

# Conformational Changes of the *BS2* Operator DNA upon Complex Formation with the *Antennapedia* Homeodomain Studied by NMR with $^{13}\text{C}/^{15}\text{N}$ -labeled DNA

César Fernández<sup>1</sup>, Thomas Szyperski<sup>1</sup>, Martin Billeter<sup>1</sup>, Akira Ono<sup>2</sup>  
 Hideo Iwai<sup>1</sup>, Masatsune Kainosho<sup>2\*</sup> and Kurt Wüthrich<sup>1\*</sup>

<sup>1</sup>*Institut für Molekularbiologie und Biophysik, Eidgenössische Technische Hochschule Hönggerberg, CH-8093, Zürich Switzerland*

<sup>2</sup>*Department of Chemistry Faculty of Science Tokyo Metropolitan University 1-1 Minamiohsawa, Hachioji 192-0397, Japan*

The NMR structures have been determined for a  $^{13}\text{C}/^{15}\text{N}$  doubly labeled 14 base-pair DNA duplex comprising the *BS2* operator sequence both free in solution and in the complex with the *Antennapedia* homeodomain. The impact of the DNA labeling is assessed from comparison with a previous structure of the same complex that was determined using isotope labeling only for the protein. Differences between the two structure determinations are nearly completely limited to the DNA, which retains the global *B*-conformation of the free DNA also in the complex. Local protein-induced conformational changes are a narrowing of the minor groove due to the interaction with the N-terminal arm of the homeodomain, and changes of the sugar puckers of the deoxyriboses G5 and C6, which are apparently induced by van der Waals interactions with Tyr25, and with Gln50 and Arg53, respectively. The high conservation of these amino acid residues in homeodomains suggests that protein-induced shifts in some sugar puckers contribute to the affinity of homeodomains to their cognate DNA. The data obtained here with the *Antennapedia* homeodomain-DNA complex clearly show that nucleic acid isotope-labeling can support detailed conformational characterization of DNA in complexes with proteins, which will be indispensable for structure determinations of complexes containing globally distorted DNA conformations.

© 1999 Academic Press

**Keywords:** protein-DNA complexes; homeodomains; NMR; isotope labeling of DNA

\*Corresponding authors

## Introduction

Three classes of genes have been described that specify the body plan (Gehring, 1987) of the fruit fly

Present addresses: T. Szyperski, Department of Chemistry, State University of New York at Buffalo, 816 Natural Sciences Complex, Buffalo, NY 14260, USA; M. Billeter, Biochemistry and Biophysics, Lundberg Laboratory, Göteborg University, Box 462, SE 405 30 Göteborg, Sweden; H. Iwai, Biochemisches Institut, Universität Zürich, Winterthurerstr. 190, CH-8057 Zürich, Switzerland.

Abbreviations used: *Antp*, *Antennapedia*; *BS2*, operator binding site 2; COSY, correlation spectroscopy; HD, homeodomain; NOE, nuclear Overhauser effect; NOESY, NOE spectroscopy; RMSD, root-mean-square deviation.

*Drosophila melanogaster*. Genes from all three classes contain a highly conserved coding fragment of about 180 bp, which is commonly referred to as the *homeobox* (McGinnis *et al.*, 1984; Scott & Weiner, 1984). The *homeobox* encodes the 60 amino acid residue homeodomain (HD), which was first found in the *Drosophila* homeotic mutant *Antennapedia* (*Antp*) (McGinnis *et al.*, 1984; Scott & Weiner, 1984). Homeodomains have since been discovered in organisms ranging from yeast to humans (Scott *et al.*, 1989), and mediate DNA-binding of the entire protein encoded by the *homeobox*-containing genes; for example, the *Antp* protein. Because of their interest for developmental biology, homeodomain-DNA interactions have been extensively studied, and have actually become a paradigm for studies of general principles governing protein-DNA interactions. A recombinant *Antp* homeodomain

polypeptide of 68 amino acid residues has been shown to retain the ability of specific binding to the cognate DNA (Müller *et al.*, 1988), and the NMR structures of this polypeptide (Quian *et al.*, 1989; Billeter *et al.*, 1990) as well as its complex with a 14 bp DNA duplex containing the BS2 binding site have been determined (Otting *et al.*, 1990; Quian *et al.*, 1993a,b; Billeter *et al.*, 1993, 1996). Recently, the structure of the *Antp* HD-DNA complex has been solved by X-ray crystallography, and from comparison with the NMR data the authors conclude that there is a high level of compatibility between the results obtained with the two techniques (Fraenkel & Pabo, 1998).

So far, the *Antp* HD-DNA complex has been studied with isotope labeling of the protein (Quian *et al.*, 1993; Billeter *et al.*, 1993). Here, we use this system to evaluate the impact of additional isotope-labeling of the DNA in NMR structure determinations of protein-DNA complexes, in particular for structure refinement of the bound DNA. The 14 bp DNA duplex used to prepare the *Antp* HD complex was either fully  $^{13}\text{C},^{15}\text{N}$ -labeled, or uniformly  $^{13}\text{C},^{15}\text{N}$ -labeled at the nucleotides implicated in intermolecular DNA-protein NOEs in the

previously determined NMR solution structure of the *Antp* HD-DNA complex (Billeter *et al.*, 1993). To identify the conformational changes of the DNA that arise from complex formation, the NMR structure of the labeled free DNA duplex was determined and compared with the structure of the labeled DNA duplex when bound to the homeodomain.

