Available online at www.sciencedirect.com
ScienceDirect
Journal of Magnetic Resonance 187 (2007) 105–111
JMR Journal of Magnetic Resonance
www.elsevier.com/locate/jmr

Proton chemical shift prediction of A·A mismatches in B-DNA duplexes

Sik Lok Lam *, Kin Fung Lai, Lai Man Chi

Department of Chemistry, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

Received 10 February 2007; revised 29 March 2007
Available online 19 April 2007

Abstract

A proton chemical shift prediction method has been developed for double helical DNAs containing A·A mismatches. This method makes use of the chemical shift prediction scheme for normal B-DNA duplexes developed by Altona and co-workers and a set of A·A mismatch triplet chemical shift values and corrections factors extracted from reference sequences. The triplet values are used for predicting chemical shifts of A·A mismatches whereas the normal B-DNA chemical shifts and correction factors are used for the flanking residues of A·A mismatches. Both 5′- and 3′-correction factors have been determined from the chemical shift differences upon replacing the A·A mismatch in a duplex with an A·T base pair. Based on 560 sets of predicted and experimental chemical shifts, the overall prediction accuracy for various types of protons has been determined to be 0.07 ppm with an excellent correlation coefficient of 0.9996.
© 2007 Elsevier Inc. All rights reserved.

Keywords: Proton chemical shift; Prediction; DNA; A·A mismatch

1. Introduction

Chemical shift methods have been used extensively in determining secondary structures of proteins [1–4]. However, the situation is much more complex in nucleic acids and much more structure–chemical shift information is needed to advance this area. Prediction of nucleic acid chemical shifts provide quick reference guide for resonance assignments based on conventional NOESY and COSY type experiments, thus facilitating solution structure determination. In addition, the prediction results can provide useful information for studying structure–chemical shift relationship, identifying unstructured or right-handed double helical regions, monitoring DNA–drug or DNA–protein binding, and investigating conformational details of special features in DNA structures. At present, several methods have been established to predict chemical shifts of random coil DNAs [5–7], double helical B-DNAs [8–10] and RNAs [11]. Briefly, chemical shifts can be predicted from structures using the same types of electrostatic and ring-current models that have been applied to proteins [8,9,11] or from measured chemical shifts within a set of given sequences that adopt stable and well-defined conformations [10]. The former approach is usually preferred because it relates chemical shifts directly to conformation. For the latter approach, no detailed information on conformational differences can be obtained once chemical shift difference is identified between predicted and experimental values. However, this approach has been shown to be more precise and accurate for predicting DNA chemical shifts [10]. A hybrid approach of the above prediction methods have also been applied to DNAs [9] and RNAs [11] and it has been found that sequence triplets define proton chemical shifts of the middle residue quite precisely [10,11]. In order to widen the applications of DNA chemical shifts, this work aims to develop chemical shift prediction methods for DNA duplexes containing mismatches.

DNA mismatches can occur *in vivo* due to misincorporation of bases [12] or strand misalignment during replication [13], heteroduplex formation during homologous recombination [14], spontaneous deamination [15], damage by

* Corresponding author. Fax: +852 2603 5057.
E-mail address: lams@cuhk.edu.hk (S.L. Lam).

1090-7807/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmr.2007.04.005

mutagenic chemicals or ionizing radiation [16–19]. Recently, mismatches in triplet repeat structures have been hypothesized to be the origin of genetic instabilities that lead to DNA mutations [20–25]. During DNA replication, repair and recombination, slippage in single strands of CAG and CTG repeats can occur, forming slipped strand structures with A·A and T·T mismatches, respectively [26–29]. Therefore, structural information about these mismatches is useful for understanding the triplet repeat expansion process. To facilitate structural studies of mismatches, an initial attempt has been carried out to establish a chemical shift prediction method for A·A mismatches in DNA duplexes.

Fig. 1a shows an A·A mismatch containing DNA duplex in which W and W′, X and X′, Y and Y′, Z and Z′ are complementary Watson–Crick base pairs. W, X, Y and Z can be any one of four bases. This mismatch sequence is similar to the duplex sequences in Fig. 1b, namely duplexA·T (XAY) and duplexT·A(XAY), which contain an A·T and T·A base pair instead of the A·A mismatch, respectively. Since only the nearest neighbor effect has been considered to be important in predicting proton chemical shift in B-DNA duplexes [9,10], it is expected that the effect of replacing A·T or T·A with A·A on the next nearest neighbor chemical shifts will be negligibly small. Therefore, the present work will focus on predicting the chemical shifts of A·A mismatch and its flanking residues in B-DNA duplexes.

