Use of $^{13}C,^{15}N$-Labeled DNA in a Non-Sequence-Specific Protein−DNA Complex Resolves Ambiguous Assignments of Intermolecular NOEs

James E. Masse,[†] Frédéric H-T. Allain,[†] Yi-Meng Yen,[§] Reid C. Johnson,[‡,§] and Juli Feigon*[,†,‡]

Department of Chemistry and Biochemistry Molecular Biology Institute and Department of Biological Chemistry University of California, Los Angeles, California 90095-1569

Received November 18, 1998
Revised Manuscript Received February 17, 1999

To date, NMR structures of protein−DNA complexes have been determined by using solely homonuclear data[1] or with only the protein component isotopically labeled.[2−4] All of these complexes have involved sequence-specific DNA binding proteins which make specific contacts with the DNA bases, with one recent exception.[4] The recent development of protocols for the production of uniformly $^{15}N,^{13}C$-labeled DNA in sufficient quantity for multidimensional NMR[5,6] now makes the acquisition of the complementary data possible. We have applied heteronuclear multidimensional NMR spectroscopy to study the interaction of the 93 amino acid yeast Non-Histone Protein 6A (NHP6A),[7,8] a nonsequence specific HMG box[9] protein, with the 15 base-pair DNA duplex d(GGGGTGATTGTTCAG)•d(CTGAACAATC-ACCCC). The DNA sequence contains the seven nucleotide recognition sequence of the sequence-specific HMG box protein SRY (a sex-determining factor in mammals),[10] and was chosen to compare the mode of binding of a sequence-specific and non-sequence-specific HMG box protein on the same DNA sequence. Assignment of intermolecular NOEs in protein−DNA complexes is often particularly difficult to obtain due in part to line broadening at the protein DNA interface, in addition to spectral overlap especially for the deoxyribose resonances. Here we show that the use of isotopically labeled DNA in this protein−DNA complex was essential to obtain unambiguous assignment of intermolecular NOEs which were unresolved in the spectra of the complex using uniformly $^{15}N,^{13}C$-labeled protein alone. These NOEs, which could not have been assigned without the use of labeled DNA, allowed a precise positioning of the protein on the DNA. This is the first time that intermolecular NOEs have been reported and a precise side-chain localization achieved in a protein−DNA complex with a non-sequence-specific HMG box protein, despite several attempts.[11] Surprisingly, the protein binds at a different site and in an opposite orientation on the DNA to what is observed for the sequence-specific proteins.

Structure determination of the free NHP6A protein reveals that it has the expected L-shaped structure of an HMG domain protein, with the unusual feature of a kink in helix 3.[12] For the NMR investigation of the NHP6A−DNA complex, four different samples were prepared, one with $^{13}C,^{15}N$-labeled NHP6A[8] complexed with unlabeled DNA and three with $^{13}C,^{15}N$-labeled NHP6A complexed with the SRY−DNA uniformly $^{13}C,^{15}N$-labeled in one strand, the other strand, or both strands. The labeled DNA was prepared by enzymatic synthesis with Taq polymerase as previously described.[6] Although the majority of the base and deoxyribose DNA resonances in the complex could be assigned with homonuclear data only, most of the H4′, H5′, and H5″ could not be unambiguously assigned. Analysis of the 3D $^{1}H-^{13}C$ NOESY-HMQC[13] and HCCH-TOCSY[14] spectra run on the three complexes with $^{15}N,^{13}C$-labeled DNA samples resolved all of the ambiguities and made it possible to obtain complete resonance assignments of the DNA in complex with NHP6A.

To observe intermolecular NOEs between the protein and the DNA in the complex, 3D double-half-filtered $^{1}H-^{13}C$ HMQC-NOESY[15,16] experiments were run on all four samples. Figure 1A shows a slice of such a 3D spectrum from the sample where only the protein is $^{13}C,^{15}N$-labeled. The slice shows several intermolecular NOEs from the side chain methyl of Leu 25 of NHP6A to several DNA sugar resonances: between 5 and 6 ppm to two H1's and/or Cyt H5's (weak), between 4.6 and 5 ppm to two H3's (very weak), and between 3.8 and 4.4 ppm to an unresolved set of H4′, H5′, and H5″ resonances (medium). The H4′, H5′, and H5″ from *at least 20 nucleotides* resonate at the frequency of this latter cross-peak. This overlap could not be resolved without the use of labeled DNA. As shown in Figure 1B, the same sets of intermolecular NOEs are observed in the 3D double-half-filtered HMQC-NOESY spectrum recorded on the complex with uniformly $^{13}C,^{15}N$-labeled SRY DNA at the slice corresponding to the ^{1}H frequency of Leu 25 side chain methyl (0.34 ppm). However, in this case the ^{13}C dimension

* To whom correspondence should be addressed.
[†] Department of Chemistry and Biochemistry.
[‡] Molecular Biology Institute.
[§] Department of Biological Chemistry.

