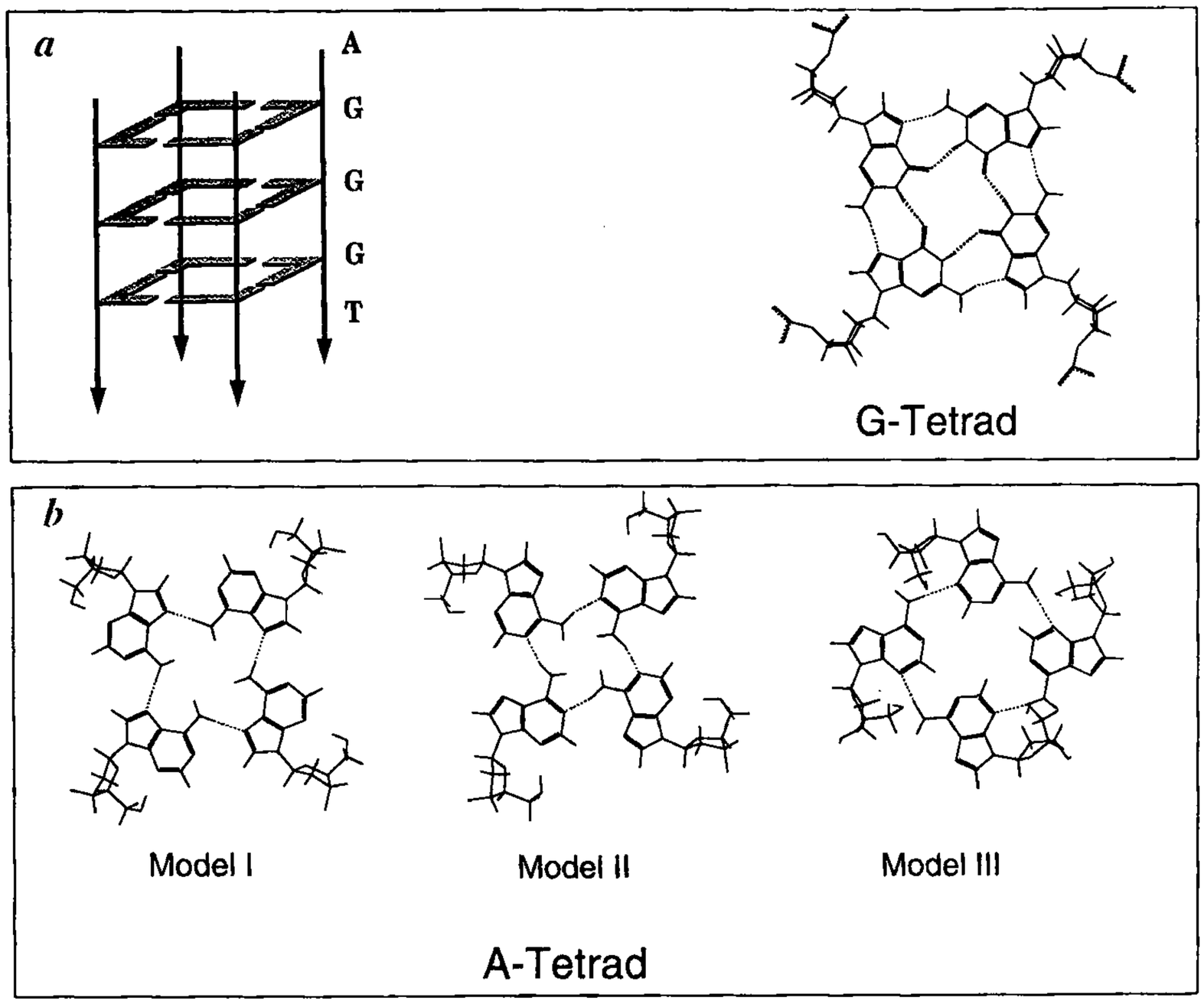


Figure 4. *a*, Schematic parallel stranded G-quadruplex structure formed by the central three G-nucleotides in AGGGT along with the characteristic H-bonding pattern in a G-tetrad. The bases are shaded to indicate the directional sense of the H-bonds. *b*, Three models for A-tetrad formation. The dotted lines indicate the H-bonds. Model I has opposite sense of H-bond directions compared to models II and III.

shielded from the solvent than is the G4-amino, and this explains the lower line width of the G2-amino resonance (Figure 3).