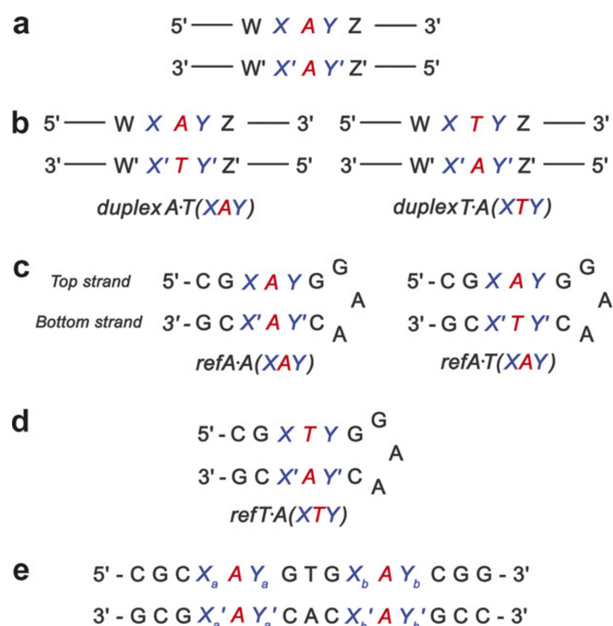

Fig. 1. (a) A general double helical B-DNA sequence containing a single A·A mismatch. (b) Normal B-DNA duplex sequences corresponding to the sequence in (a) in which the A·A mismatch has been replaced by an A·T or T·A base pair. (c) Reference A·A and A·T hairpin sequences used to derive 5′- and 3′-A·A correction factors. The nomenclature of sequences is in accord with the top strand triplets. (d) Reference T·A hairpin sequence used for verifying the applicability of the proposed prediction method. (e) Testing sequences used for determining the prediction accuracy.

2. Experimental

2.1. Sample design

Fig. 1c shows the design of two types of reference samples which are required for determining A·A mismatch triplet chemical shifts and correction factors. They are all 15-nucleotide DNA hairpins containing a 5′-GAA loop which connects the top and bottom strands in order to simplify the sample preparative work. In each type of the reference samples, there are 16 sequences containing different XAY triplets. The nomenclature of these sequences is in accord with the triplets in their top strands. The first type, refA·A(XAY), contains an A·A mismatch in the stem region. The "XAY" in parenthesis represents the triplet that contains the mismatched A in the top strand of the reference sample. The second type, refA·T(XAY), corresponds to refA·A(XAY) in which the A·A mismatch has been replaced by an A·T base pair. In this case, the "XAY" in parenthesis represents the triplet that contains the A of the A·T base pair.

Fig. 1d shows an additional type of reference sequences, refT·A(XTY), which is used for verifying the applicability of the base pair replacement approach used in the proposed prediction method. These sequences correspond to refA·A(XAY) in which the A·A mismatch has been replaced by an T·A base pair.

For testing the prediction accuracy of the proposed method, Fig. 1e shows the design of eight double helical B-DNA testing sequences. The exact sequences have been listed in Supplementary material S1. Each strand of these duplexes contains 15 nucleotides with two A.A mismatches separated by 5 nucleotides. Thereby, chemical shifts of a total of 32 XAY triplets can be used for testing purpose.

2.2. DNA samples

All DNA samples were synthesized using an Applied Biosystems Model 392 DNA synthesizer and purified using denaturing polyacrylamide gel electrophoresis and diethylaminoethyl Sephacel anion exchange chromatography. Centricon-3 concentrators were used in the last purification step to remove the high salt contents from the samples. DNA quantities were determined using UV absorbance data at 260 nm. NMR samples were prepared by dissolving 0.5 μmole purified DNA samples into 500 μl buffer solution containing 150 mM sodium chloride, 10 mM sodium phosphate at pH 7 and 0.1 mM 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS).

2.3. NMR measurements

All NMR experiments were performed on a Bruker ARX-500 or AV-500 spectrometer operating at 500.13 MHz and acquired at 25 °C unless stated otherwise. For sequential assignments of non-labile protons, the solvent was exchanged to 99.96% D_2O and conventional two

dimensional homonuclear nuclear Overhauser effect spectroscopy (NOESY), double-quantum-filtered correlation spectroscopy (DQF-COSY) and total correlation spectroscopy (TOCSY) were performed with a 2-s presaturation pulse to suppress the residual HDO signal. NOESY and TOCSY were acquired at 300 and 75 ms mixing time, respectively. A total of 512 free induction decays, each consisting of 4096 complex data points were collected. The most upfield signal of DSS was set at 0 ppm to serve as an internal chemical shift reference.