## Results

### Impact of DNA $^{13}\text{C}/^{15}\text{N}$ -labeling for the NMR structure determination

The use of the labeled DNA resulted in a three-fold increase of the number of intra-DNA NOEs, and the  $^{13}\text{C}/^{15}\text{N}$  double-labeling of the DNA further enabled measurement of a large number of otherwise inaccessible spin-spin coupling constants, i.e.  $^3J_{\text{C}2'\text{P}}$  (Szyperski *et al.*, 1997),  $^3J_{\text{HH}}$  (Szyperski *et al.*, 1998),  $^3J_{\text{C}4'\text{P}}$ ,  $^3J_{\text{H}3'\text{P}}$  (Szyperski *et al.*, 1999a) and  $^3J_{\text{H}5'/\text{H}5'\text{P}}$ . In view of the resulting large number of conformational constraints (Table 1), we decided to use a structure calculation protocol similar to the standard protocol for protein struc-

**Table 1.** Statistics for the determination of the energy-refined NMR structure of the free DNA duplex of Figure 2 and the *Antp* homeodomain-DNA complex with the program DYANA

	Free DNA duplex	DNA duplex bound to the <i>Antp</i> HD	<i>Antp</i> HD-DNA complex
<b>A. Input for the DYANA structure calculation</b>			
NOE upper distance limits <sup>a</sup>	570	475 <sup>b</sup>	1359 <sup>b</sup>
Hydrogen bond distance constraints <sup>c</sup>	112	112	112
Ring closure distance constraints <sup>d</sup>	280	280	280
Dihedral angle constraints <sup>e</sup>	293	284	622
Stereospecific assignments, DNA <sup>f</sup>	H2'/H2''	24	24
	H5'/H5''	5	6
<b>B. Residual violations of upper distance constraints and dihedral angle constraints</b>			
DYANA target function <sup>g</sup> ( $\text{\AA}^2$ )	Average 0.94	0.69	5.26
	Range 0.84 ... 0.98	0.61 ... 0.74	3.93 ... 6.42
NOE constraint violations <sup>h</sup> ( $\text{\AA}$ )	$n > 0.1 \text{\AA}$	$0.10 \pm 0.30$	$0.60 \pm 0.73$
	Max	$0.09 \pm 0.00$	$0.10 \pm 0.01$
Dihedral angle violations <sup>h</sup> (deg.)	$n > 2^\circ$	$2.70 \pm 1.79$	$8.70 \pm 2.35$
	Max	$2.46 \pm 1.13$	$3.29 \pm 0.51$
<b>C. AMBER energies<sup>h</sup></b>			
Van der Waals ( $\text{kcal mol}^{-1}$ )	$-235 \pm 8$	$-213 \pm 6$	$-433 \pm 22$
Electrostatic ( $\text{kcal mol}^{-1}$ )	$1273 \pm 77$	$964 \pm 38$	$-5521 \pm 244$
<b>D. Average RMS deviations for different atom selections<sup>i</sup> (<math>\text{\AA}</math>)</b>			
DNA 2-13, 16-27	Backbone	1.38 (3.30)	1.49
	Heavy-atoms	1.20 (2.84)	1.34
DNA 2-13, 16-27; <i>Antp</i> 8-56	Backbone	-	1.20 (1.63)
	Heavy-atoms	-	1.35 (1.69)
DNA 7-11, 18-22; <i>Antp</i> 43-54 (interface)	Backbone	-	0.56 (1.05)
	Heavy-atoms	-	0.78 (1.11)

<sup>a</sup> Number of NOE constraints after considering pseudo-atom corrections with the program DYANA (Güntert *et al.*, 1997).

<sup>b</sup> The corresponding values in the previous NMR structure determination of the *Antp* HD-DNA complex without isotope labeling of the DNA were: 151 (DNA duplex bound to the *Antp* HD) and 1045 (*Antp* HD-DNA complex) (Billeter *et al.*, 1993).

<sup>c</sup> Hydrogen bond constraints (two distance limit constraints for each hydrogen bond) were introduced for the 12 non-terminal Watson-Crick base-pairs of the duplex as inferred from the imino proton exchange rates (see Materials and Methods, and Figure 2).

<sup>d</sup> Five distance limits per deoxyribose moiety were applied to enforce ring closure (see Materials and Methods).

<sup>e</sup> Output of the grid search performed with the FOUND (Güntert *et al.*, 1997) module of DYANA.

<sup>f</sup> Stereospecific assignments were obtained using the FOUND and GLOMSA modules of DYANA.

<sup>g</sup> Before energy minimization.

<sup>h</sup> Average over the 20 energy-refined DYANA conformers and standard deviation.

<sup>i</sup> The values in parentheses were obtained for a structure determination performed without inclusion of the dihedral angle constraints derived from scalar spin-spin couplings into the input (see the text).

ture determination (Wüthrich, 1986; Szyperski *et al.*, 1999b). To assess the impact of the newly measured scalar couplings on the structure determination, DYANA (Güntert *et al.*, 1997) calculations were performed with and without inclusion of these conformational constraints. Comparison of the RMSD values (Table 1) revealed that the additional measurement of these couplings resulted in significantly improved precision of the DNA structure determination in the free and HD-bound forms.

As expected, no additional intra-protein constraint resulted from the use of the labeled DNA, and when compared with the structure determination based on labeling of the protein (Billeter *et al.*, 1993), only seven new intermolecular protein-DNA NOEs could be observed. This outcome was anticipated from a previous theoretical analysis of the HD-DNA interface (Billeter & Wüthrich, 1993).

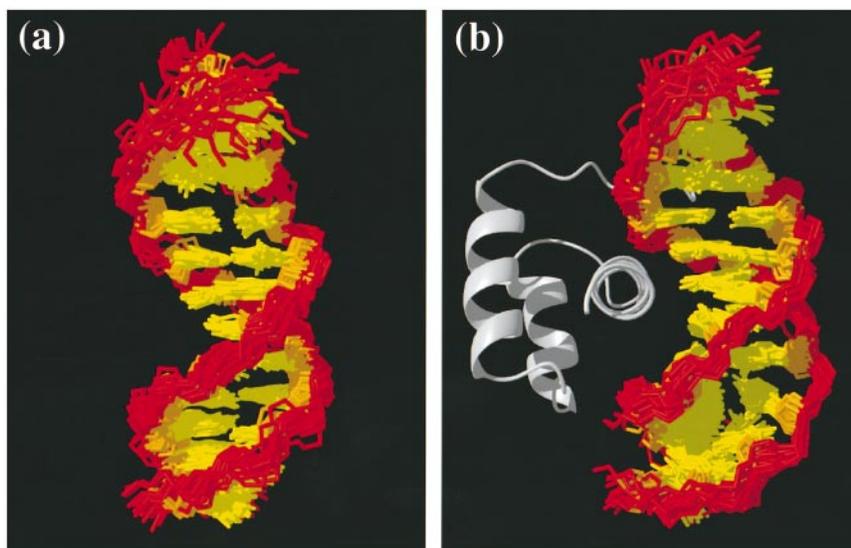
### NMR structure of the free DNA duplex in solution

