Journal of Biomolecular NMR, **24:** 329–337, 2002.
KLUWER/ESCOM

Random coil proton chemical shifts of deoxyribonucleic acids

Sik Lok Lam*, Lai Nang Ip, Xiaodai Cui & Cheuk Nang Ho
Department of Chemistry, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Received 12 July 2002; Accepted 9 October 2002

Key words: DNA, proton chemical shift, random coil, sequence effect

Abstract

Sixteen 17-nucleotide DNA sequences have been used to determine the sequence effect on random coil DNA proton chemical shifts. Based on the proton chemical shifts measured for the central nucleotides in 64 triplets and the correction factors determined for the next nearest neighbor effects, a parameter set has been derived for predicting random coil DNA proton chemical shifts. The root-mean-square deviation (RMSD) between the predicted and the observed aromatic H6/H8 proton chemical shifts of ∼200 data from 22 random coil DNA sequences was determined to be 0.02 ppm with a correlation coefficient of 0.998. For the H1′, H2′, H2″ and H3′ sugar protons, the RMSD values between the predicted and the experimental shifts were found to be 0.02, 0.03, 0.03 and 0.02 ppm, respectively.

Introduction

The 'structural' chemical shifts, which determine the differences between the chemical shifts of a well-defined structure and a reference state, have been shown to correlate well with protein secondary structures (Case et al., 1994; Szilagyi, 1995; Williamson and Asakura, 1997; Wishart and Sykes, 1994), DNA helical structures (Altona et al., 2000; Wijmenga et al., 1997) and RNA helical structures (Cromsigt et al., 2001). In proteins, a simple maximum and minimum relationship exists between the main chain atom structural shifts and secondary structures. For example, the difference between the H^{α} chemical shift and the random coil value is maximum in the β-sheet conformation but minimum in the α-helical conformation. In contrast, this simple maximum and minimum relationship between the structural shifts and secondary structures does not exist in nucleic acids. In order to investigate the structure and chemical shift relationship of nucleic acids, it is useful to have as much nucleic acid chemical shift information as possible. This includes the random coil DNA chemical shifts, which are of interest for NMR structural studies and theoretical studies of DNA chemical shifts. For proteins, the random coil chemical shifts of the main-chain H^{α} are independent of the amino acid sequence and thus were derived from short and disordered peptides (Bundi and Wüthrich, 1979; Merutka et al., 1995; Wishart et al., 1995). However, the random coil DNA proton chemical shifts are sequence-dependent because the conformations, the aromatic ring current effects and the electrostatic contributions from phosphates are variable in short oligonucleotide fragments (Dejaegere et al., 1999; Wüthrich, 1986). Therefore, random coil DNA proton chemical shifts cannot be readily derived from short and disordered oligonucleotides.

In this study, sixteen 17-nucleotide DNA sequences containing all 64 different types of triplets have been used to determine the sequence effect on random coil DNA proton chemical shifts. A parameter set has been derived for predicting random coil DNA proton chemical shifts. The predicted shifts for the aromatic H6/H8 protons and the H1′, H2′, H2″ and H3′ sugar protons have been shown to agree well with the experimental values and a significant improvement has been achieved when compared with the results obtained from a previously established procedure based on RNA-derived neighboring parameters (Bell et al., 1983, 1985; Hader et al., 1982).

*To whom correspondence should be addressed. E-mail: lams@cuhk.edu.hk

Table 1. DNA sequences used in studying the sequence effect on random coil DNA chemical shifts

Sequence name	Nucleotide position[a]																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
SS1	**T**	C	C	C	**T**	T	A	G	**T**	A	G	A	**T**	G	C	T	**T**
SS2	**T**	G	C	C	**T**	C	T	G	**T**	T	G	A	**T**	A	C	T	**T**
SS3	**T**	A	C	C	**T**	G	T	G	**T**	C	A	A	**T**	T	C	T	**T**
SS4	**T**	T	C	C	**T**	A	T	G	**T**	G	A	A	**T**	C	G	T	**T**
SS5	**T**	C	G	C	**T**	T	T	G	**T**	A	A	A	**T**	G	G	T	**T**
SS6	**T**	G	G	C	**T**	C	C	G	**T**	T	A	A	**T**	A	G	T	**T**
SS7	**T**	A	G	C	**T**	G	C	G	T	C	T	A	**T**	T	G	T	**T**
SS8	**T**	T	G	C	**T**	A	C	G	**T**	G	T	A	**T**	C	A	T	**T**
SS9	**T**	C	A	C	**T**	T	C	G	**T**	A	T	A	**T**	G	A	T	**T**
SS10	**T**	G	A	C	**T**	C	G	G	**T**	T	T	A	**T**	A	A	T	**T**
SS11	**T**	A	A	C	**T**	G	G	G	**T**	C	C	A	**T**	T	A	T	**T**
SS12	**T**	T	A	C	**T**	A	G	G	**T**	G	C	A	**T**	C	T	T	**T**
SS13	**T**	C	T	C	**T**	T	G	G	**T**	A	C	A	**T**	G	T	T	**T**
SS14	**T**	G	T	C	**T**	C	A	G	**T**	T	C	A	**T**	A	T	T	**T**
SS15	**T**	A	T	C	**T**	G	A	G	**T**	C	G	A	**T**	T	T	T	**T**
SS16	**T**	T	T	C	**T**	A	A	G	**T**	G	G	A	**T**	C	C	T	**T**

[a]Nucleotide position starts from 5′-end.