In case of ambiguous proton resonance assignments, ^1H–^{13}C heteronuclear single-quantum correlation (HSQC) [30,31] and heteronuclear multiple bond correlation (HMBC) [32] experiments were performed. Heteronuclear decoupling in HSQC was executed by the GARP-1 sequence [33]. Carbon chemical shifts were indirectly referenced to DSS using the derived nucleus-specific ratio (Ξ) of 0.251449530 [34]. All NMR data were processed using Bruker Topspin 1.3 software.

3. Results and discussion

Chemical shifts of non-labile base protons, including H6/H8, adenine H2 (A-H2), cytosine H5 (C-H5) and thymine methyl H7 (T-H7) as well as sugar protons, including H1′, H2′, H2″ and H3′, of all reference and testing sequences have been extracted. As the chemical shifts of H4′, H5′ and H5″ usually overlap seriously, the analysis of these protons have been excluded to minimize the uncertainty contributions due to ambiguous assignments to the prediction method.

3.1. Chemical shift prediction of A·A mismatch

Table 1 shows the chemical shifts of mismatched A in all different types of XAY triplets extracted from refA·A(XAY) sequences. Based on the nearest neighbor model, the predicted chemical shifts of mismatched A in B-DNA duplexes will be equal to the extracted chemical shifts of mismatched A from the corresponding refA·A(XAY) sequences, i.e.

$$\delta_{\text{pred}}(\text{A}) = \delta_{\text{refA·A(XAY)}}(\text{A}) \quad (1)$$

Using the extracted triplet values in Table 1, a total of 192 predicted chemical shifts have been obtained for various types of protons in eight testing sequences (Fig. 1e). By comparing with the experimental results, the root-mean-square deviation (RMSD) value has been determined to be 0.06 ppm with a correlation coefficient of 0.9998.

Table 1
Chemical shifts of A in XAY triplets extracted from the top strands of refA·A(XAY)

XAY triplet	$\delta_{\text{refA·A(XAY)}}$(A), ppm					
	H8	H1′	H2′	H2″	H3′	H2
AAA	8.046	5.730	2.601	2.601	4.977	7.673
AAC	8.170	6.105	2.738	2.771	5.032	8.184
AAG	8.034	5.815	2.580	2.580	4.961	7.879
AAT	8.087	6.072	2.697	2.785	5.015	8.042
CAA	8.116	5.515	2.518	2.518	4.896	7.512
CAC	8.250	6.209	2.839	2.839	5.017	8.172
CAG	8.089	5.613	2.497	2.567	4.864	7.630
CAT	8.200	6.175	2.785	2.785	4.998	7.870
GAA	8.147	5.840	2.768	2.768	5.028	7.874
GAC	8.254	6.237	2.861	2.861	5.057	8.348
GAG	8.135	5.926	2.741	2.741	5.013	8.114
GAT	8.216	6.205	2.863	2.863	5.045	8.218
TAA	8.109	5.647	2.570	2.570	4.921	7.602
TAC	8.253	6.226	2.868	2.868	5.026	8.218
TAG	8.090	5.705	2.538	2.538	4.895	7.696
TAT	8.171	6.166	2.828	2.828	5.010	7.967

3.2. Chemical shift prediction of 5′-flanking residue of A·A mismatch

For predicting the chemical shift of X, which locates at the 5′-side of A·A mismatch, it is necessary to use the chemical shift of X in the corresponding normal duplex in Fig. 1b and a correction factor which accounts for the change in chemical shift of X due to replacement of the A·A mismatch with an A·T base pair, i.e.

$$\delta_{\text{pred}}(\text{X}) = \delta_{\text{duplexA·T(XAY)}}(\text{X}) + 5'\text{-}\Delta_{\text{A·A/A·T(XAY)}}(\text{X}) \quad (2)$$

where $\delta_{\text{duplexA·T(XAY)}}$(X) is the chemical shift of X in the corresponding duplexA·T(XAY) sequence and $5'\text{-}\Delta_{\text{A·A/A·T(XAY)}}$(X) is the 5′-A·A correction factor. $\delta_{\text{duplexA·T(XAY)}}$(X) can be predicted from the chemical shifts in the database of sequence triplets in B-DNA derived by Altona and co-workers [10] whereas the 5′-correction factor can be determined using the chemical shifts of X extracted from refA·A(XAY) and refA.T(XAY), i.e.

$$5'\text{-}\Delta_{\text{A·A/A·T(XAY)}}(\text{X}) = \delta_{\text{refA·A(XAY)}}(\text{X}) - \delta_{\text{refA·T(XAY)}}(\text{X}) \quad (3)$$

Using these correction factors as shown in Table 2, the RMSD value using 184 sets of testing data has been found to be 0.08 ppm with a correlation coefficient of 0.9994.