The NMR structure of the free 14mer DNA duplex was obtained from 1255 conformational constraints using the program DYANA (Table 1). Figure 1(a) shows a superposition of the 20 best energy-refined DYANA conformers that were selected to represent the solution structure. The high quality of the structure determination is reflected by the small residual constraint violations, and by the RMSD value of 1.2 Å relative to the mean coordinates calculated for all heavy-atoms of the non-terminal base-pairs formed by nucleotides 2 to 13 and 16 to 27 (Figure 2). The presence of a stable duplex is clearly manifested by the slow imino proton exchange rates, which further reflect increased fraying of base-pairs 1-28, 2-27, 3-26, 13-16 and 14-15 (Figure 2(a)).

The ranges of the backbone and glycosidic dihedral angle values found in the 20 energy-refined DYANA conformers are very similar to those of a standard *B*-helix (data not shown). The same is reflected by the average helical twist of  $35.9(\pm 1.2)^\circ$  (standard *B*-DNA  $35.6^\circ$ ), which corresponds to 10 bp per turn (standard *B*-DNA 10.1), and an average rise per base-pair of  $3.54(\pm 0.10)$  Å (standard *B*-DNA 3.38 Å) (Hartmann & Lavery, 1996). Moreover, only canonical  $B_1$  phosphodiester backbone conformations ( $\epsilon^t\zeta^-$ ) were observed (Szyperski *et al.*, 1997).

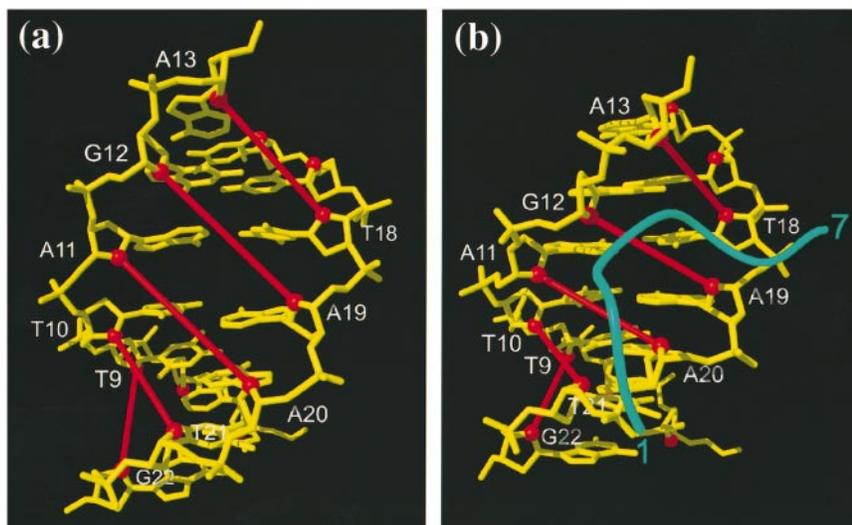
### NMR solution structure of the *Antp* homeodomain-DNA complex

The NMR structure of the *Antp* HD-DNA complex was obtained from 2373 conformational constraints using the program DYANA (Table 1 and Figure 1(b)). Compared to the previous structure determination without DNA labeling, the DNA-protein interface is better defined, as evidenced by the low RMSD values for the atom selection from the interface (Table 1), for which the values 0.84 Å for the backbone and 1.01 Å for all heavy-atoms had been obtained without DNA-labeling (Billeter *et al.*, 1993). The orientation of the side-chains of Gln50 and Asn51 is still not well defined, but intermolecular hydrogen bonds are observed in part of the energy-refined DYANA conformers. These two residues form direct or water-mediated contacts with the DNA in all 3D structures of homeodomain-DNA complexes known to date (Billeter *et al.*, 1993; Fraenkel & Pabo, 1998; Tucker-Kellog *et al.*, 1998; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Klem *et al.*, 1994; Hirsch & Aggarwal, 1995; Li *et al.*, 1995; Wilson *et al.*, 1995; Wolberger, 1996). Multiple local conformations of Gln50 and Asn51 were observed in most of these crystal structures, which provides a rationale for the apparent exchange



**Figure 1.** NMR structures determined with  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled 5'-GAA AGCCATTAGAG-3' · 5'-CTCTAATG GCTTTC-3'. (a) Free DNA in solution. (b) DNA in the *Antp* HD-DNA complex. All heavy-atoms of the 12 non-terminal base-pairs in the 20 best energy-refined DYANA conformers used to represent the NMR structure have been superimposed for minimal RMSD (Table 1). Color code: phosphodiester backbone, red; deoxyribose rings, orange; bases, yellow. In (b), the backbone of the *Antp* HD is represented by a grey ribbon.