(1) Boelens, R.; Scheek, R. M.; van Boom, J. H.; Kaptein, R. *J. Mol. Biol.* **1987**, *193*, 213-216.
(2) (a) Billeter, M.; Qian, Y. Q.; Otting, G.; Muller, M.; Gehring, W.; Wüthrich, K. *J. Mol. Biol.* **1993**, *234*, 1084−1094. (b) Omichinski, J. G.; Clore, G. M.; Schaad, O.; Felsenfeld, G.; Trainor, C.; Appella, E.; Stahl, S. J.; Gronenborn, A. M. *Science* **1993**, *261*, 438−446. (c) Zhang, H.; Zhao, D. Q.; Revington, M.; Lee, W. T.; Jia, X.; Arrowsmith, C.; Jardetzky, O. *J. Mol. Biol.* **1994**, *238*, 592−614. (d) Ogata, K.; Morikawa, S.; Nakamura, H.; Sekikawa, A.; Inoue, T.; Kanai, H.; Sarai, A.; Ishii, S.; Nishimura, Y. *Cell* **1994**, *79*, 639−648. (e) Werner, M. H.; Clore, G. M.; Fisher, C. L.; Fisher, R. J.; Trinh, L.; Shiloach, J.; Gronenborn, A. M. *J. Biomol. NMR* **1997**, *10*, 317−328. (f) Starich, M. R.; Wikstrom, M.; Arst, H. N.; Clore, G. M.; Gronenborn, A. M. *J. Mol. Biol.* **1998**, *277*, 605−620. (g) Wuttke, D. S.; Foster, M. P.; Case, D. A.; Gottesfeld, J. M.; Wright, P. E. *J. Mol. Biol.* **1997**, *273*, 183−206. (h) Huth, J. R.; Bewley, C. A.; Nissen, M. S.; Evans, J. N. S.; Reeves, R.; Gronenborn, A. M.; Clore, G. M. *Nature Struct. Biol.* **1997**, *4*, 657−665. (i) Omichinski, J. G.; Pedone, P. V.; Felsenfeld, G.; Gronenborn, A. M.; Clore, G. M. *Nature Struct. Biol.* **1997**, *4*, 122−132. (j) Allen, M. D.; Yamasaki, K.; Ohme-Takagi, M.; Tateno, M.; Suzuki, M. *EMBO J.* **1998**, *17*, 5484−5496.
(3) Love, J. J.; Li, X. A.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. *Nature* **1995**, *376*, 791−795.
(4) Agback, P.; Baumann, H.; Knapp, S.; Landenstein, R.; Hard, T. *Nature Struct. Biol.* **1998**, *5*, 579−584.
(5) (a) Louis, J. M.; Martin, R. G.; Clore, G. M.; Gronenborn, A. M. *J. Biol. Chem.* **1998**, *273*, 2374−2378. (b) Mer, G.; Chazin, W. J. *J. Am. Chem. Soc.* **1998**, *120*, 607−608. (c) Ono, A.; Tate, S.; Ishido, Y.; Kainosho, M. *J. Biomol. NMR* **1994**, *4*, 581−586. (d) Zimmer, D. P.; Crothers, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 3091−3095. (e) Smith, D. E.; Su, J. Y.; Jucker, F. M. *J. Biomol. NMR* **1997**, *10*, 245−253. (f) Kainosho, M. *Nature Struct. Biol.* **1997**, *Suppl. 4*, 858−871. (g) Fernandez, C.; Szyperski, T.; Ono, A.; Iwai, H.; Tate, S.-I.; Kainosho, M.; Wüthrich, K. *J. Biomol. NMR* **1998**, *12*, 25−37.
(6) Masse, J.; Bortmann, P.; Dieckmann, T.; Feigon, J. *Nucleic Acids Res.* **1998**, *26*, 2618−2624.
(7) Kolodrubetz, D.; Burgum, A. *J. Biol. Chem.* **1990**, *265*, 3234−3239.
(8) Yen, Y.-M.; Wong, B.; Johnson, R. C. *J. Biol. Chem.* **1998**, *273*, 4424−4435.
(9) Landsman, D.; Bustin, M. *Bioessays* **1993**, *15*, 539−546.
(10) Haqq, C. M.; King, C. Y.; Donahoe, P. K.; Weiss, M. A. *Proc. Natl. Acad. Sci. U.S.A.* **1993**, *90*, 1097−1101.
(11) (a) Hardman, C. H.; Broadhurst, R. W.; Raine, A. R. C.; Grasser, K. D.; Thomas, J. O.; Laue, E. D. *Biochemistry* **1995**, *34*, 16596−16607. (b) Churchill, M. E. A.; Jones, D. N. M.; Glaser, T.; Hefner, H.; Searles, M. A.; Travers, A. A. *EMBO J.* **1995**, *14*, 1264−1275. (c) Berners-Price, S. J.; Corazza, A.; Guo, Z. J.; Barnham, K. J.; Sadler, P. J.; Ohyama, Y.; Leng, M.; Locker, D. *Eur. J. Biochem.* **1997**, *243*, 782−791.
(12) Allain, F. H.-T.; Yen, Y.-M.; Masse, J. E.; Schultze, P.; Dieckmann, T.; Johnson, R. C.; Feigon, J. *EMBO J.* In press.
(13) Marion, D.; Kay, L. E.; Sparks, S. W.; Torchia, D. A.; Bax, A. *J. Am. Chem. Soc.* **1989**, *111*, 1515−1517.
(14) Bax, A.; Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Ikura, M.; Kay, L. E. *J. Magn. Reson.* **1990**, *87*, 620−627. (b) Nikonowicz, E. P.; Pardi, A. *J. Mol. Biol.* **1993**, *232*, 1141−1156.
(15) Lee, W.; Revington, M. J.; Arrowsmith, C.; Kay, L. E. *FEBS Lett.* **1994**, *350*, 87−90.
(16) Otting, G.; Wüthrich, K. *Q. Rev. Biophys.* **1990**, *23*, 39−96.