3.3. Chemical shift prediction of 3′-flanking residue of A·A mismatch

Similarly, for predicting the chemical shift of Y, which locates at the 3′-side of A·A mismatch, the following equation can be used:

$$\delta_{\text{pred}}(\text{Y}) = \delta_{\text{duplexA·T(XAY)}}(\text{Y}) + 3'\text{-}\Delta_{\text{A·A/A·T(XAY)}}(\text{Y}) \quad (4)$$

where $\delta_{\text{duplexA·T(XAY)}}$(Y) is the chemical shift of Y in the corresponding duplexA·T(XAY) sequence and can be predicted using Altona method [10]. $3'\text{-}\Delta_{\text{A· A/A·T(XAY)}}$(Y) is the 3′-A·A correction factor, indicating the change in chemical shift of Y upon replacing A·A with an A·T base pair. It can be determined from the chemical shift difference of Y between refA·A(XAY) and refA·T(XAY), i.e.

$$3'\text{-}\Delta_{\text{A·A/A·T(XAY)}}(\text{Y}) = \delta_{\text{refA·A(XAY)}}(\text{Y}) - \delta_{\text{refA·T(XAY)}}(\text{Y}) \quad (5)$$

These correction factors have been summarized in Table 3. Based on 184 sets of testing data, the RMSD value has

Table 2
5′-A·A correction factors extracted from the top strands of refA·A(XAY) and refA·T(XAY)

XAY triplet	$5'$-$\Delta_{A\cdot A/\ A\cdot T(XAY)}$(X), ppm[a]							
	H6/H8	H1′	H2′	H2″	H3′	A-H2	C-H5	T-H7
AAA	−0.137	−0.103	−0.211	−0.260	−0.091	0.276	—	—
AAC	−0.114	−0.239	−0.202	−0.223	−0.060	0.185	—	—
AAG	−0.146	−0.071	−0.231	−0.279	−0.097	0.262	—	—
AAT	−0.140	−0.156	−0.162	−0.095	−0.051	0.221	—	—
CAA	−0.002	0.382	−0.095	−0.196	−0.105	—	0.045	—
CAC	−0.083	−0.026	−0.391	−0.315	−0.079	—	0.004	—
CAG	−0.014	0.419	−0.079	−0.168	−0.118	—	0.030	—
CAT	−0.105	0.016	−0.391	−0.255	−0.091	—	−0.018	—
GAA	−0.093	−0.199	−0.219	−0.339	−0.090	—	—	—
GAC	−0.084	−0.378	−0.163	−0.286	−0.075	—	—	—
GAG	−0.097	−0.168	−0.245	−0.313	−0.084	—	—	—
GAT	−0.115	−0.327	−0.194	−0.175	−0.059	—	—	—
TAA	−0.039	0.277	−0.217	−0.245	−0.107	—	—	0.038
TAC	−0.122	−0.036	−0.438	−0.339	−0.088	—	—	−0.008
TAG	−0.072	0.353	−0.189	−0.212	−0.103	—	—	0.021
TAT	−0.158	0.011	−0.427	−0.232	−0.092	—	—	−0.018

[a] $5'$-$\Delta_{A\cdot A/A\cdot T(XAY)}(X) = \delta_{refA\cdot A(XAY)}(X) - \delta_{refA\cdot T(XAY)}(X)$.

Table 3
3′-A·A correction factors extracted from the top strands of refA·A(XAY) and refA·T(XAY)