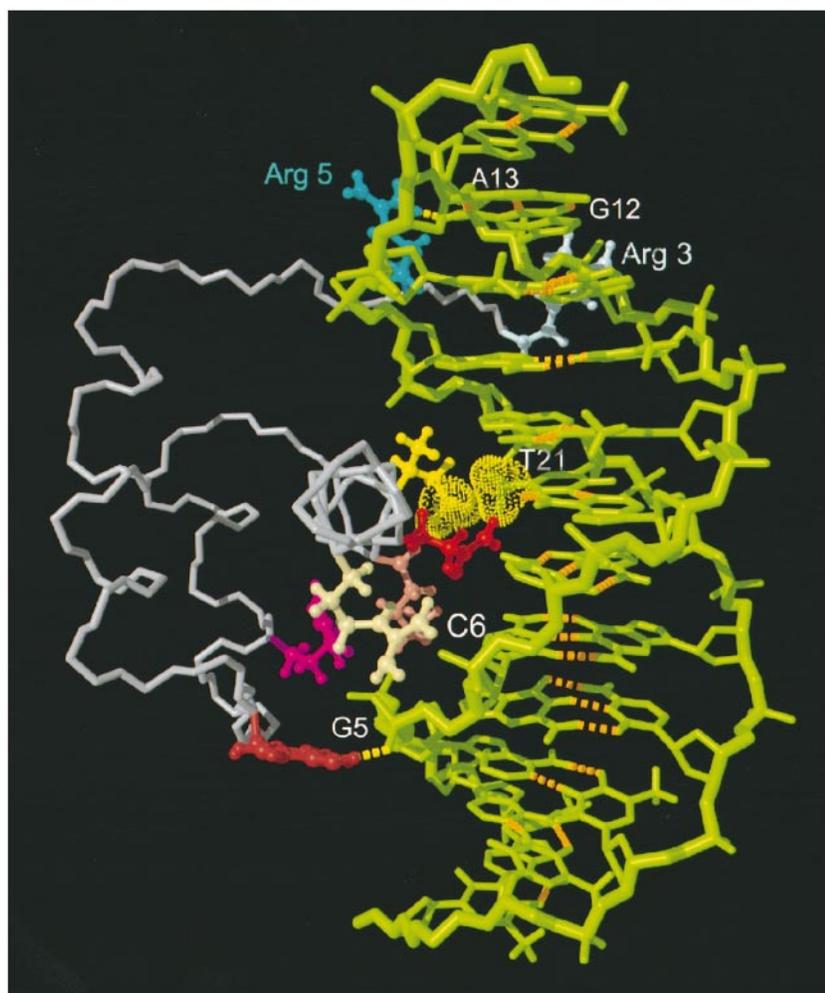




**Figure 3.** View into the minor groove of the DNA segment comprising bases 7 to 13 and 16 to 22. The DYANA conformer with the smallest RMSD from the mean coordinates is shown. (a) Free DNA. (b) DNA in the *Antp* HD-DNA complex. In (b), the N-terminal arm of the protein with residues 1 to 7 is represented by a blue spline function drawn through the  $\alpha$ -carbon atoms. The shortest O4'-O4' distances across the minor groove are indicated by red bars (see also Figure 2(b)).

these conformational shifts, i.e. the phosphate group of G5 is contacted by Arg31 and Lys46, and the backbone atoms of G12 and A13 interact with Arg3 and Arg5 in the minor groove, respectively (Figure 4). Since Arg5, Arg31 and Lys46 are con-

served in more than 97% of all homeodomain sequences known to date (Billeter, 1996), it is tempting to speculate that these intermolecular contacts are important for stabilizing the HD-DNA complexes.



**Figure 4.** Intermolecular contacts in the *Antp* HD-DNA complex described in the text. In the DNA major groove, Ile47 (yellow) shows hydrophobic contacts with the methyl group of T21, Gln50 (red) and Arg53 (beige) contact nucleotide C6, and Tyr25 (brown), Arg31 (magenta) and Lys46 (pink) interact with nucleotide G5. In the minor groove, Arg5 (blue) forms a hydrogen bond to backbone atoms of A13, and Arg3 (light blue) shows contacts to the phosphate group of G12. The amino acid side-chains are represented as ball-and-stick models. Hydrogen bonds within the DNA duplex are shown as broken orange lines, two intermolecular hydrogen bonds as broken yellow lines. The van der Waals contacts between Ile47 and T21 are indicated with yellow dotted spheres. The DNA is colored green and the backbone of the *Antp* HD is displayed in gray.

Analysis of the deoxyribose conformations in the free DNA and the protein-DNA complex was performed within the framework of a two-state model (Hartmann & Lavery, 1996; Widmer & Wüthrich, 1987), assuming rapid interconversion between the C2'-endo and C3'-endo states (pucker amplitude 35°). The nearly complete sets of  $^3J_{\text{HH}}$  scalar couplings (Table 2) (Szyperski *et al.*, 1998) show that all deoxyribose rings, on average, adopt about 90% C2'-endo, except for G5 in the *Antp* HD-DNA complex, which exhibits 45% to 70% C2'-endo conformer. Although no dihedral angle constraints in the complex were derived for  $\nu_1$  and  $\nu_2$  in G5 (see Materials and Methods), NOE data yielded S conformation (C2'-endo to C1'-exo), which is similar to all other sugar rings. Intriguingly, the sugar ring of G5 interacts with the aromatic ring of Tyr25 (Figure 4). Similarly, shifts of  $^3J_{\text{H1}'\text{H2}'}$  observed for C6 and T21 upon complexation (Table 2) can be correlated with intermolecular van der Waals interactions. In particular, C6 contacts directly the side-chains of Gln50 and Arg53, and T21 is involved in numerous intermolecular contacts, although no direct van der Waals contacts are seen from its deoxyribose atoms to any particular protein residue. The apparent conformational adaptability of the deoxyriboses may contribute to improved intermolecular contacts in the *Antp* HD-DNA complex, thereby increasing its stability. The aforemen-

tioned contacts are found in all three-dimensional structures of HD-DNA complexes known to date (Billeter *et al.*, 1993; Fraenkel & Pabo, 1998; Tucker-Kellog *et al.*, 1998; Kissinger *et al.*, 1990; Wolberger *et al.*, 1991; Klem *et al.*, 1994; Hirsch & Aggarwal, 1995; Li *et al.*, 1995; Wilson *et al.*, 1995; Wolberger, 1996). Furthermore, Tyr25, Gln50 and Arg53 are among the most highly conserved residues in homeodomains (Billeter, 1996), suggesting that the presently identified "induced fit" interactions involving conformational shifts of sugar puckers might be a general feature of the formation of homeodomain-DNA complexes.