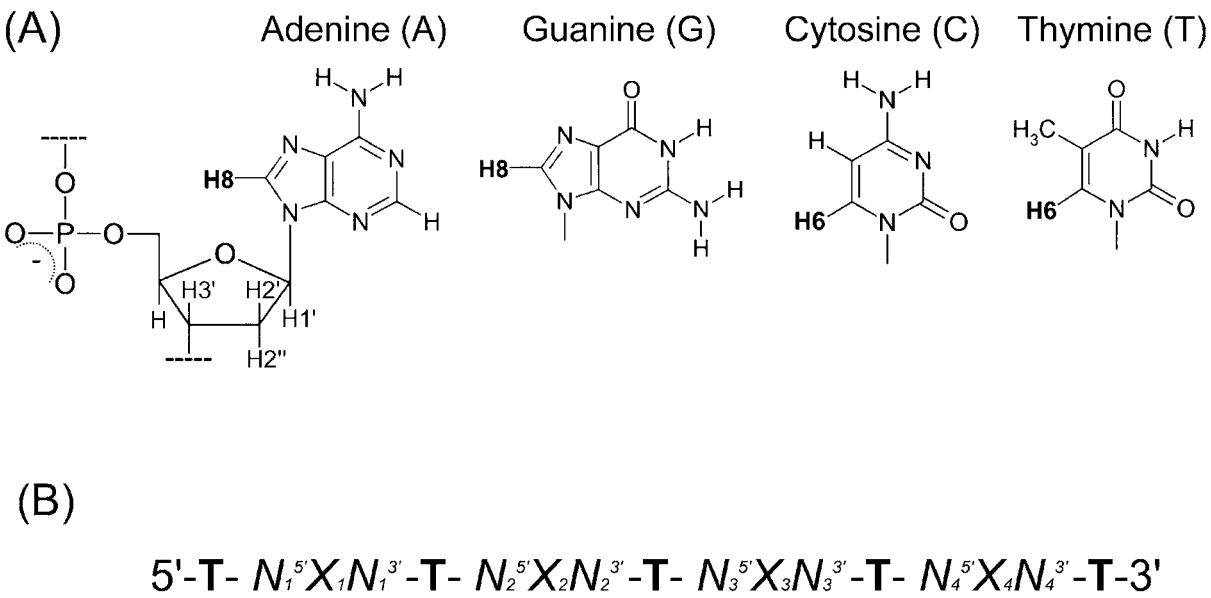

Figure 1. (A) Chemical structures of purines (adenine and guanine) and pyrimidines (cytosine and thymine) deoxyribonucleotides. The sequence effects on the chemical shifts of the aromatic protons (H6 and H8) and the sugar protons (H1′, H2′, H2″ and H3′) are investigated in this study. (B) Design of 17-nucleotide DNA sequences. The symbols $N_i^{5'}$ and $N_i^{3'}$ represent the nearest neighbor of the central nucleotide X_i on its 5′-end and 3′-end, respectively.

Table 2. Third nearest neighbor effect on DNA aromatic proton chemical shifts

Sequence name	Nucleotide position 12	13	**14**[a]	15	16	17[b]	Δ_3 (ppm)[c]
Group A							
SS16	A	T	**C**	C	T	T (7.633)	0.003
SS1	A	T	G	C	T	T (7.634)	0.004
SS2	A	T	**A**	C	T	T (7.630)	0.000
SS3	A	T	**T**	C	T	T (7.643)	0.013
Group B							
SS4	A	T	**C**	G	T	T (7.623)	0.011
SS5	A	T	**G**	G	T	T (7.617)	0.005
SS6	A	T	**A**	G	T	T (7.612)	0.000
SS7	A	T	**T**	G	T	T (7.615)	0.003
Group C							
SS8	A	T	**C**	A	T	T (7.611)	0.012
SS9	A	T	**G**	A	T	T (7.603)	0.004
SS10	A	T	**A**	A	T	T (7.599)	0.000
SS11	A	T	**T**	A	T	T (7.612)	0.013
Group D							
SS12	A	T	**C**	T	T	T (7.644)	0.012
SS13	A	T	**G**	T	T	T (7.637)	0.005
SS14	A	T	**A**	T	T	T (7.632)	0.000
SS15	A	T	**T**	T	T	T (7.646)	0.014

[a]The bolded nucleotides at position 14 are the third nearest neighbors of T17.
[b]The aromatic proton chemical shifts of T17 (in ppm) are reported in parentheses.
[c]Δ_3 represents the third nearest neighbor effect on T17. $\Delta_3 = \delta_{T17} - \delta_{T17A}$ where δ_{T17} is the T17 chemical shift and δ_{T17A} is the T17 chemical shift with an adenine nucleotide as the third nearest neighbor in the group.

Materials and methods

DNA samples

Sixteen 17-nucleotide DNA sequences, namely SS1 to SS16 (Table 1), were synthesized using solid-phase phosphoramidite chemistry in an Applied Biosystems Model 392 DNA synthesizer followed by polyacrylamide gel electrophoresis purification. The concentrations of purified DNA samples were kept at 1.0 mM by dissolving 0.50 μmoles of DNA in 500 μl solution containing 8 M urea, 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS), 150 mM NaCl and 10 mM sodium phosphate at pH 7. The samples were then dried and re-dissolved in 99.9% D_2O twice and finally put into 5 mm Wilmad PP528 NMR tubes. Apart from the 16 sequences, four other 17-nucleotide DNA, one 15-nucleotide DNA, one 12-nucleotoide DNA and three 11-nucleotide DNA sequences were also synthesized and purified for testing the accuracy of the newly developed chemical shift prediction protocol.