XAY triplet	$3'$-$\Delta_{A\cdot A/\ A\cdot T(XAY)}$(Y), ppm[a]							
	H6/H8	H1′	H2′	H2″	H3′	A-H2	C-H5	T-H7
AAA	0.096	−0.005	0.123	−0.005	−0.004	−0.042	—	—
AAC	0.019	0.018	−0.089	−0.022	−0.018	—	−0.037	—
AAG	0.199	0.025	0.170	0.060	0.010	—	—	—
AAT	0.036	0.055	−0.026	−0.017	−0.007	—	—	−0.108
CAA	0.140	−0.072	0.214	0.039	−0.010	−0.090	—	—
CAC	0.019	−0.062	−0.129	−0.016	−0.040	—	0.077	—
CAG	0.250	−0.024	0.274	0.105	0.002	—	—	—
CAT	0.076	−0.064	0.013	0.019	−0.032	—	—	0.095
GAA	0.042	0.008	0.037	−0.021	−0.009	−0.032	—	—
GAC	0.002	0.005	−0.175	−0.037	−0.034	—	0.003	—
GAG	0.096	0.032	0.055	0.015	−0.007	—	—	—
GAT	−0.019	−0.001	−0.128	−0.048	−0.037	—	—	−0.040
TAA	0.117	−0.037	0.164	0.016	−0.016	−0.029	—	—
TAC	0.006	−0.049	−0.126	−0.020	−0.031	—	0.052	—
TAG	0.239	−0.010	0.235	0.089	0.005	—	—	—
TAT	0.051	−0.010	−0.007	0.004	−0.013	—	—	0.030

[a] $3'$-$\Delta_{A\cdot A/A\cdot T(XAY)}(Y) = \delta_{refA\cdot A(XAY)}(Y) - \delta_{refA\cdot T(XAY)}(Y)$.

been determined to be 0.06 ppm with a correlation coefficient of 0.9997.

3.4. Triplet values and correction factors from bottom strands of reference sequences

Apart from the set of A·A mismatch triplet values and correction factors shown in Tables 1–3, another set of data has also been extracted from the triplets in the bottom strands of the reference sequences in Fig. 1c and they have been summarized in Supplementary materials S2–S4. Similarly, the new set of triplet values can be used directly to predict chemical shifts of A·A mismatches. For predicting the chemical shifts of the flanking nucleotides X and Y using this new set of correction factors, Eqs. (2) and (4) have to be modified to:

$$\delta_{pred}(X) = \delta_{duplexT\cdot A(XTY)}(X) + 5'\text{-}\Delta_{A\cdot A/T\cdot A(XAY)}(X) \quad (6)$$

$$\delta_{pred}(Y) = \delta_{duplexT\cdot A(XTY)}(Y) + 3'\text{-}\Delta_{A\cdot A/T\cdot A(XAY)}(Y) \quad (7)$$

where $\delta_{duplexT\cdot A(XTY)}(X)$ and $\delta_{duplexT\cdot A(XTY)}(Y)$ are the chemical shifts of X and Y in the normal duplex containing an T·A base pair (Fig. 1b), respectively. $5'$-$\Delta_{A\cdot A/T\cdot A(XAY)}(X)$ is the new 5′-correction factors which can be determined from the chemical shift difference of X in the bottom strands of refA·A(Y′AX′) and refA·T(Y′AX′) whereas $3'$-$\Delta_{A\cdot A/T\cdot A(XAY)}(Y)$ is the new 3′-correction factor which can be determined from the chemical shift difference of Y in the bottom strands of refA·A(Y′AX′) and refA·T(Y′AX′), i.e.

$$5'\text{-}\Delta_{\text{A·A/T·A(XAY)}}(\text{X}) = \delta_{\text{refA·A(Y'AX')}}(\text{X}) - \delta_{\text{refA·T(Y'AX')}}(\text{X}) \quad (8)$$

$$3'\text{-}\Delta_{\text{A·A/T·A(XAY)}}(\text{Y}) = \delta_{\text{refA·A(Y'AX')}}(\text{Y}) - \delta_{\text{refA·T(Y'AX')}}(\text{Y}) \quad (9)$$

In these two data sets, the positions of nucleotides used for extracting the 3′-correction factors from the top strands and the 5′-correction factors from the bottom strands are only one nucleotide apart from the GAA loop in the reference sequences. Therefore, these correction factors may be influenced by the loop. In order to investigate the loop effect, comparison of the RMSD values using different sets of correction factors has been made in Table 4. No significant difference has been found in the RMSD values between the two sets of 5′ and 3′-correction factors, revealing the loop effect is insignificant. In addition, the RMSD values of results predicted using the top and bottom strand triplet chemical shifts, and the overall RMSD values are all within 0.06–0.08 ppm with excellent correlation coefficients, indicating both data sets can be used reliably in the prediction method.

Table 4
Comparison of prediction accuracy using values extracted from the top and bottom strands of refA·A(XAY) and refA·T(XAY)

Value extracted	No. of test data	Prediction accuracy (correlation coefficient), ppm[a]	
		From top strand	From bottom strand
Triplet XAY	192	0.06 (0.9998)	0.08 (0.9998)
5′-Correction factor	184	0.08 (0.9994)	0.06 (0.9996)
3′-Correction factor	184	0.06 (0.9997)	0.06 (0.9998)
Overall	560	0.07 (0.9996)	0.07 (0.9997)

[a] Prediction accuracy was determined from RMSD between the predicted and experimental values.