Three models for symmetrical A-tetrad formation can be conceived and these are shown in Figure 4 *b*. However, the NMR data do not favour Model III, since here, the amino groups lie on the exterior and would exchange very rapidly with the solvent. Besides, in Model III several short distances appear between the A-amino protons and the backbone protons for which no NOE correlations are observable. Model I is supported by the ANH_2–AH8 cross peak and Model II is supported by the ANH_2–AH2 cross peak (Figure 2 *b*); notice that different A-amino protons are involved in the pairing in the two models. Thus if both the models are contributing, each of the ANH_2 protons would exist partly in the H-bonded state and partly in the free state and then the observation of a single resonance (~6.8 ppm) for each of them (and also for the protons on the adjacent G which would have different chemical shifts for the two models), must be explained by a rapid interconversion between the two alignments; the peak position would be at the average of H-bonded (~7.5–9.0 ppm) and free amino (~5.5–6.5 ppm) proton chemical shifts. We may point out that this interconversion between the two alignments in I and II can be brought about very easily by a small concerted rotation of the A-base orientations with respect to each other in the tetrad.

In summary, our observation of an A-tetrad capping the G-quadruplex in a telomeric DNA sequence is unprecedented. This not only enhances the variety in quadruplex structures at the telomere ends but also generates new structural thinking for other regions of DNA as well and thus may lead to new functional roles for DNA. This speculation finds support in a recent observation of a two base A–A platform in the crystal structure of the P4–P6 domain of the *Tetrahymena* self-splicing group I ribozyme[30]. This involves a side-by-side alignment of two adenine bases.

1. Zakian, V. A., *Annu. Rev. Genet.*, 1989, **23**, 579–604.
2. Blackburn, E. H., *J. Biol. Chem.*, 1990, **265**, 5919–5921.
3. Blackburn, E. H., *Nature*, 1991, **350**, 569–573.
4. Guo, Q., Lu, M. and Kallenbach, N. R., *J. Biol. Chem.*, 1992, **267**, 15293–15300.
5. Blackburn, E. H., *Cell*, 1994, **77**, 621–623.
6. Konig, P., Fairall, L. and Rhodes, D., *Nucleic Acids Res.*, 1998, **26**, 1731–1740.
7. Sen, D. and Gilbert, W., *Nature*, 1988, **334**, 364–366.
8. Sundquist, W. I. and Klug, A., *Nature*, 1989, **342**, 825–829.
9. Williamson, J. R., Raghuraman, M. K. and Cech, T. R., *Cell*, 1989, **59**, 871–880.
10. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. and Blackburn, E. H., *Cell*, 1987, **51**, 899–908.

11. Murchie, A. I. H. and Lilley, D. M., *EMBO J.*, 1994, **13**, 993–1001.
12. Schultze, P., Smith, F. W. and Feigon, J., *Structure*, 1994, **2**, 221–233.
13. Wang, K. Y., Swaminathan, S. and Bolton, P. H., *Biochemistry*, 1994, **33**, 7517–7527.
14. Wang, Y. and Patel, D. J., *J. Mol. Biol.*, 1993, **234**, 1171–1183.
15. Gupta, G., Garcia, A. E., Guo, Q., Lu, M. and Kallenbach, N. R., *Biochemistry*, 1994, **32**, 7098–7103.
16. Wang, Y. and Patel, D. J., *Biochemistry*, 1992, **31**, 8112–8119.
17. Wang, Y. and Patel, D. J., *Structure*, 1993, **1**, 263–282.
18. Wang, Y. and Patel, D. J., *J. Mol. Biol.*, 1995, **251**, 76–94.
19. Smith, F. W. and Feigon, J., *Nature*, 1992, **356**, 164–168.
20. Kang, C. H., Zhang, X., Ratcliff, R., Moyzis, R. and Rich, A., *Nature*, 1992, **356**, 126–131.
21. Abdelali, K., Bouaziz, S., Wang, W., Jones, R. A. and Patel, D. J., *Nat. Struct. Biol.*, 1997, **4**, 382–389.
22. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. and Luisi, B., *J. Mol. Biol.*, 1997, **273**, 171–182.
23. Mohanty, D. and Bansal, M., *Biophys. J.*, 1995, **69**, 1046–1067.
24. Fang, G. and Cech, T. R., *Cell*, 1993, **74**, 875–885.
25. Liu, Z. and Gilbert, W., *Cell*, 1994, **77**, 1083–1092.
26. Liu, Z., Lee, A. and Gilbert, W., *Proc. Natl. Acad. Sci. USA*, 1995, **92**, 6002–6006.
27. Rhodes, D. and Giraldo, R., *Curr. Opin. Struct. Biol.*, 1995, **5**, 311–322.
28. Wuthrich, K., *NMR of Proteins and Nucleic Acids*, Wiley, New York, 1986.
29. Hosur, R. V., Govil, G. and Miles, H. T., *Magn. Reson. Chem.*, 1988, **26**, 927–944.
30. Gate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R. and Doudna, J. A., *Science*, 1996, **273**, 1696–1699.