### Implications for the use of isotope-labeled nucleic acids in studies of protein-nucleic acid complexes

The presently studied *Antp* HD-DNA complex is, by today's standards, a quite simple system, and the previous NMR structure determination without DNA-labeling (Otting *et al.*, 1990; Qian *et al.*, 1993a,b; Billeter *et al.*, 1993) have been fully re-validated by a recent crystallographic study (Fraenkel & Pabo, 1998). An important advance obtained with the use of the labeled DNA is that both the DNA and the protein part of the structure have now been determined *de novo*, starting from randomized coils, whereas the previous structure

**Table 2.**  $^3J_{\text{HH}}$  Couplings (Hz) measured for the DNA duplex of Figure 2 free in solution and in the *Antp*(C395) HD-DNA complex

Nucleotide <sup>a</sup>	$^3J_{\text{H1}'\text{H2}'}$	$^3J_{\text{H1}'\text{H2}''}$	$^3J_{\text{H2}'\text{H3}'}$	$^3J_{\text{H2}''\text{H3}''}$	$^3J_{\text{H3}'\text{H4}'}$
G1	9.3/9.1	5.6/5.2	4.6/4.2	1.0/0.3	1.4/ <2.0
A2	9.4/8.9	-/4.8	4.1/3.5	<2.0/1.6	<2.0/ <2.0
A3	9.1/9.1	5.0/4.5	5.0/4.0	<2.0/ <2.0	2.3/ <2.0
<u>A4</u>	8.0/8.5	5.9/5.4	5.4/3.6	<2.0/1.2	2.5/ <2.0
<u>G5</u>	<b>9.5/5.6<sup>b</sup></b>	5.9/5.8	5.4/6.5	<b>&lt;2.0/4.3<sup>b</sup></b>	<b>2.5/4.1<sup>b</sup></b>
<u>C6</u>	<b>8.6/6.8<sup>b</sup></b>	5.8/4.6	5.3/-	2.5/-	3.6/-
<u>C7</u>	8.9/8.1	5.8/5.4	5.5/6.1	1.7/2.1	2.8/4.1
A8	8.9/8.1	5.5/4.3	4.6/-	0.8/-	<2.0/-
T9	8.4/8.3	5.8/5.4	7.0/5.7	1.8/ <2.0	4.0/2.7
T10	9.4/8.6	5.9/4.8	5.0/-	0.7/-	3.2/-
A11	9.7/- <sup>c</sup>	5.6/- <sup>c</sup>	4.2/- <sup>c</sup>	2.1/- <sup>c</sup>	2.2/ <2.0
<u>G12</u>	9.8/- <sup>c</sup>	4.3/- <sup>c</sup>	4.8/- <sup>c</sup>	<2.0/- <sup>c</sup>	1.9/ <2.0
<u>A13</u>	8.5/9.8	5.5/6.1	4.3/-	<2.0/-	<2.0/ <2.0
G14	7.0/8.6	5.3/5.7	6.5/6.4	<6.5/1.8	4.1/3.1
C15	5.4/5.4	6.6/6.4	6.3/6.5	6.0/5.5	5.9/5.7
T16	9.4/9.1	5.7/6.1	4.9/4.3	<2.0/ <2.0	1.4/2.7
<u>C17</u>	8.4/8.4	5.7/4.6	6.0/6.5	2.0/ <2.0	2.6/1.6
<u>T18</u>	8.9/9.2	5.9/5.2	5.7/-	1.2/-	2.1/-
<u>A19</u>	8.8/-	4.4/-	5.0/-	<2.0/-	<2.0/-
<u>A20</u>	8.9/9.2	5.4/3.9	5.5/5.5	<2.0/ <2.0	<2.0/3.0
<u>T21</u>	<b>9.2/7.2<sup>b</sup></b>	5.8/6.2	6.6/6.7	2.0/1.3	4.2/4.5
<u>G22</u>	9.6/8.4	4.9/-	4.5/-	<2.0/-	1.3/ <2.0
G23	9.7/10.1	4.4/4.6	4.8/-	4.8/-	<2.0/ <2.0
C24	7.3/7.8	5.6/5.9	6.0/-	3.1/-	4.5/-
T25	8.9/8.4	5.8/5.5	5.7/4.7	1.0/ <2.0	2.5/ <2.0
T26	8.4/-	6.3/-	6.6/5.1	2.3/1.0	3.7/2.5
T27	7.7/-	5.8/-	6.3/4.8	2.8/ <4.3	3.9/2.4
C28	- <sup>c</sup> /- <sup>c</sup>	4.9/4.0			

The first value is for the free DNA and the second entry is for the *Antp* HD-DNA complex.

<sup>a</sup> Labeled nucleotides in the partially labeled duplex are underlined.

<sup>b</sup> Bold if  $|\Delta^3J_{\text{HH}}(\text{free DNA} - \text{complex})| > 1.5$  Hz.

<sup>c</sup> No data given, since the chemical shifts of the geminal protons were degenerate or nearly degenerate.

calculation for the complex included a refinement of an *ad hoc* assumed standard B-DNA in the starting structures. In the present study we could thus investigate protein-induced conformational changes of the DNA by comparing high-quality NMR solution structures of both the free and the bound DNA. This yielded novel insights into subtle HD-induced local conformational changes, which extend present understanding of HD-DNA interactions. From the present experience, one can predict that isotope labeling of DNA fragments in their complexes with proteins will enable precise DNA structure determinations also in instances where protein binding causes major rearrangements of the nucleic acid conformation. This prediction receives support from recent demonstrations (Dingley & Grzesiek, 1998; Pervushin *et al.*, 1998) that  $^{15}\text{N}$ -labeling of nucleic acids enables direct identification of nucleotide pairing by scalar couplings across the Watson-Crick hydrogen bonds, and thus opens an avenue, for the first time, of direct NMR-identification of base-pairing in irregular nucleic acid structures.

## Materials and Methods

### Sample preparation and NMR spectroscopy

The NMR sample preparation of *Antp* HD-DNA complexes comprising the  $^{13}\text{C}$ ,  $^{15}\text{N}$  doubly labeled DNA and the NMR experiments were performed as described (Fernández *et al.*, 1998; Szyperski *et al.*, 1997, 1998).