Table 5
Comparison of prediction accuracy using values extracted from different sets of reference sequences

Nucleus	No. of test data	Prediction accuracy (correlation coefficient), ppm[a]	
		From refA·T(XAY) set[b]	From refT·A(XTY) set[b]
H6/H8	96	0.07 (0.9938)	0.06 (0.9956)
H1′	96	0.08 (0.9564)	0.06 (0.9784)
H2′	96	0.08 (0.9827)	0.08 (0.9821)
H2″	96	0.05 (0.9777)	0.04 (0.9857)
H3′	96	0.03 (0.9720)	0.03 (0.9591)
A-H2	48	0.08 (0.9763)	0.09 (0.9714)
C-H5	16	0.08 (0.9395)	0.05 (0.9598)
T-H7	16	0.06 (0.9500)	0.06 (0.9425)
Overall	560	0.07 (0.9996)	0.06 (0.9997)

[a] Prediction accuracy was determined from RMSD between the predicted and experimental values.
[b] Predictions were made using triplet values extracted from the top strands of refA·A(XAY) and correction factors extracted from the top strands of refA·A(XAY) and refA·T(XAY) or refT·A(XTY).

The prediction accuracy of this method using 5′ and 3′-correction factors is limited by the prediction accuracy of Altona method because predicted chemical shifts of normal B-DNA duplexes are needed. Nevertheless, the prediction accuracy of this newly established method and Altona method are comparable and they are in the same order of magnitude. The prediction accuracy has also been examined according to the types of protons. In Table 5, the RMSD values from refA·T(XAY) set vary from 0.03 to 0.08 ppm among different types of protons. Fig. 2a shows an excellent correlation plot ($r = 0.9996$) between 560 sets of predicted and experimental chemical shifts and the overall RMSD value has been found to be 0.07 ppm, indicating the high reliability of this prediction method.

3.5. Verification of the base pair replacement approach

To verify the applicability of the base pair replacement approach adopted in this prediction method, chemical shifts of a new set of reference sequences, refT·A(XTY) (Fig. 1d) have also been measured. In analog to refA·T (XAY), 5′- and 3′-A·A correction factors have been determined from the top strands of refA·A(XAY) and refT·A(XTY) and the results have been tabulated in Supplementary materials S5 and S6. In this case, these correction factors represent the changes in chemical shifts upon

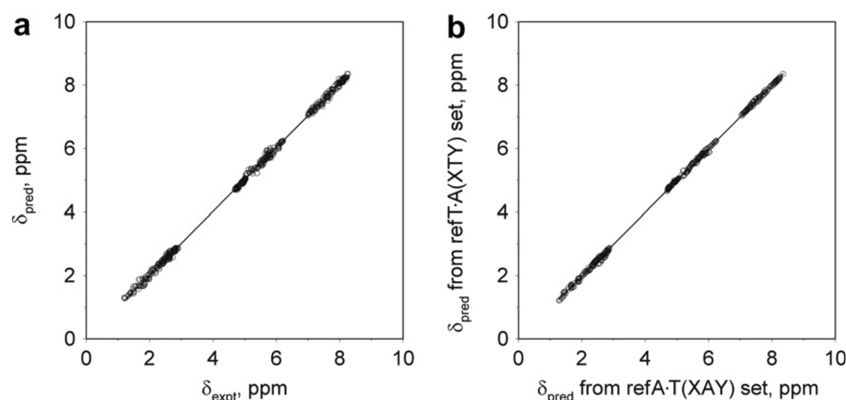

Fig. 2. (a) Plot of predicted chemical shifts of A·A mismatch and its flanking residues using values extracted from refA·T(XAY) set versus experimental chemical shifts. (b) Plot of predicted chemical shifts of A·A mismatch and its flanking residues using correction factors extracted from refA·T(XAY) set versus those using correction factors extracted from refT·A(XTY) set.

replacing the A·A mismatch in refA·A(XAY) with a T·A base pair, i.e.

