*Proc. Natl. Acad. Sci. USA*
Vol. 89, pp. 10738–10742, November 1992
Biophysics

# NMR structure determination reveals that the homeodomain is connected through a flexible linker to the main body in the *Drosophila* Antennapedia protein

(Antennapedia homeodomain/protein–DNA interactions/transcriptional regulation/transverse-relaxation time)

YAN QIU QIAN*, GOTTFRIED OTTING*, KATSUO FURUKUBO-TOKUNAGA†, MARKUS AFFOLTER†, WALTER J. GEHRING†, AND KURT WÜTHRICH*

*Institut für Molekularbiologie und Biophysik, Eidgenössiche Technische Hochschule-Hönggerberg CH-8093 Zürich, Switzerland; and †Biozentrum der Universität Basel, Abteilung Zellbiologie, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

*Contributed by Walter J. Gehring, August 10, 1992 (received for review June 5, 1992)*

ABSTRACT **The secondary structure of an N-terminally elongated Antennapedia (Antp) homeodomain (HD) polypeptide containing residues −14 to 67, where residues 1–60 constitute the HD, has been determined by NMR in solution. This polypeptide contains the conserved motif -Tyr-Pro-Trp-Met- (YPWM) at positions −9 to −6. Despite the hydrophobic nature of this tetrapeptide motif, the N-terminal arm consisting of residues −14 to 6 is flexibly disordered, and the well-defined part of the HD structure with residues 7–59 is indistinguishable from that of the shorter Antp HD polypeptide (where positions 0, 1, and 67 are methionine, arginine, and glycine, respectively). *In vitro* biochemical studies showed that the stability and specificity of the DNA binding previously observed for the shorter Antp HD polypeptide is preserved in the elongated polypeptide. These results strongly support the view that the HD is connected through a flexible linker to the main body in the Antp protein and that the minor groove contacts by the N-terminal arm (residues 1–6) in the Antp HD–DNA complex are an intrinsic feature of the DNA-binding interactions of the intact Antp protein.**

The identification of the homeobox in developmental regulatory genes of eukaryotes (1–3) was followed by numerous studies on homeodomain (HD) structure and gene regulatory functions. Homeobox sequences were detected in species ranging from yeast to humans (4, 5), and HD-containing proteins were found to act as transcriptional regulators in which the HD is responsible for sequence-specific DNA recognition (6, 7). The three-dimensional NMR solution structure of the Antp HD (8) contains four helices, whereby helix 2 and helix 3 form a helix–turn–helix motif. The three-dimensional structures of three HD–DNA complexes (Antp, engrailed, and MATα2) determined by NMR (9) or by x-ray crystallography (10, 11) show the same protein fold and similar HD–DNA contacts. The second helix of the helix–turn–helix motif is located in the major groove of the DNA, and further protein–DNA contacts are present between the N-terminal heptapeptide segment and the minor groove of the DNA. Comparison of the different complexes suggests that, overall, a relatively small subset of conserved amino acid residues is responsible for defining the characteristic DNA-binding mode of HDs (11).

Comparison of the protein sequences revealed that, in addition to the HD, many of the *Drosophila* homeotic gene products contain another highly conserved peptide region, the -Tyr-Pro-Trp-Met- (YPWM) motif, which is separated from the N terminus of the HD by various distances (ref. 5 and Table 1). Strikingly, vertebrate *HOX* genes, which are very similar to *Drosophila* homeotic genes in several aspects (for a review, see ref. 22), also encode the YPWM sequence (23). The conservation of the YPWM sequence in HD-containing gene products of such diverse species indicates that this peptide region may also play structural and/or functional roles in developmental gene regulation.

None of the HD-containing polypeptides so far used for structure determinations (8–11) includes the tetrapeptide YPWM. In this project we investigated an elongated Antp protein fragment, Antp(YPWM), that corresponds to the polypeptide segment with residues 279–359 of the intact Antp protein containing the YPWM peptide and the complete HD (see Fig. 2). The experiments with the Antp(YPWM) polypeptide described in this paper were designed for a dual purpose, (*i*) to obtain evidence for possible structural and functional roles of the conserved tetrapeptide motif YPWM and (*ii*) to investigate whether the length of the N-terminal extension has any impact on the previously defined behavior of the HD N-terminal arm that makes direct minor groove protein–DNA contacts in the Antp HD–DNA complex (9).