### Collection of structural constraints

The sequence-specific resonance assignments for the presently used DNA duplexes were as described (Fernández *et al.*, 1998).  $^1\text{H}$ - $^1\text{H}$  upper limit distance constraints for the DNA and the protein-DNA interface were derived from two 750 MHz 3D  $^{13}\text{C}$ -resolved [ $^1\text{H}$ ,  $^1\text{H}$ ]-NOESY spectra centered on the deoxyribose region and the base resonances, respectively, and from 750 MHz 2D [ $^1\text{H}$ ,  $^1\text{H}$ ]-NOESY spectra recorded in  $\text{H}_2\text{O}$  or  $^2\text{H}_2\text{O}$  (Anil-Kumar *et al.*, 1980). These newly determined distance constraints were added to the previously derived  $^1\text{H}$ - $^1\text{H}$  upper limit distance constraints for the protein in the *Antp* HD-DNA complex (Quian *et al.*, 1993a,b; Billeter *et al.*, 1993). The central 12 bp in the DNA duplex have typical imino proton exchange rates for non-terminal Watson-Crick base-pairs (Guéron & Leroy, 1995), as estimated from the relative intensities of the diagonal peaks and the cross-peaks at  $\omega_1(\text{H}_2\text{O})$  in a 2D [ $^1\text{H}$ ,  $^1\text{H}$ ]-NOESY spectrum recorded in  $\text{H}_2\text{O}$ . Accordingly, hydrogen bond constraints for these base-pairs were added to the input for the structure calculation.

$^3J_{\text{HH}}$  scalar coupling constants for the deoxyribose rings obtained from 3D HCH-COSY spectra (Szyperski *et al.*, 1998; Grzesiek *et al.*, 1995) (Table 2) were translated into constraints for the dihedral angles  $\nu_1$   $37(\pm 2)^\circ$  and  $\nu_2$   $-35(\pm 6)^\circ$  (Kim *et al.*, 1992; Saenger, 1984), which corresponds to a pseudorotation phase angle range of  $155(\pm 15)^\circ$  for the present situation with C3'-*endo* conformation in excess of 80%. For G5 in the complex, the  $^3J_{\text{HH}}$  couplings reflect rapid repuckering for the deoxyribose ring (Szyperski *et al.*, 1998), and therefore no  $\nu_1$  and  $\nu_2$  constraints were included in the input.

Dihedral angle constraints for  $\epsilon$  and  $\beta$  were obtained from  $^3J_{\text{C}2'\text{P}}$  (Szyperski *et al.*, 1997),  $^3J_{\text{H}3'\text{P}}$  (Tate *et al.*, 1995; Szyperski *et al.*, 1999a) and  $^3J_{\text{C}4'\text{P}}$  (Szyperski *et al.*, 1999a) using the corresponding Karplus curves (Plavec & Chatopadhyahya, 1995; Mooren *et al.*, 1994). In addition, dihedral angle constraints for  $\beta$  were derived from a qualitative evaluation of  $^3J_{\text{H}5'\text{P}}$  and  $^3J_{\text{H}5''\text{P}}$  couplings (Varani *et al.*, 1996). Dihedral angle constraints for  $\gamma$  were derived from the H4' resonance line-widths (Kim *et al.*, 1992).

For the free DNA duplex (for the *Antp* HD-DNA complex), a total of 766 (1555) upper limit distance constraints were used as input for DYANA (Güntert *et al.*, 1997) (570 (1359) NOE distance constraints, 56 (56) hydrogen bond constraints and 140 (140) ring closure constraints) and 196 (196) lower limit distance constraints (56 (56) hydrogen bond constraints and 140 (140) ring closure constraints). For the *Antp* HD-DNA complex, the 1555 upper limit distance constraints consist of 671 and 820 intramolecular distance constraints for the DNA and the protein, respectively, and 64 intermolecular DNA-protein distance constraints. For the free DNA duplex (for the *Antp* HD-DNA complex), 24 (16)  $^3J_{\text{C}2'\text{P}}$  couplings (Szyperski *et al.*, 1997) and 20 (17)  $^3J_{\text{H}3'\text{P}}$  couplings (Szyperski *et al.*, 1999a) were measured (see the Supplementary Material). Combination with 17 (18)  $^3J_{\text{C}4'\text{P}}$  couplings yielded a total of 16 (14) constraints for the backbone dihedral angle  $\beta$ . 56 (54) constraints for the angles  $\nu_1$  and  $\nu_2$  were derived from measurement of 135 (97)  $^3J_{\text{HH}}$  scalar couplings (Szyperski *et al.*, 1998) (Table 2). Additionally, a total of 18 dihedral angle constraints for  $\gamma$  were derived from measurement of H4' resonance line-widths (Kim *et al.*, 1992), and 12  $^3J_{\text{NH}\beta}$  couplings were estimated from 3D HNNHB (Archer *et al.*, 1991) for the *Antp* HD-DNA complex. These data yielded a total of 293 (622) dihedral angle constraints in the input for the final DYANA calculation (see Table 1).

### Calculation and analysis of the three-dimensional NMR structures

The intranucleotide and sequential upper limit distance constraints, and the scalar couplings were translated into dihedral angle constraints with the FOUND module (Güntert *et al.*, 1998) of DYANA, which performs a grid search for allowed conformations in the space spanned by the seven torsion angles and the two ring puckers describing a dinucleotide segment. FOUND provided also stereospecific assignments for some H5'/H5'' protons. Additional stereospecific assignments for H5'/H5'' protons were obtained with the GLOMSA module of DYANA during the refinement procedure (Güntert *et al.*, 1997). The input for the final DYANA structure calculation included constraints to close the sugar rings (C3'-O4' 2.28 Å; C4'-O4' 1.41 Å; C4'-C1' 2.40 Å; C5'-O4' 2.39 Å; and H4'-O4' 2.12 Å) (Luxor & Gorenstein, 1995). The structure calculations with DYANA made use of "pseudolinkers" to connect the protein and the two DNA chains. The final calculation was started with 300 randomized structures. The 20 DYANA conformers with the smallest residual target function values were subjected to restrained energy minimization in a water shell of 15 Å thickness, using the AMBER all-atom force-field (Cornell *et al.*, 1995) as implemented in the program OPAL (Luginbühl *et al.*, 1996). Helix parameters were determined with the program CURVES (Lavery & Sklenar, 1988). All color Figures were generated with the program MOLMOL (Koradi *et al.*, 1996).