$$5'\text{-}\Delta_{\text{A}\cdot\text{A/T}\cdot\text{A(XAY)}}(\text{X}) = \delta_{\text{refA}\cdot\text{A(XAY)}}(\text{X}) - \delta_{\text{refT}\cdot\text{A(XTY)}}(\text{X}) \quad (10)$$

$$3'\text{-}\Delta_{\text{A}\cdot\text{A/T}\cdot\text{A(XAY)}}(\text{Y}) = \delta_{\text{refA}\cdot\text{A(XAY)}}(\text{Y}) - \delta_{\text{refT}\cdot\text{A(XTY)}}(\text{Y}) \quad (11)$$

where $\delta_{\text{refT}\cdot\text{A(XTY)}}(\text{X})$ and $\delta_{\text{refT}\cdot\text{A(XTY)}}(\text{Y})$ are the chemical shifts of X and Y in refT·A(XTY), respectively. In order to predict the chemical shifts of the flanking residues X and Y of A·A mismatch, Eqs. (6) and (7) can be used again. Together with the predicted chemical shifts of mismatched A, an excellent correlation plot ($r = 0.9998$) between the prediction results from refT·A(XTY) set and refA·T(XAY) set has been made in Fig. 2b. The prediction accuracy from these two reference sets on various types of protons have also been found to be very similar (Table 5), indicating the base pair replacement approach is applicable in this newly established prediction method.

4. Conclusions

A reliable chemical shift prediction method for A·A mismatch and its flanking residues in B-DNA duplexes has been established with a good prediction accuracy of 0.07 ppm. Excellent correlations between the predicted and experimental values have been obtained. The base pair replacement approach is also applicable to further development of chemical shift prediction method on other types of mismatches.

Acknowledgment

We thank the valuable suggestions from F.K. Lo, Y. Zhan, D. Fenn and H. Yang. 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.: CUHK400704) and a direct grant from CUHK Research Committee Funding (Project Code: 2060252).

Appendix A. Supplementary data

A list of testing sequences and five tables containing chemical shifts of A in XAY triplets extracted from the bottom strands of refA·A(Y′AX′), 5′- and 3′-correction factors extracted from the bottom strands of refA·A(Y′AX′) and refA·T(Y′AX′), and from the top strands of refA·A(XAY) and refT·A(XTY). Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmr.2007.04.005.