## MATERIALS AND METHODS

**Cloning of pAop2CS-YPWM.** The plasmid used to express the elongated YPWM-containing Antp polypeptide was constructed using overlap-extension PCR (polymerase chain reaction) techniques. As the presence of cysteinyl residues may cause difficulties due to autooxidation (24), the following strategy was employed in cloning a DNA fragment from which the Antp(YPWM) polypeptide could be expressed. The N-terminal sequences coding for an Antp-derived methionine, a 13-amino acid N-terminal extension of the HD (containing the YPWM sequence), and the first 17 amino acid residues of the HD itself were derived from the Antp cDNA 303 described in Schneuwly *et al.* (12). The C-terminal part of the Antp HD was derived from the plasmid pAop2CS (ref. 25; see also ref. 24). The N-terminal sequences were amplified from Antp cDNA 303 by using the 5′ primer 5′-TCG-CAGTCCTCGCATATGCCGTCTC-3′ and the 3′ primer 5′-CTCTAGCTCTAGAGTCTGGTACCGGGTGTA-3′. The C-terminal sequences were independently amplified, starting from pAop2CS and using the 5′ primer 5′-TACACCCGG-TACCAGACTCTAGAGCTAGAG-3′ and the 3′ primer 5′-AGCAGCCGGTTAACCCGGCTC-3′. Subsequently, the two amplified fragments were purified and ligated by overlap extension PCR using the two outer primers. The resulting DNA fragment was cut at the ends with *Nde* I and *Hpa* I and

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HD, homeodomain; NOE, nuclear Overhauser effect; NOESY, two-dimensional NOE spectroscopy; COSY, two-dimensional correlated spectroscopy; TOCSY, two-dimensional total correlation spectroscopy.

Table 1. Conservation of the YPWM tetrapeptide segment in *Drosophila* gene products

| Gene | Sequence | Linker peptide | Ref(s). |
|---|---|---|---|
| *Antp* | PLYPWMRS | 5/9 | 12 |
| *pb* | -E----KE | 19–29 | 13 |
| *Dfd* | II----KK | 18 | 14 |
| *Scr* | QI----KR | 15 | 15 |
| *Ubx* | TF----AI | 8–51 | 16, 17 |
| *abd-A* | -R----TL | 25 | 18 |
| *lab* | -T-K--QL | 105/111 | 19 |
| *ems* | Q----LL- | 23 | 20 |
| *cad* | -YFD--KK | 16 | 21 |

Amino acids identical to those in Antp are indicated by a dash. Linker peptide refers to the number of amino acid residues separating the YPWM motif and the N terminus of the HD. For *Antp* and *lab*, two transcripts that differ in the length of the linker peptide have been identified. For *pb* and *Ubx*, several transcripts that are the result of alternative splicing between the YPWM motif and the HD have been identified. The numbers shown indicate the shortest and the longest linker peptides identified for these two gene products.

cloned in pAop2CS from which the natural *Nde* I–*Hpa* I insert had previously been removed. The structure of the final plasmid was verified by DNA sequencing using the dideoxynucleotide chain-terminator method (26).

**Expression and Purification of the Antp(YPWM) Polypeptide.** The Antp(YPWM) polypeptide was expressed in *Escherichia coli* using a T7 expression system (27). BL21(DE3) LysE was used as a host. Cells were grown in 4 liters of super broth (12 g of Bacto tryptone/24 g of yeast extracts/5 ml of glycerol/2.3 g of $KH_2PO_4$/12.5 g of $K_2HPO_4$ per liter with ampicillin at 0.1 mg/ml) at 25°C in a high-density fermenter (Lab-Line Instruments). After reaching an OD value of 5 at 595 nm, the culture was induced with 8 ml of isopropyl β-D-thiogalactopyranoside (100 mg/ml) and allowed to grow overnight at 25°C. Cells were harvested by centrifugation and lysed in a lysis buffer [50 mM Tris·HCl, pH 8.0/10 mM EDTA/10% (wt/vol) sucrose/0.2 M KCl/0.5% Triton X-100/1 mM phenylmethanesulfonyl fluoride/30 mM spermidine/1 mM dithiothreitol/lysozyme (0.4 mg/ml)] on ice for 1 h followed by a high salt extraction in 0.6 M KCl at 4°C for 30 min. Cell debris and the chromosomal DNA were removed by centrifugation at 30,000 rpm in a 50 T; Beckman rotor at 4°C for 2 h, and the clear supernatant was dialyzed overnight against a buffer solution (50 mM potassium phosphate, pH 7.0/1 mM dithiothreitol/1 mM phenylmethylsulfonyl fluoride/0.5 M KCl). The HD was retained and washed on a Bio-Rex 70 column with the above buffer and eluted with a salt gradient of 0.5–1.5 M KCl in the same buffer. The peak fractions were pooled and diluted twice with the buffer solution. The protein was further purified through a Mono S 10/10 FPLC (Pharmacia) column with a gradient elution from 0.4 to 1.0 M KCl in the above buffer. After extensive dialysis against pure $H_2O$, the purified protein fraction was further desalted by ultrafiltration and lyophilized. About 50 mg of pure protein was recovered from 4 liters of culture in the induced super broth. The protein yield was determined by measuring the absorption at 280 nm by using the extinction coefficients for tryptophan and tyrosine [$2.22 \times 10^4$ $M^{-1} \cdot cm^{-1}$ for the Antp(YPWM) polypeptide].