## Acknowledgments

Financial support was obtained from the Schweizerischer Nationalfonds (Project 31.49047.96), from the ETH Zürich for a special grant within the framework of the Swiss-Japanese R & D Roundtable Collaboration, and by Crest (Core Research for Evolutional Science and Technology) of the Japan Science and Technology Corporation (JST). C.F. is indebted to the "Schweizerische Bundesstipendienkommission" for a fellowship.

## References

- Anil-Kumar, Ernst, R. R. & Wüthrich, K. (1980). A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun.* **95**, 1-6.
- Archer, S. J., Ikura, M., Torchia, D. A. & Bax, A. (1991). An alternative 3D NMR technique for correlating backbone  $^{15}\text{N}$  with side-chain  $\text{H}\beta$  resonances in larger proteins. *J. Magn. Reson.* **95**, 636-641.
- Billeter, M. (1996). Homeodomain-type DNA recognition. *Prog. Biophys. Mol. Biol.* **66**, 211-225.
- Billeter, M. & Wüthrich, K. (1993). Model studies relating nuclear magnetic resonance data with the three-dimensional structure of protein-DNA complexes. *J. Mol. Biol.* **234**, 1094-1097.
- Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990). Determination of the three-dimensional structure of the *Antennapedia* homeodomain from *Drosophila* in solution by  $^1\text{H}$  nuclear magnetic resonance spectroscopy. *J. Mol. Biol.* **214**, 183-197.
- Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1993). Determination of the nuclear magnetic resonance solution structure of an *Antennapedia* homeodomain-DNA complex. *J. Mol. Biol.* **234**, 1084-1093.
- Billeter, M., Güntert, P., Luginbühl, P. & Wüthrich, K. (1996). Hydration and DNA recognition by homeodomains. *Cell*, **85**, 1057-1065.
- Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Jr, Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117**, 5179-5197.
- Dingley, A. J. & Grzesiek, S. (1998). Direct observation of hydrogen bonds in nucleic acid base-pairs by internucleotide  $^2J_{\text{NN}}$ -couplings. *J. Am. Chem. Soc.* **120**, 8293-8297.
- Fernández, C., Szyperki, T., Ono, A., Iwai, H., Tate, S., Kainosho, M. & Wüthrich, K. (1998). NMR with  $^{13}\text{C}$ ,  $^{15}\text{N}$ -doubly-labeled DNA: the *Antennapedia* homeodomain complex with a 14mer DNA duplex. *J. Biomol. NMR*, **12**, 25-37.
- Fraenkel, E. & Pabo, C. O. (1998). Comparison of X-ray and NMR structures for the *Antennapedia* homeodomain-DNA complex. *Nature Struct. Biol.* **5**, 692-697.
- Gehring, W. J. (1987). Homeoboxes in the study of development. *Science*, **236**, 1245-1252.
- Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Pérez, D., Affolter, M., Otting, G. & Wüthrich, K. (1994). Homeodomain-DNA recognition. *Cell*, **78**, 211-223.
- Grzesiek, S., Kuboniwa, H., Hinck, A. P. & Bax, A. (1995). Multiple-quantum line narrowing for measurement of  $\text{H}^2\text{-H}^\beta$   $J$  couplings in isotopically enriched proteins. *J. Am. Chem. Soc.* **117**, 5312-5315.
- Guéron, M. & Leroy, J. L. (1995). Studies of base-pair kinetics by NMR measurement of proton exchange. *Methods Enzymol.* **261**, 383-413.
- Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol.* **273**, 283-298.
- Güntert, P., Billeter, M., Ohlenschläger, O., Brown, L. R. & Wüthrich, K. (1998). Conformational analysis of protein and nucleic acids fragments with the new grid search algorithm FOUND. *J. Biomol. NMR*, **12**, 543-548.
- Han, G. W., Kopka, M. L., Cascio, D., Grzeskowiak, K. & Dickerson, R. E. (1997). Structure of a DNA analog of the primer for HIV-1 RT second strand synthesis. *J. Mol. Biol.* **269**, 811-826.
- Hartmann, B. & Lavery, R. (1996). DNA structural forms. *Rev. Biophys.* **29**, 309-368.
- Hirsch, J. & Aggarwal, A. K. (1995). Structure of the even-skipped homeodomain complexed to AT-rich DNA: new perspectives on homeodomain specificity. *EMBO J.* **14**, 6280-6291.
- Kim, S.-G., Lin, L.-J. & Reid, B. R. (1992). Determination of nucleic acid backbone conformation by  $^1\text{H}$  NMR. *Biochemistry*, **31**, 3564-3574.
- Kissinger, C. R., Lia, B., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990). Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution: a framework for understanding homeodomain-DNA interactions. *Cell*, **63**, 579-590.
- Klem, J. D., Rould, M. A., Rajeev, A., Herr, W. & Pabo, C. O. (1994). Crystal structure of the Oct-1 POU domain bound to an octamer site: DNA recognition with tethered DNA-binding. *Cell*, **77**, 21-32.
- Koradi, R., Billeter, M. & Wüthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14**, 51-55.
- Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature of irregular nucleic acids. *Biomol. Struct. Dynam.* **6**, 63-91.
- Li, T., Stark, M. R., Johnson, A. D. & Wolberger, C. (1995). Crystal structure of the MATA1/MAT $\alpha$ 2 homeodomain heterodimer bound to DNA. *Science*, **270**, 262-269.
- Luginbühl, P., Güntert, P., Billeter, M. & Wüthrich, K. (1996). The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. *J. Biomol. NMR*, **8**, 136-146.
- Luxor, B. A. & Gorenstein, D. G. (1995). Comparison of X-ray and NMR-determined nucleic acid structures. *Methods Enzymol.* **261**, 45-89.
- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A. & Gehring, W. J. (1984). A conserved DNA sequence in homeotic genes of the *Drosophila Antennapedia* and *bithorax* complexes. *Nature*, **308**, 428-433.
- Mooren, M. M. W., Wijmenga, S. S., van der Marel, G. A., van Boom, J. H. & Hilberts, C. W. (1994). The solution structure of the circular trinucleotide cr(GpGpGp) determined by NMR and molecular mechanics calculations. *Nucl. Acids Res.* **22**, 2658-2666.
- Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988). Isolation and