References

[1] D.A. Case, H.J. Dyson, P.E. Wright, Use of chemical shifts and coupling constants in nuclear magnetic resonance structural studies on peptides and proteins, Methods Enzymol. 239 (1994) 392–416.
[2] L. Szilagyi, Chemical shifts in proteins come of age, Prog. Nucl. Magn. Reson. Spectrosc. 27 (1995) 325–443.
[3] M.P. Williamson, T. Asakura, Protein chemical shifts, Methods Mol. Biol. 60 (1997) 53–59.
[4] D.S. Wishart, B.D. Sykes, Chemical shifts as a tool for structure determination, Methods Enzymol. 239 (1994) 363–392.
[5] S.L. Lam, L.N. Ip, X. Cui, C.N. Ho, Random coil proton chemical shifts of deoxyribonucleic acids, J. Biomol. NMR 24 (2002) 329–337.
[6] C.W. Kwok, C.N. Ho, L.M. Chi, S.L. Lam, Random coil carbon chemical shifts of deoxyribonucleic acids, J. Magn. Reson. 166 (2004) 11–18.
[7] C.N. Ho, S.L. Lam, Random coil phosphorus chemical shift of deoxyribonucleic acids, J. Magn. Reson. 171 (2004) 193–200.
[8] A. Dejaegere, R.A. Bryce, D.A. Case, in: J.C. Facelli, A.C. de Dios (Eds.), Modeling NMR Chemical Shifts. Gaining Insight into Structure and Environment, American Chemical Society, Washington, DC, 1999, pp. 194–206.
[9] S.S. Wijmenga, M. Kruithof, C.W. Hilbers, Analysis of chemical shifts in DNA: assessment of the reliability of ^{1}H chemical shift calculations for use in structure refinement, J. Biomol. NMR 10 (1997) 337–350.
[10] C. Altona, D.H. Faber, A.J.A. Westra Hoekzema, Double-helical DNA ^{1}H chemical shifts: an accurate and balanced predictive empirical scheme, Magn. Reson. Chem. 38 (2000) 95–107.
[11] J.A. Cromsigt, C.W. Hilbers, S.S. Wijmenga, Prediction of proton chemical shifts in RNA. Their use in structure refinement and validation, J. Biomol. NMR 21 (2001) 11–29.
[12] M.F. Goodman, S. Creighton, L.B. Bloom, J. Petruska, Biochemical basis of DNA replication fidelity, Crit. Rev. Biochem. Mol. Biol. 28 (1993) 83–126.
[13] L.M. Chi, S.L. Lam, NMR investigation of DNA primer-template models: structural insights into dislocation mutagenesis in DNA replication, FEBS Lett. 580 (2006) 6496–6500.
[14] A. Bhattacharyya, D.M. Lilley, Single base mismatches in DNA. Long- and short-range structure probed by analysis of axis trajectory and local chemical reactivity, J. Mol. Biol. 209 (1989) 583–597.
[15] T. Lindahl, Instability and decay of the primary structure of DNA, Nature 362 (1993) 709–715.
[16] G.A. Leonard, E.D. Booth, T. Brown, Structural and thermodynamic studies on the adenine·guanine mismatch in B-DNA, Nucleic Acids Res. 18 (1990) 5617–5623.
[17] G.E. Plum, A.P. Grollman, F. Johnson, K.J. Breslauer, Influence of the oxidatively damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex, Biochemistry 34 (1995) 16148–16160.
[18] F. Hutchinson, Chemical changes induced in DNA by ionizing radiation, Prog. Nucleic Acid Res. Mol. Biol. 32 (1985) 115–154.
[19] J.F. Ward, DNA damage produced by ionizing radiation in mammalian cells: identities, mechanisms of formation, and reparability, Prog. Nucleic Acid Res. Mol. Biol. 35 (1988) 95–125.
[20] C.T. Ashley Jr., S.T. Warren, Trinucleotide repeat expansion and human disease, Annu. Rev. Genet. 29 (1995) 703–728.
[21] R. Gellibolian, A. Bacolla, R.D. Wells, Triplet repeat instability and DNA topology: an expansion model based on statistical mechanics, J. Biol. Chem. 272 (1997) 16793–16797.
[22] C.E. Pearson, Y.H. Wang, J.D. Griffith, R.R. Sinden, Structural analysis of slipped-strand DNA (S-DNA) formed in (CTG)n.(CAG)n repeats from the myotonic dystrophy locus, Nucleic Acids Res. 26 (1998) 816–823.
[23] J.B. Vincent, A.D. Paterson, E. Strong, A. Petronis, J.L. Kennedy, The unstable trinucleotide repeat story of major psychosis, Am. J. Med. Genet. 97 (2000) 77–97.
[24] A.K. Mosemiller, J.C. Dalton, J.W. Day, L.P. Ranum, Molecular genetics of spinocerebellar ataxia type 8 (SCA8), Cytogenet. Genome Res. 100 (2003) 175–183.
[25] B.A. Lenzmeier, C.H. Freudenreich, Trinucleotide repeat instability: a hairpin curve at the crossroads of replication, recombination, and repair, Cytogenet. Genome Res. 100 (2003) 7–24.
[26] M.J. Hartenstine, M.F. Goodman, J. Petruska, Base stacking and even/odd behavior of hairpin loops in DNA triplet repeat slippage

and expansion with DNA polymerase, J. Biol. Chem. 275 (2000) 18382–18390.

[27] J. Petruska, M.J. Hartenstine, M.F. Goodman, Analysis of strand slippage in DNA polymerase expansions of CAG/CTG triplet repeats associated with neurodegenerative disease, J. Biol. Chem. 273 (1998) 5204–5210.

[28] K. Ohshima, R.D. Wells, Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes, J. Biol. Chem. 272 (1997) 16798–16806.

[29] L.M. Chi, S.L. Lam, Structural roles of CTG repeats in slippage expansion during DNA replication, Nucleic Acids Res. 33 (2005) 1604–1617.

[30] G. Bodenhauser, D.J. Ruben, Natural abundance ^{15}N NMR by enhanced heteronuclear spectroscopy, Chem. Phys. Lett. 69 (1980) 185–189.

[31] T.J. Norwood, J. Boyd, J.E. Heritage, N. Soffe, I.D. Campbell, Comparison of techniques for proton-detected heteronuclear proton–nitrogen-15 spectroscopy, J. Magn. Reson. 87 (1990) 488–501.

[32] M.J. van Dongen, S.S. Wijmenga, R. Eritja, F. Azorin, C.W. Hilbers, Through-bond correlation of adenine H2 and H8 protons in unlabeled DNA fragments by HMBC spectroscopy, J. Biomol. NMR 8 (1996) 207–212.

[33] A.J. Shaker, P.B. Barker, R. Freeman, Computer-optimized decoupling scheme for wideband applications and low-level operation, J. Magn. Reson. 64 (1985) 547–552.

[34] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, J. Biomol. NMR 5 (1995) 67–81.