NMR measurements

All NMR experiments were performed on a Bruker ARX-500 NMR spectrometer operating at 500.13 MHz. A 5-mm inverse broadband probe was used and the acquired spectral data was processed using Bruker XWIN-NMR software. One-dimensional ^1H spectra were simply acquired with a presaturation pulse to suppress the residual HDO peak. The most upfield signal of DSS was set at 0 ppm to serve as an internal chemical shift reference. Two-dimensional NOESY experiments were performed at 1 sec mixing time with the time-proportional phase incrementation (TPPI) phase cycling method (Marion and Wuthrich, 1983). The recycling delay was set to two seconds and 512 FIDs with each consisting of 4096 complex data points were collected. At least 24 scans were collected for each FID and the acquired data matrix were finally zero-filled to give a 4k × 4k data set with cosine window function applied to both dimensions. All 2D NOESY experiments were acquired at 25 °C except for SS7 and SS16, which were acquired at 35 °C. For the resonance assignments of the 15-nucleotide hairpin sample and the 12-nucleotide self-complementary duplex at 85 °C, the samples were prepared in 90% H_2O/10% D_2O solution and two-dimensional ROESY (Bax and Davis, 1985) experiments at 600 ms spin-lock time were performed.

Results and discussion

Among the different types of non-exchangeable protons in deoxyribonucleotides, the aromatic base proton resonances have a wider spectral dispersion than the sugar protons. Therefore, the aromatic H8 of purines and H6 of pyrimidines (Figure 1A), which are located between 7.0 to 8.5 ppm in random coil DNA sequences, serve as good initial candidates for determining the sequence effect on aromatic proton chemical shifts. The DNA sequences SS1 to SS16 have the same sequence pattern as shown in Figure 1B and they were designed to avoid duplex or hairpin formation. In addition, these DNA samples were prepared

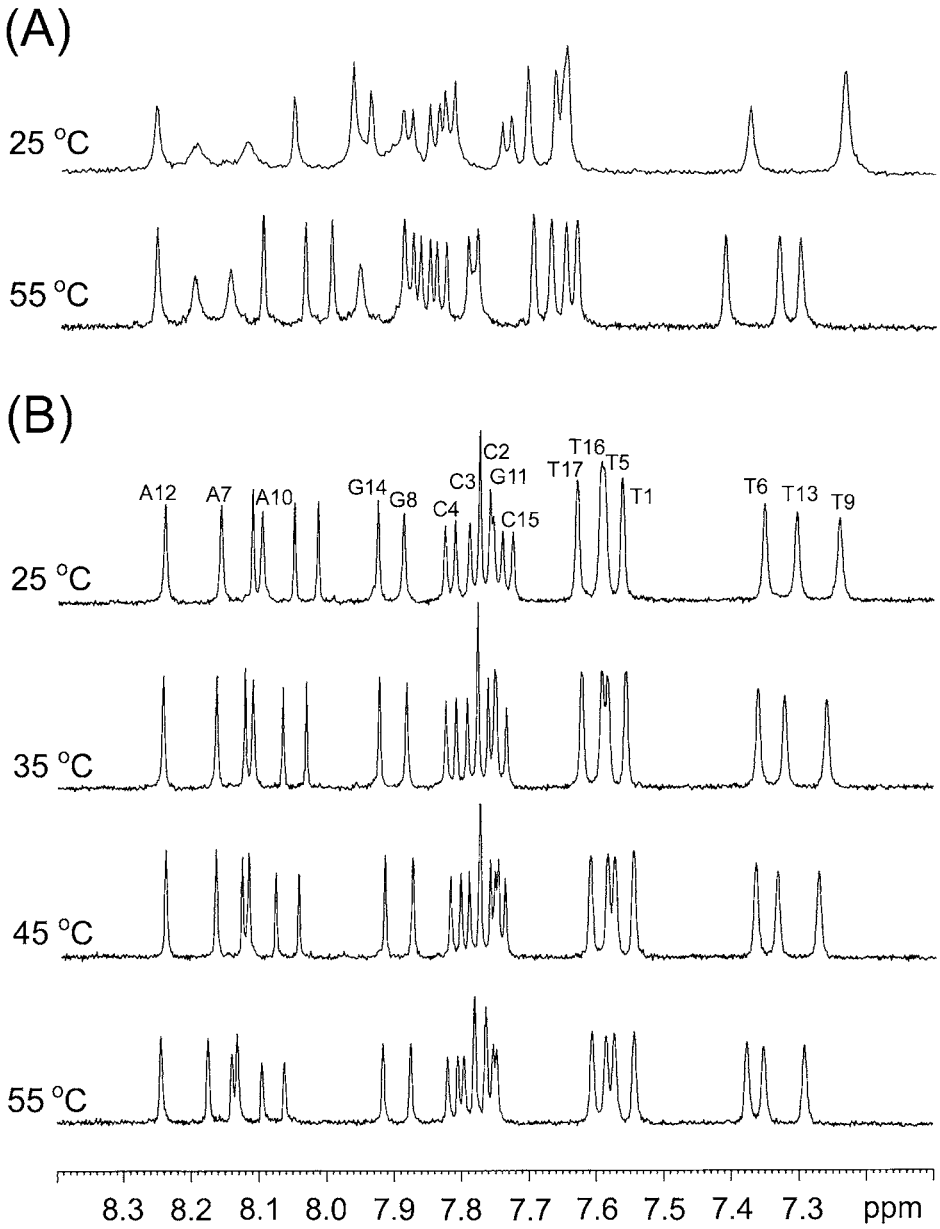

Figure 2. Aromatic region of the ^{1}H NMR spectra of SS1 in (A) water and (B) 8 M urea solution.

in 8 M urea solution to remove any residual structures. The ^{1}H spectrum of SS1 prepared in water has been acquired (Figure 2A) and compared with the one prepared in 8 M urea solution (Figure 2B). At 25 °C, considerable structures were found in the sample prepared in water. Most of these residual structures were removed when the temperature was increased to 55 °C. Therefore, preparing the DNA samples in 8 M urea is important in studying the random coil DNA chemical shifts.