- sequence-specific DNA binding of the *Antennapedia* homeodomain. *EMBO J.* **7**, 4299-4304.
- Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990). Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance in solution. *EMBO J.* **9**, 3085-3092.
- Pervushin, K., Ono, A., Fernández, C., Szyperski, T., Kainosho, M. & Wüthrich, K. (1998). NMR scalar couplings across Watson-Crick base-pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. *Proc. Natl Acad. Sci. USA*, **95**, 14147-14151.
- Plavec, J. & Chattopadhyaya, J. (1995). Parametrization of Karplus equation relating  $^3J_{C-C-O-P}$  to torsion angle. *Tetrahedron Letters*, **36**, 1949-1952.
- Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989). The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution: comparison with prokaryotic repressors. *Cell*, **59**, 573-580.
- Qian, Y. Q., Otting, G., Billeter, M., Müller, M., Gehring, W. J. & Wüthrich, K. (1993a). NMR spectroscopy of a DNA complex with the uniformly  $^{13}C$ -labeled *Antennapedia* homeodomain and structure determination of the DNA-bound homeodomain. *J. Mol. Biol.* **234**, 1070-1083.
- Qian, Y. Q., Otting, G. & Wüthrich, K. (1993b). NMR detection of hydration water in the intermolecular interface of a protein-DNA complex. *J. Am. Chem. Soc.* **115**, 1189-1190.
- Saenger, W. (1984). *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
- Scott, M. P. & Weiner, A. J. (1984). Structural relationship among genes that control development: Sequence homology between the *Antennapedia*, *Ultrabithorax* and *fushi tarazu* loci of *Drosophila*. *Proc. Natl Acad. Sci. USA*, **81**, 4115-4119.
- Scott, M. P., Tamkun, J. W. & Hartzell, G. W., III (1989). The structure and function of the homeodomain. *BBA Rev. Cancer*, **989**, 25-49.
- Szyperski, T., Ono, A., Fernández, C., Iwai, H., Tate, S., Wüthrich, K. & Kainosho, M. (1997). Measurement of  $^3J_{C2'P}$  scalar couplings in a 17 kDa complex with  $^{13}C,^{15}N$ -labeled DNA distinguishes the  $B_I$  and  $B_{II}$  phosphate conformations of DNA. *J. Am. Chem. Soc.* **119**, 9901-9902.
- Szyperski, T., Fernández, C., Ono, A., Kainosho, M. & Wüthrich, K. (1998). Measurement of deoxyribose  $^3J_{HH}$  scalar couplings reveals protein binding-induced changes in the sugar puckers of the DNA. *J. Am. Chem. Soc.* **120**, 821-822.
- Szyperski, T., Fernández, C., Ono, A., Wüthrich, K. & Kainosho, M. (1999a). The 2D  $\{^31P\}$ -spin-echo-difference constant-time- $[^{13}C,^1H]$ -HMQC experiment for simultaneous determination of  $^3J_{H3'P}$  and  $^3J_{C4'P}$  in  $^{13}C$ -labeled nucleic acids and their protein complexes. *J. Magn. Reson.* in the press.
- Szyperski, T., Götte, M., Billeter, M., Perola, E., Cellai, L., Heumann, H. & Wüthrich, K. (1999b). NMR structure of the chimeric hybrid duplex r(gcagggc)·r(gcca)d(CTGC) comprising the tRNA-DNA junction formed during initiation of HIV-1 reverse transcription. *J. Biomol. NMR*, **13**, 343-355.
- Tate, S., Kubo, Y., Ono, A. & Kainosho, M. (1995). Stereospecific measurements of the vicinal  $^1H$ - $^{31}P$  coupling constraints for the diastereotopic C5' methylene protons in a DNA dodecamer with a  $^{13}C/^2H$  doubly labeled residue. Conformational analysis of the torsion angle  $\beta$ . *J. Am. Chem. Soc.* **117**, 7277-7278.
- Tucker-Kellogg, L., Rould, M. A., Chambers, K. A., Ades, S. E., Sauer, R. T. & Pabo, C. O. (1998). Engrailed (Gln50  $\rightarrow$  Lys) homeodomain-DNA complex at 1.9 Å resolution: structural basis for enhanced affinity and altered specificity. *Structure*, **5**, 1047-1054.
- Varani, G., Aboul-ela, F. & Allain, F. H.-T. (1996). NMR investigation of RNA structure. *Prog. NMR Spectrosc.* **29**, 51-127.
- Widmer, H. & Wüthrich, K. (1987). Simulated two-dimensional NMR cross peak fine structures for  $^1H$  spin systems in polypeptides and polydeoxynucleotides. *J. Magn. Reson.* **74**, 316-336.
- Wilson, D. S., Guenter, B., Desplan, C. & Kurijan, J. (1995). High resolution crystal structure of a paired (PAX) class cooperative homeodomain dimer on DNA. *Cell*, **82**, 709-719.
- Wolberger, C. (1996). Homeodomain interactions. *Curr. Opin. Struct. Biol.* **6**, 62-68.
- Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D. & Pabo, C. O. (1991). Crystal structure of a *Mat*  $\alpha 2$  homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell*, **67**, 517-528.
- Wüthrich, K. (1986). *NMR of Proteins and Nucleic Acids*, Wiley, New York.

Edited by I. Tinoco

(Received 31 March 1999; received in revised form 2 July 1999; accepted 2 July 1999)