**NMR Measurements and Data Processing.** NMR spectra were recorded on Bruker AMX 500 and AMX 600 spectrometers by using 9 mM Antp(YPWM) polypeptide in a mixed solvent of 90% $H_2O$/10% $^2H_2O$ at pH 4.3 and 20°C. A two-dimensional nuclear Overhauser effect spectroscopy (NOESY) spectrum (28) with a mixing time of 100 ms and a clean two-dimensional total correlation spectroscopy (TOCSY) spectrum (29) with a mixing time of 60 ms recorded at 500 MHz were used to obtain complete sequence-specific $^1H$ NMR assignments of the protein. For a detailed comparison between the three-dimensional structures of the Antp(YPWM) polypeptide and the previously studied Antp(C39S) HD, soft-NOESY spectra with a mixing time of 40 ms were recorded with a selective inversion pulse centered at $\omega_1 = 4.8$ ppm (30–32).

Amide proton exchange rates were determined from the time course of the signal intensity of cross peaks with the amide protons after dissolving the lyophilized protein in $^2H_2O$. A series of NOESY spectra with a 2-ms homospoil pulse in the mixing period to shorten the phase cycle were performed to obtain amide proton exchange data for the Antp(YPWM) polypeptide. These NOESY spectra were recorded at 600 MHz with a mixing time of 100 ms and a total recording time of 1.5 h for each experiment. Amide proton exchange data for the uniformly $^{15}N$-labeled Antp(C39S) HD were obtained from a series of two-dimensional [$^{15}N$,$^1H$]-correlated spectroscopy (COSY) spectra (33, 34) recorded at 500 MHz with a 3.3 mM protein sample and a total recording time of 10.5 min for each experiment.

A method was developed to measure the transverse relaxation times $T_2$ of the amide protons from a series of clean TOCSY spectra (29) with a constant mixing time preceded by different relaxation delays, $\tau$ (Fig. 1). This scheme is closely related to the $T_1$ measurements described by Arseniev *et al.* (35). The $T_2$ relaxation of the amide protons is reflected in the time course of the amide proton($\omega_1$)–α-proton($\omega_2$) and amide proton($\omega_1$)–β-proton($\omega_2$) cross peaks. To refocus both the chemical shifts and the scalar couplings of the amide protons, a selective inversion pulse was applied to the amide protons in the middle of the relaxation delay (Fig. 1). The selective pulse consisted of a series of 24 pulses separated by delays of 120 μs with the pulse widths proportional to the approximate sinc function calculated by Mao *et al.* (39) for optimum inversion. The total width of the selective pulse was 5 ms and its carrier frequency was set at 8.2 ppm. With this pulse, all amide protons could be inverted without significantly affect-

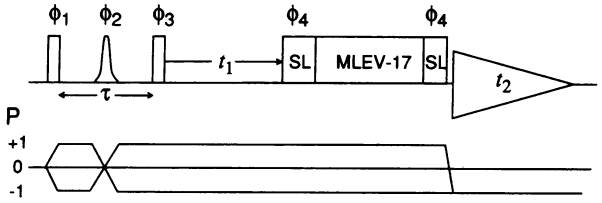


FIG. 1. Experimental scheme used for the measurement of the transverse amide proton relaxation times $T_2$. It consists of a clean TOCSY experiment preceded by a relaxation delay $\tau$. $T_2$ relaxation during the delay $\tau$ leads to reduced amounts of amide proton magnetization available at the beginning of the $t_1$ evolution and to correspondingly reduced intensities of the cross peaks arising from the amide proton magnetization present at the end of the evolution time. The $T_2$ relaxation times of the amide protons are determined from the cross-peak intensities recorded in a series of experiments with increasing waiting periods $\tau$. A selective pulse is applied to the amide protons in the middle of $\tau$ to refocus their chemical shifts and their scalar couplings with the α protons. The coherence transfer pathway for the amide protons is indicated below the pulse sequence. To avoid echo effects, the two spin-lock pulses (SL) are of different length, typically 2 ms and 500 μs. The $(\pi/2)_{\phi_3}$ pulse serves to purge any magnetization that is not aligned along the $y$ axis. Water suppression is achieved by selective irradiation of the water resonance during the delays of 1 s between successive scans and throughout the delay $\tau$. The phase cycling used was as follows: $\phi_1 = 8(x)$; $\phi_2 = 2(x, y, -x, -y)$; $\phi_3 = [4y, 4(-y)]$; $\phi_4 = 8(x)$; receiver = $4(x, -x)$. This phase cycle was extended to 16 steps by additional phase alternation of $\phi_4$ and the phase of the MLEV-17 mixing sequence (36) and further to 64 steps to include CYCLOPS by the simultaneous incrementation of all phases by $\pi/2$ (37). Time proportional phase incrementation may be applied to the pulses $\phi_1$, $\phi_2$, and $\phi_3$ for quadrature detection in $\omega_1$ (38).