To ensure a random coil state has been obtained for the DNA samples, variable temperature ^{1}H NMR spectra of all sequences were also acquired (Figure 2B). Within the temperature range from 25 to 55 °C, the chemical shift changes are in general less than 0.01 ppm per 10 °C, indicating that all 16 sequences are unstructured. For SS7 and SS16, the

Table 3. Nearest neighbor effect on DNA aromatic proton chemical shifts

Sequence name	Nucleotide position[a] 1[b]	2	3	4	5	$\Delta_1^{3'}$ (ppm)[c]	Sequence name	Nucleotide positon[a] 13	14	15	16[b]	17	$\Delta_1^{5'}$ (ppm)[d]
Group A							*Group E*						
SS1	T (7.567)	**C**	C	C	T	0.197	SS16	T	C	**C**	T (7.626)	T	0.150
SS2	T (7.431)	**G**	C	C	T	0.061	SS4	T	C	**G**	T (7.512)	T	0.036
SS3	T (7.370)	**A**	C	C	T	0.000	SS8	T	C	**A**	T (7.476)	T	0.000
SS4	T (7.597)	**T**	C	C	T	0.227	SS12	T	C	**T**	T (7.611)	T	0.135
Group B							*Group F*						
SS5	T (7.551)	**C**	G	C	T	0.193	SS1	T	G	**C**	T (7.598)	T	0.147
SS6	T (7.403)	**G**	G	C	T	0.045	SS5	T	G	**G**	T (7.497)	T	0.046
SS7	T (7.358)	**A**	G	C	T	0.000	SS9	T	G	**A**	T (7.451)	T	0.000
SS8	T (7.571)	**T**	G	C	T	0.213	SS13	T	G	**T**	T (7.592)	T	0.141
Group C							*Group G*						
SS9	T (7.523)	**C**	A	C	T	0.177	SS2	T	A	**C**	T (7.586)	T	0.152
SS10	T (7.394)	**G**	A	C	T	0.048	SS6	T	A	**G**	T (7.480)	T	0.046
SS11	T (7.346)	**A**	A	C	T	0.000	SS10	T	A	**A**	T (7.434)	T	0.000
SS12	T (7.556)	**T**	A	C	T	0.210	SS14	T	A	**T**	T (7.581)	T	0.147
Group D							*Group H*						
SS13	T (7.563)	**C**	T	C	T	0.171	SS3	T	T	**C**	T (7.618)	T	0.135
SS14	T (7.443)	**G**	T	C	T	0.051	SS7	T	T	**G**	T (7.526)	T	0.043
SS15	T (7.392)	**A**	T	C	T	0.000	SS11	T	T	**A**	T (7.483)	T	0.000
SS16	T (7.593)	**T**	T	C	T	0.197	SS15	T	T	**T**	T (7.616)	T	0.133

[a]The bolded nucleotides at position 2 and position 15 are the nearest neighbors on the 3′-end of T1 and on the 5′-end of T16, respectively.
[b]The aromatic proton chemical shifts of T1 and T16 (in ppm) are reported in parentheses.
[c]$\Delta_1^{3'}$ represents the nearest neighbor effect on T1. $\Delta_1^{3'} = \delta_{T1} - \delta_{T1A}$ where δ_{T1} is the chemical shift of T1 and δ_{T1A} is the T1 chemical shift with an adenine nucleotide as the nearest neighbor in the group.
[d]$\Delta_1^{5'}$ represents the nearest neighbor effect on T16. $\Delta_1^{5'} = \delta_{T16} - \delta_{T16A}$ where δ_{T16} is the chemical shift of T16 and δ_{T16A} is the T16 chemical shift with an adenine nucleotide as the nearest neighbor in the group.

chemical shifts were recorded at 35 °C because some of the aromatic peaks are broad at 25 °C but sharpened at 35 °C. The use of chemical shift data at 35 °C for these two sequences does not affect the following data analysis, as the changes in the aromatic proton chemical shifts over 10 °C were negligibly small (< 0.01 ppm/10 °C).

Among the 16 DNA sequences, each DNA contains four different types of $N_i^{5'}X_iN_i^{3'}$ triplets separated by thymine nucleotides. The central nucleotides X_i in different triplets experience the same type of next nearest neighbor effect from thymines. The neighboring effect from the third nearest neighbor is assumed to be negligibly small. To validate this, SS1-SS16 sequences were arranged into four groups such that nucleotides 12 to 17 (except nucleotide 14, which is the third nearest neighbor of T17) in each group are the same (Table 2). The T17 chemical shift with an adenine nucleotide as the third nearest neighbor (δ_{T17A}) in each group (usually the most upfield shift) was subtracted from the T17 chemical shift (δ_{T17}) of each sequence. The chemical shift difference ($\delta_{T17} - \delta_{T17A}$) indicates the significance of the third nearest neighbor effect (Δ_3). All Δ_3 values were determined to be less than 0.02 ppm with an average of 0.007 ppm, indicating that the third nearest neighbor effect is negligibly small.