10.1021/ja9839926 CCC: $18.00
Published on Web 03/26/1999

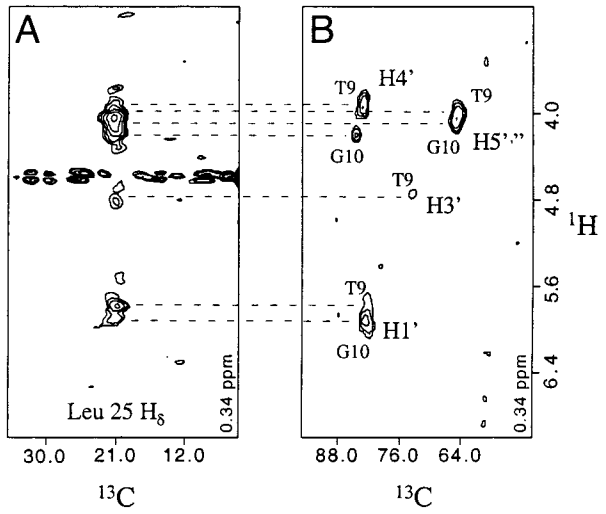

Figure 1. Planes from 600-MHz ^{13}C double-half-filtered HMQC-NOESY[15] spectra of the NHP6A-SRY DNA complex: (A) F1 plane at 0.34 ppm of ^{13}C,^{15}N-labeled NHP6A bound to unlabeled SRY DNA and (B) F3 plane at 0.34 ppm of ^{13}C,^{15}N-labeled NHP6A bound to ^{13}C,^{15}N-labeled SRY DNA. Seven of the eight assigned NOEs are labeled in part B. An additional very weak L25 methyl-G10 H3′ cross-peak is not visible at the contour levels shown. Samples are 2 mM in protein−DNA complex at 37 °C in 10 mM NaCl, 2 mM phosphate pH 5.5 in 99.9% D_2O. For part A, the spectral widths in F1, F2, and F3 were 6000, 10500, and 5700 Hz, respectively. 160 increments were acquired in t1 and 68 in t2 in States-TPPI mode with 16 scans and 512 complex points in t3. The spectra were processed with 512*128*512 complex points after apodization with a shifted squared sine bell (shifted by $\pi/4$ in t1 and t2, and $\pi/3$ in t3). Spectra for part B were acquired and processed as for those in part A, except the spectral widths in F1, F2, and F3 were 6000, 9100, and 5900 Hz, respectively, 168 increments were acquired in t1, and 68 in t2.

resolves the sets of intermolecular NOEs found between 3.8 and 4.4 ppm into four unambiguously assignable intermolecular NOE cross-peaks between Leu 25 methyl to two H4′ (at 4.18 and 4.02 ppm) and two H5′,H5″ pairs (Figure 1B). In addition, the two cross-peaks between 5 and 6 ppm can be unambiguously assigned to H1′ rather than Cyt H5 resonances. As indicated in Figure 1B, the 8 intermolecular NOEs are assigned to the H1′, H3′, H4′, and H5′,H5″ of T9 and G10, which are the only two sugar spin systems that can simultaneously match this ensemble of resonances.