ing the $\alpha$-proton resonances. As a result of the phase cycling of the selective inversion pulse (see Fig. 1), only the diagonal peaks of and the cross peaks with the amide protons appear in the spectrum. Four spectra were recorded at 600 MHz, using a mixing time of 60 ms and relaxation delays ($\tau$) of 5.2 ms, 35 ms, 65 ms, and 125 ms. Each spectrum took 9 h to record.

The amide proton exchange measurements and the $T_2$ relaxation measurements were analyzed using the same techniques. The program EASY (40) was used for peak picking and integration of the cross peaks, and all the spectra of a time series were processed identically. The rate constants for the amide proton exchange and the $T_2$ relaxation times were obtained with a nonlinear fit routine included in the program NOLIFI (unpublished program).

## RESULTS

Between the conserved tetrapeptide segment YPWM and the N terminus of the HD, the *Drosophila Antp* gene has two alternative splicing sites (12) producing two types of functional protein products that differ by the presence or absence of 4 amino acid residues between the YPWM motif and the HD. The 4-amino acid insert in the longer Antp protein contains a cysteinyl residue. As the presence of such residues may cause difficulties due to autooxidation (24), we used the shorter protein in this study.

**$^1$H NMR Assignments for the Elongated Antp(C39S) HD and Secondary Structure Determination.** Complete sequence-specific $^1$H NMR assignments were obtained using the standard sequential assignment strategy for small proteins (41–44). A survey of the sequential and medium-range NOE connectivities is presented in Fig. 2. With three exceptions, all pairs of sequentially neighboring amino acid residues are connected by at least one of the sequential connectivities $d_{NN}$, $d_{\beta N}$, or $d_{\alpha N}$ (for Xaa-Pro, $d_{\alpha\delta}$ instead of $d_{\alpha N}$). The exceptions are Leu-16–Glu-17, Ala-37–Leu-38, and Gln-44–Ile-45, where the two amide proton resonances of each pair of residues have virtually identical chemical shifts. The $^1$H chemical shifts and NOEs observed for the elongated HD are virtually identical to those for the corresponding protons of the 68-amino acid HD fragment Antp(C39S) (25). This is illustrated by Fig. 3, which shows the spectral region ($\omega_1$, 4.0–4.6 ppm; $\omega_2$, 7.1–8.4 ppm) from two soft NOESY spectra recorded with the Antp(C39S) HD (Fig. 3*A*) and with the Antp(YPWM) polypeptide (Fig. 3*B*), respectively. Nearly all cross peaks seen for the Antp(C39S) HD are also present in the spectrum of Antp(YPWM), with nearly identical chemical shifts and intensities. The absence in Fig. 3*B* of some of the weak cross peaks seen in Fig. 3*A* can be attributed to a somewhat different selection of the lowest contour levels plotted in the two spectra. The near identity of corresponding NOESY cross-peak patterns in the entire NOESY spectrum implies that nearly identical three-dimensional structures prevail for the well-ordered HD core of residues 7–59 (8) in the Antp(YPWM) polypeptide and in Antp(C39S). Additional intense cross peaks in Fig. 3*B* arise from the N-terminal elongation polypeptide segment of residues −1 to −13 in Antp(YPWM). Arrows in Fig. 3*B* identify the sequential assignment pathway for the residues −11 to 1 via the $d_{\alpha N}$ connectivities. In addition to the $d_{\alpha N}$ connectivities, quite strong sequential $d_{NN}$ connectivities were observed between the same residues (Fig. 2), which is indicative of flexibly disordered nonregular secondary structure (44). Furthermore, all cross peaks involving these N-terminal residues show narrow line shapes, which is again characteristic of relatively long transverse relaxation times due to increased mobility.