The significance of the nearest neighbor effects from the 3′-end ($\Delta_1^{3'}$) and from the 5′-end ($\Delta_1^{5'}$) was investigated by comparing the T1 and T16 aromatic

Table 4. Next nearest neighbor effect on DNA aromatic proton chemical shifts

Sequence name	Nucleotide position[a] 1[b]	2	3	4	5	$\Delta_2^{3'}$ (ppm)[c]	Sequence name	Nucleotide positon[a] 13	14	15	16[b]	17	$\Delta_2^{5'}$ (ppm)[d]
Group A							*Group E*						
SS1	T (7.567)	C	**C**	C	T	0.044	SS16	T	C	**C**	T	T (7.633)	0.022
SS5	T (7.551)	C	**G**	C	T	0.028	SS4	T	C	**G**	T	T (7.623)	0.012
SS9	T (7.523)	C	**A**	C	T	0.000	SS8	T	C	**A**	T	T (7.611)	0.000
SS13	T (7.563)	C	**T**	C	T	0.040	SS12	T	C	**T**	T	T (7.644)	0.033
Group B							*Group F*						
SS2	T (7.431)	G	**C**	C	T	0.037	SS1	T	G	**C**	T	T (7.634)	0.031
SS6	T (7.403)	G	**G**	C	T	0.009	SS5	T	G	**G**	T	T (7.617)	0.014
SS10	T (7.394)	G	**A**	C	T	0.000	SS9	T	G	**A**	T	T (7.603)	0.000
SS14	T (7.443)	G	**T**	C	T	0.049	SS13	T	G	**T**	T	T (7.637)	0.034
Group C							*Group G*						
SS3	T (7.370)	A	**C**	C	T	0.024	SS2	T	A	**C**	T	T (7.630)	0.031
SS7	T (7.358)	A	**G**	C	T	0.012	SS6	T	A	**G**	T	T (7.612)	0.013
SS11	T (7.346)	A	**A**	C	T	0.000	SS10	T	A	**A**	T	T (7.599)	0.000
SS15	T (7.392)	A	**T**	C	T	0.046	SS14	T	A	**T**	T	T (7.632)	0.033
Group D							*Group H*						
SS4	T (7.597)	T	**C**	C	T	0.041	SS3	T	T	**C**	T	T (7.643)	0.031
SS8	T (7.571)	T	**G**	C	T	0.015	SS7	T	T	**G**	T	T (7.615)	0.003
SS12	T (7.556)	T	**A**	C	T	0.000	SS11	T	T	**A**	T	T (7.612)	0.000
SS16	T (7.589)	T	**T**	C	T	0.033	SS15	T	T	**T**	T	T (7.646)	0.034

[a]The bolded nucleotides at position 3 and position 15 are the nearest neighbors on the 3′-end of T1 and on the 5′-end of T17 respectively.
[b]The aromatic proton chemical shifts of T1 and T17 (in ppm) are reported in parentheses.
[c] $\Delta_2^{3'}$ represents the next nearest neighbor effect on T1. $\Delta_2^{3'} = \delta_{T1} - \delta_{T1A}$ where δ_{T1} is the chemical shift of T1 and δ_{T1A} is the T1 chemical shift with an adenine nucleotide as the next nearest neighbor in the group.
[d] $\Delta_2^{5'}$ represents the next nearest neighbor effect on T17. $\Delta_2^{5'} = \delta_{T17} - \delta_{T17A}$ where δ_{T17} is the chemical shift of T17 and δ_{T17A} is the T17 chemical shift with an adenine nucleotide as the next nearest neighbor in the group.

proton chemical shifts, respectively (Table 3). The most upfield T1 and T16 chemical shifts were usually found in cases where adenine nucleotides were positioned as the 3′- and the 5′-nearest neighbors, respectively. The average 3′-nearest neighbor effects for C, G and T on substituting A were found to be 0.185, 0.051 and 0.212 ppm, respectively and the magnitude of shielding effect from the 3′-nearest neighbors follows the order A > G > C > T. For the 5′-nearest neighbors, the average effects for C, G and T on substituting A were found to be 0.146, 0.043, and 0.139 ppm, respectively and the magnitude of shielding effect from the 5′-nearest neighbors follows the order A > G > C ∼ T. Comparing with the 3′- nearest neighbor effect, the 5′-effect appears to be less significant. In addition, purine-type nearest neighbors usually cause a more upfield aromatic proton chemical shift than pyrimidine-type nearest neighbors.

For the 3′-next nearest neighbor effect ($\Delta_2^{3'}$) and the 5′-next nearest neighbor effect ($\Delta_2^{5'}$), the aromatic proton chemical shifts of T1 and T17 of the 16 DNA sequences were compared, respectively (Table 4). The largest next nearest neighbor effect was found to be 0.049 ppm. Although this value is already an order of magnitude less than the largest value determined for the nearest neighbor effect, the next nearest neighbor effect remains significant in affecting aromatic proton chemical shifts. The average 3′-next nearest neighbor effects for C, G and T on substituting A were found to be 0.037, 0.016, and 0.042 ppm, respectively whereas

the corresponding average values for the 5′-next nearest neighbor effects were 0.029, 0.011 and 0.033 ppm, respectively. Unlike the nearest neighbor effect, the next nearest neighbor effects from both the 3′- and 5′-ends are similar and the magnitude of shielding effect follows the order A > G > C ∼ T.