The use of NMR data from complexes prepared with ^{15}N,^{13}C-labeled DNA was essential not only for unambiguous assignment of intermolecular NOEs involving the deoxyribose 4′, 5′, 5″ resonances, but also for precisely localizing the Leu 25 side-chain, since eight intermolecular distance constraints could be calibrated and used for the structure calculations versus four if only protein isotopic labeling were used. The strong cross-peaks from Leu 25 methyl to the H5′, H5″ and H4′ resonances of the deoxyribose localize Leu 25 in the minor groove of the DNA, roughly equidistant from the sugar rings of T9 and G10 (Figure 2). In contrast, for the SRY and LEF1 protein−DNA complexes, the residue equivalent to Leu 25 was localized between A22 and A23, on the opposite strand of the helix and one base pair displaced. A more complete analysis of the NHP6A−DNA complex

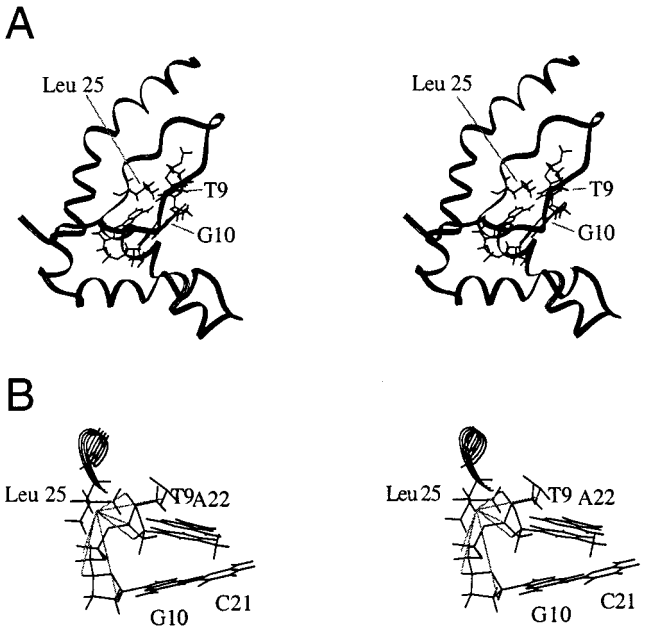

Figure 2. Stereoviews of the contacts between Leu 25 and T9 and G10 observed in a model[12] of the NHP6A-SRY DNA complex of the structure of the free NHP6A protein,[12] the 8 intermolecular NOEs obtained from the spectra of Figure 1, and the structure of the sequence-specific HMG box LEF1-DNA complex.[3] (A) Shown in the context of the protein structure represented by its backbone ribbon. (B) Close-up view of the interaction. The dotted lines illustrate the distance constraints derived for the 8 intermolecular NOEs shown assigned in Figure 1.

compared to the sequence-specific HMG box proteins SRY[17] and LEF1[3] will be presented elsewhere.[12]

In summary, the use of isotopically labeled DNA in a protein−DNA complex provides complementary information to that obtained with isotopically labeled proteins in double-half-filtered HMQC-NOESY experiments. This not only provides confirmation of the intermolecular NOEs obtained from spectra of complexes with isotopically labeled protein, which is important since double-half-filtered and half-filtered spectra tend to have some artifacts, but also allows unambiguous assignment of intermolecular cross-peaks which could not otherwise be resolved, especially in the crowded H4′, H5′−H5″ region, due to spectral overlap in both the carbon and proton dimensions. This is particularly essential for non-sequence-specific DNA−protein interactions, since most of the protein−DNA contacts would be expected to involve the DNA backbone and few, if any, base contacts are expected to be present. In the example presented here, spectra obtained on the complex with ^{13}C,^{15}N-labeled DNA led to unambiguous assignment of intermolecular NOEs involving the phosphodiester backbone that could never have been obtained with ^{13}C,^{15}N-labeled protein alone. This provided additional constraints which could be included in structure calculations, making it possible for the first time to precisely localize a non-sequence-specific HMG−box protein onto DNA.

Acknowledgment. This work was supported by NIH grants GM48123 to J.F. and GM38509 to R.C.J. and European Molecular Biology Organization and Human Frontiers Sciences Program Organization postdoctoral fellowships to F.H.T.A. The authors thank Dr. Thorsten Dieckmann for assistance with the NMR spectroscopy.

JA9839926

(17) Werner, M. H.; Huth, J. R.; Gronenborn, A. M.; Clore, G. M. *Trends Biochem. Sci.* **1996**, *21*, 302−308.