ACKNOWLEDGEMENTS. The facilities provided by the National Facility for High Field NMR at TIFR are gratefully acknowledged.

Received 1 April 1998; revised accepted 4 May 1998

Pollen production in some terrestrial angiosperms

Amal Kumar Mondal and Sudhendu Mandal
Department of Botany, Visva-Bharati, Santiniketan 731 235, India

Pollen production in terms of number per anther along with the particulars of anther number per flower and anther length was determined for 54 angiospermous plant species collected from different areas of Burdwan district, West Bengal. Pollen production varied widely from genus to genus and from species to species within the same genus of a family. There exists a correlation between pollen production and habit of plant. An increase in pollen production from herbs → shrubs → trees was observed and analysed. Variation in the number of pollen grains produced has been noticed in anemophilous, entomophilous and amphiphilous plant taxa. It is presumed that high-pollen producers are cross-pollinated, whereas low-pollen producers are either self-pollinated or apomictic.

A plant during its entire flowering period produces large amounts of pollen grains most of which are not involved in fertilization. These large amount of pollen released may float in air or water, and finally get deposited on the earth's surface[1]. It has been proved that pollen grains of some plant species are bio-pollutants as they cause various types of allergic diseases in human beings[2]. The knowledge of the quantitative production and methods of dispersal of pollen grains is significant as these factors, directly or indirectly, are involved in causing pollution of the environment, and, in the interpretation of data on the pollen content of the atmosphere, honey and sedimentary deposits[3]. These data also give some idea about the frequency of presence of particular plant pollen grains in the atmosphere and hydrosphere.

In flowering plants, pollen productivity is referred to as number of pollen grains produced per anther of the flower. The total production of pollen grains in a particular plant depends on number of anthers per flower and number of flowers per plant. This has significance in aerobiological work dealing with allergy problems. The allergenic pollens in the atmosphere if present in abundance in the air cause allergic symptoms according to Thommen's postulate[4]. The dispersal in air is controlled by factors such as morphology of the flowers, the season of flowering, and characters like anemophily, entomophily and amphiphilly. Various types of allergenic manifestation in human beings are related with dispersal of pollen grains in the air. It has been proved that plants known for high-pollen production are significant for causing pollen allergy problems. Pollen production in anemophilous plants has been studied by Reddi and Reddi[5]. Further, the pollen profile of a particular area reflects the pattern of vegetation of the area under investigation.

In this study, pollen production of 54 selected angiospermic plants profusely growing in Burdwan district, West Bengal, were investigated to establish their supposed role in pollen allergy problems. Pollen production was studied following the methods proposed by Nair and Rastogi[6] and Mandal and Chanda[2]. Pollen grains from the 54 selected angiospermic plant taxa were collected from different biozones of Burdwan district. Unopened flowers were collected between 8.30 and 9.30 a.m. From each flower one anther was crushed and dispersed uniformly in 50 drops of 50% glycerine. One drop of this mixture was placed on a slide and covered with a 18 × 18 mm cover glass. The number of pollen grains