The above analysis on neighboring effects indicates that the aromatic proton chemical shift of a specific nucleotide in a DNA sequence is affected not only by its nearest neighbors but also its next nearest neighbors on both of its 3′- and 5′-ends. To make use of the measured aromatic proton chemical shifts of the central nucleotide X_i ($\delta_{triplet}$) in all 64 types of $N_i^{5'}X_iN_i^{3'}$ triplets (Table 5) to predict random coil DNA aromatic proton chemical shifts, the next nearest neighbor effects from the next nearest thymine nucleotides on the 5′-end ($\Delta_2^{5'T}$) and from the 3′-end ($\Delta_2^{3'T}$) have to be subtracted and the 'real' next nearest neighbor effects from the 5′-end ($\Delta_2^{5'N}$) and from the 3′-end ($\Delta_2^{3'N}$) have to be added, i.e.,

$$\delta_{pred} = \delta_{triplet} - \Delta_2^{5'T} - \Delta_2^{3'T} + \Delta_2^{5'N} + \Delta_2^{3'N.}$$

Each type of 'real' next nearest neighbor effect has been calculated by averaging the 5′- and 3′-next neighbor effects (Table 6). Thus, to predict the random coil H8 chemical shift of A3 in 5′-G1-T2-**A3**-G4-C5-T6-A7-G8-G9-T10-G11-3′, the $\delta_{triplet}$ value of TAG (8.158 ppm) in Table 5 and the next nearest neighbor values of T (0.038), G (0.013) and C (0.033) in Table 6 are need and the predicted H8 shift equals to: $8.158 - 0.038 - 0.038 + 0.013 + 0.033 = 8.128$ ppm.

To test the accuracy and reliability of this prediction protocol, a total of 198 aromatic proton chemical shifts from 22 random coil DNA sequences containing 9 to 17 nucleotides were measured. The predicted chemical shift values (δ_{pred}) were all in excellent agreement with the experimental values (δ_{expt}) (Figure 3A). The Pearson correlation coefficient of the plot of was 0.998 and the RMSD value between the predicted and the experimental values was 0.02 ppm. These results are superior to the results predicted by the RNA-derived neighboring parameter procedure (Figure 3B) with a correlation coefficient of 0.981 and a RMSD value of 0.10 ppm (Bell et al., 1983, 1985; Hader et al., 1982).

The protocol was also used to predict the aromatic proton chemical shifts of a denatured self-complementary duplex (Tjandra et al., 2000) and a denatured 15-nucloetide DNA hairpin with a T-T-T loop (Kuklenyik et al., 1996). At 25 °C, the RMSD values between the predicted random coil chemical shifts and the measured values for the hairpin and for the duplex were 0.194 ppm and 0.151 ppm, respectively. Raising the experimental temperature to 85 °C, the RMSD was found to decrease to 0.023 ppm for the duplex and 0.047 ppm for the hairpin. At 95 °C, the RMSD value of the hairpin sample was further reduced to 0.023 ppm.

To extend the above prediction protocol to DNA sugar protons, the H1′, H2′, H2″ and H3′ sugar proton chemical shifts of the central nucleotide X_i ($\delta_{triplet}$) in all 64 types of $N_i^{5'}X_iN_i^{3'}$ triplets were also measured (Table 7). Similarly, the correction factors for the next nearest neighbor effects have also been determined following the above neighboring effect analysis method (Table 8). In order to test the accuracy and reliability of the prediction protocol on the sugar proton chemical shifts, ∼200 experimentally measured H1′,

Table 5. Aromatic proton chemical shifts[a] (ppm) of X_i in Triplet $N_i^{5'}X_iN_i^{3'}$

C_i	$\delta_{triplet}$	G_i	$\delta_{triplet}$	A_i	$\delta_{triplet}$	T_i	$\delta_{triplet}$
CCC	7.783	**CGC**	7.968	**CAC**	8.340	**CTC**	7.637
GCC	7.710	**GGC**	7.935	**GAC**	8.278	**GTC**	7.510
ACC	7.656	**AGC**	7.888	**AAC**	8.237	**ATC**	7.456
TCC	7.765	**TGC**	7.964	**TAC**	8.327	**TTC**	7.608
CCG	7.640	**CGG**	7.786	**CAG**	8.157	**CTG**	7.464
GCG	7.579	**GGG**	7.773	**GAG**	8.105	**GTG**	7.379
ACG	7.549	**AGG**	7.754	**AAG**	8.090	**ATG**	7.337
TCG	7.633	**TGG**	7.766	**TAG**	8.158	**TTG**	7.453
CCA	7.589	**CGA**	7.798	**CAA**	8.144	**CTA**	7.402
GCA	7.525	**GGA**	7.776	**GAA**	8.088	**GTA**	7.312
ACA	7.480	**AGA**	7.756	**AAA**	8.066	**ATA**	7.274
TCA	7.578	**TGA**	7.805	**TAA**	8.144	**TTA**	7.400
CCT	7.808	**CGT**	7.935	**CAT**	8.349	**CTT**	7.621
GCT	7.733	**GGT**	7.920	**GAT**	8.295	**GTT**	7.521
ACT	7.703	**AGT**	7.904	**AAT**	8.268	**ATT**	7.510
TCT	7.802	**TGT**	7.947	**TAT**	8.350	**TTT**	7.616

[a]Uncertainty in chemical shift measurement was ± 0.002 ppm.

Table 6. Correction factors for next nearest neighbor effects of aromatic protons

Next nearest neighbor	$\Delta_2^{5'N}$ or $\Delta_2^{3'N}$ (ppm)
C	0.033
G	0.013
A	0.000
T	0.038

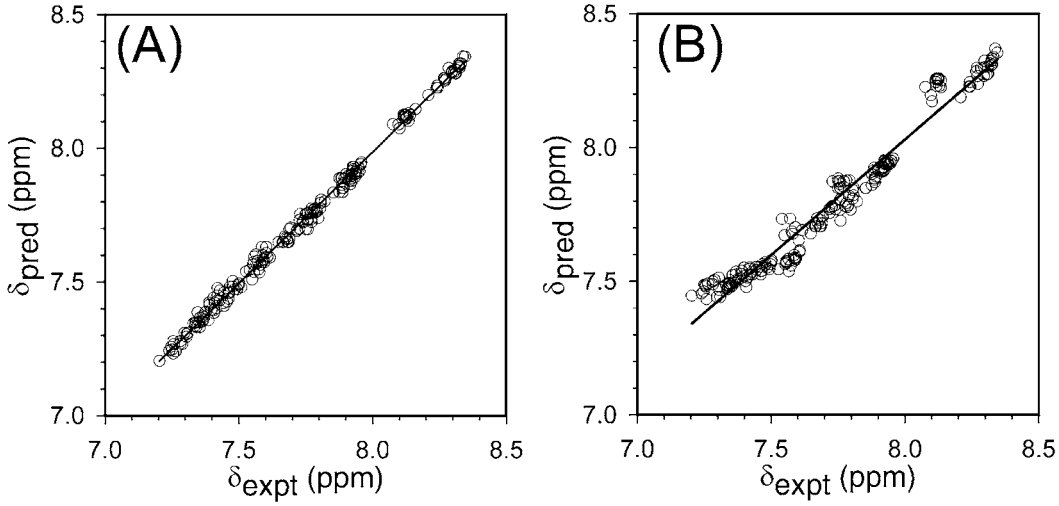

Figure 3. (A) A plot of the predicted random coil chemical shifts against the experimental values of ∼200 aromatic protons from 22 random coil DNA sequences. (B) A plot of the predicted chemical shifts using the RNA-derived neighboring parameters against the experimental values.

Table 7. Sugar proton chemical shifts[a] of X_i in triplet $N_i^{5'}X_iN_i^{3'}$

	$\delta_{triplet}$ (ppm)																		
C_i	H1′	H2′	H2″	H3′	G_i	H1′	H2′	H2″	H3′	A_i	H1′	H2′	H2″	H3′	T_i	H1′	H2′	H2″	H3′
CCC	6.269	2.252	2.519	4.833	**CGC**	6.152	2.653	2.755	4.974	**CAC**	6.375	2.764	2.824	4.994	**CTC**	6.240	2.266	2.476	4.866
GCC	6.266	2.243	2.514	4.835	**GGC**	6.086	2.629	2.708	4.986	**GAC**	6.346	2.735	2.788	5.016	**GTC**	6.238	2.280	2.468	4.860
ACC	6.218	2.213	2.477	4.825	AGC	5.996	2.601	2.661	4.975	**AAC**	6.248	2.686	2.726	4.999	**ATC**	6.181	2.243	2.433	4.847
TCC	6.268	2.256	2.513	4.868	**TGC**	6.143	2.660	2.752	4.975	**TAC**	6.360	2.735	2.812	4.994	**TTC**	6.256	2.372	2.526	4.872
CCG	6.140	2.000	2.399	4.774	**CGG**	5.918	2.537	2.537	4.898	**CAG**	6.155	2.592	2.665	4.948	**CTG**	6.113	2.051	2.368	4.796
GCG	6.107	1.982	2.394	4.762	**GGG**	5.858	2.524	2.524	4.923	**GAG**	6.125	2.562	2.645	4.967	**GTG**	6.106	2.035	2.358	4.810
ACG	6.096	1.965	2.374	4.772	**AGG**	5.762	2.480	2.480	4.920	**AAG**	6.053	2.530	2.593	4.954	**ATG**	6.050	1.979	2.302	4.778
TCG	6.145	2.000	2.405	4.782	**TGG**	5.891	2.521	2.521	4.894	**TAG**	6.125	2.597	2.643	4.944	**TTG**	6.111	2.039	2.360	4.797
CCA	6.105	1.879	2.323	4.769	**CGA**	5.776	2.496	2.496	4.899	**CAA**	6.083	2.559	2.662	4.944	**CTA**	6.059	1.927	2.294	4.782
GCA	6.084	1.864	2.332	4.768	**GGA**	5.754	2.489	2.489	4.917	**GAA**	6.056	2.540	2.638	4.960	**GTA**	6.043	1.912	2.289	4.800
ACA	6.023	1.840	2.295	4.764	**AGA**	5.588	2.445	2.445	4.912	**AAA**	5.945	2.492	2.578	4.958	**ATA**	5.987	1.864	2.238	4.781
TCA	6.101	1.881	2.322	4.766	**TGA**	5.759	2.494	2.494	4.900	**TAA**	6.050	2.559	2.632	4.947	**TTA**	6.047	1.917	2.288	4.790
CCT	6.266	2.269	2.539	4.872	**CGT**	6.152	2.688	2.814	4.992	**CAT**	6.386	2.823	2.823	5.018	**CTT**	6.241	2.316	2.518	4.880
GCT	6.244	2.257	2.533	4.861	**GGT**	6.093	2.656	2.754	4.998	**GAT**	6.357	2.753	2.753	5.038	**GTT**	6.222	2.305	2.492	4.890
ACT	6.216	2.258	2.514	4.860	**AGT**	6.006	2.621	2.718	5.003	**AAT**	6.274	2.750	2.750	5.033	**ATT**	6.178	2.279	2.479	4.880
TCT	6.256	2.275	2.531	4.865	**TGT**	6.158	2.688	2.805	4.993	**TAT**	6.377	2.790	2.847	5.016	**TTT**	6.244	2.331	2.506	4.889

[a]Uncertainty in chemical shift measurement was ± 0.002 ppm.

H2′, H2″ and H3′ chemical shifts from 22 random coil DNA sequences were compared with the predicted values. For the H1′ proton, the RMSD value was found to be 0.02 ppm with a correlation coefficient of 0.988, indicating a significant improvement in the accuracy of chemical shift prediction has been achieved when compared with the RMSD value of 0.16 ppm and the correlation coefficient of 0.813 obtained from the RNA-derived neighboring parameter procedure (Bell et al., 1983, 1985; Hader et al., 1982).

For the H2′ and H2″ protons, the correction factors for the next nearest neighbor effects were found to be similar to those of the H6/H8 protons and the H1′ proton. A RMSD value of 0.03 ppm was obtained between the prediction results and the experimental values of the H2′ and H2″ protons. Excellent corre-

Table 8. Correction factors for next nearest neighbor effects of sugar protons

Next nearest neighbor	$\Delta_2^{5'N}$ or $\Delta_2^{3'N}$ (ppm)			
	H1′	H2′	H2″	H3′
C	0.033	0.039	0.042	0.019
G	0.010	0.013	0.011	0.004
A	0.000	0.000	0.000	0.000
T	0.037	0.040	0.034	0.019

lation coefficients of 0.995 and 0.988 have also been obtained for the H2′ and H2″ protons, respectively.

For the H3′ proton, the correction factors for the next nearest neighbor effects were found to be about 50% smaller than the factors for the H6/H8, H1′, H2′ and H2″ protons. Nevertheless, pyrimidine-type correction factors remain greater than purine-type correction factors. The accuracy of the prediction protocol based on the extracted H3′ triplet chemical shift values and these next nearest neighbor effect correction factors has also been assessed. The RMSD value between the predicted H3′ shifts and the measured H3′ shifts was found to be 0.02 ppm with a correlation coefficient of 0.983, indicating the prediction results on the H3′ chemical shifts remain accurate and reliable. Due to the lack of appropriate parameters for the H2′, H2″ and H3′ protons in literature, no comparison with the results obtained from the RNA-derived neighboring parameter procedure has been made.

Conclusions

From studying the sequence effect on random coil DNA chemical shifts, the nearest neighbor and the next nearest neighbor effects have been found to play a significant role in affecting the chemical shifts of a specific nucleotide. Therefore, these effects have to be considered in developing chemical shift prediction method. In this study, a random coil DNA chemical shift parameter set which makes use of the measured proton chemical shifts of the central nucleotide X_i in all 64 types of $N_i^{5'}X_iN_i^{3'}$ triplets and the correction factors for the next nearest neighbor effect has been established and a protocol has been developed for predicting random coil DNA proton chemical shifts. The prediction results on H6/H8, H1′, H2′, H2″ and H3′ chemical shifts have been demonstrated to be highly reliable and accurate.

Acknowledgements

The work described in this paper was partially supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No.: CUHK4255/01P) and a direct grant from CUHK Research Committee Funding (Project Code: 4450006).

References

Altona, C., Faber, D.H. and Westra Hoekzema, A.J.A. (2000) *Magn. Reson. Chem.*, **38**, 95–107.

Bax, A. and Davis, D.G. (1985) *J. Magn. Reson.*, **63**, 207–213.

Bell, R.A., Alkema, D., Coddington, J.M., Hader, P.A., Hughes, D.W. and Neilson, T. (1983) *Nucl. Acids Res.*, **11**, 1143–1149.

Bell, R.A., Everett, J.R., Hughes, D.W., Coddington, J.M., Alkema, D., Hader, P.A. and Neilson, T. (1985) *J. Biomol. Struct. Dyn.*, **2**, 693–707.

Bundi, A. and Wüthrich, K. (1979) *Biopolymers*, **18**, 285–297.

Case, D.A., Dyson, H.J. and Wright, P.E. (1994) *Meth. Enzymol.*, **239**, 392–416.

Cromsigt, J.A., Hilbers, C.W. and Wijmenga, S.S. (2001) *J. Biomol. NMR*, **21**, 11–29.

Dejaegere, A., Bryce, R.A. and Case, D.A. (1999) In *Modeling NMR Chemical Shifts. Gaining Insight into Structure and Environment*, Facelli, J.C. and de Dios, A.C. (Eds.), American Chemical Society, Washington, DC, pp. 194–206.

Hader, P.A., Alkema, D., Bell, R.A. and Neilson, T. (1982) *J. Chem. Soc. Chem. Commun.*, 10–12.

Kuklenyik, Z., Yao, S. and Marzilli, L.G. (1996) *Eur. J. Biochem.*, **236**, 960–969.

Marion, D. and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.

Merutka, G., Dyson, H.J. and Wright, P.E. (1995) *J. Biomol. NMR*, **5**, 14–24.

Szilagyi, L. (1995) *Prog. Nucl. Magn. Reson. Spectrosc.*, **27**, 325–443.

Tjandra, N., Tate, S., One, A., Kainosho, M. and Bax, A. (2000) *J. Am. Chem. Soc.*, **122**, 6190–6200.

Wijmenga, S.S., Kruithof, M. and Hilbers, C.W. (1997) *J Biomol. NMR*, **10**, 337–350.

Williamson, M.P. and Asakura, T. (1997) *Methods Mol. Biol.*, **60**, 53–59.

Wishart, D.S. and Sykes, B.D. (1994) *Methods Enzymol.*, **239**, 363–392.

Wishart, D.S., Bigam, C.G., Holm, A., Hodges, R.S. and Sykes, B.D. (1995) *J. Biomol. NMR*, **5**, 67–81.

Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.