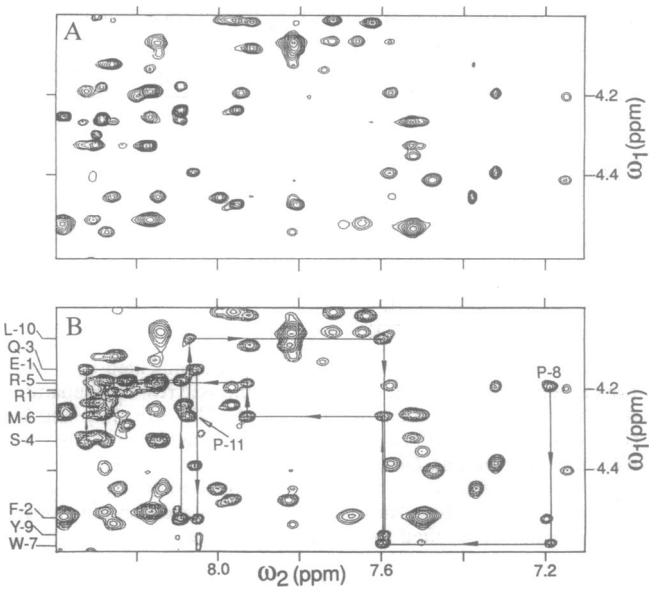


FIG. 3. Spectral region ($\omega_1$, 4.0–4.6 ppm; $\omega_2$, 7.1–8.4 ppm) of two-dimensional homonuclear soft-NOESY spectra (solvent 90% $H_2O$/10% $^2H_2O$, pH 4.3, 20°C; mixing time, 40 ms; $^1$H frequency, 600 MHz). (*A*) Antp(C39S) HD solution (11 mM). (*B*) Antp(YPWM) polypeptide solution (9 mM). The sequential assignment pathway via the $d_{\alpha N}$ connectivities is identified for the N-terminal residues −11 to 1 with arrows connecting the intraresidual and sequential $\alpha$H–NH cross peaks.

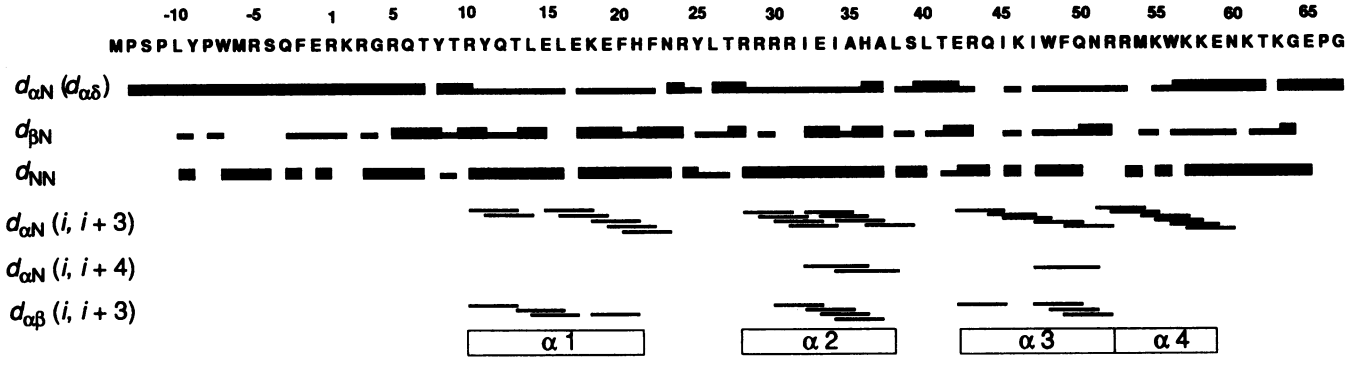


FIG. 2. Amino acid sequence of the Antp(YPWM) polypeptide and survey of sequential and medium-range NOE connectivities. Negative numbering is used for the amino acid residues preceding the HD, which consists of residues 1–60. Solid bars indicate sequential $d_{\alpha N}$, $d_{\beta N}$, and $d_{NN}$ connectivities, where the thickness of the bar reflects the intensity of the NOEs (for Xaa-Pro dipeptide segments, $d_{\alpha\delta}$ connectivities were drawn in the place of $d_{\alpha N}$ connectivities). Short medium-range distances $d_{\alpha N}(i, i + 3)$, $d_{\alpha N}(i, i + 4)$, and $d_{\alpha\beta}(i, i + 3)$ are indicated with lines between the two residues that are connected by the NOE. At the bottom, the locations of the four helices identified in the Antp HD (8) are given.

**Characterization of Internal Mobility in the Antp(YPWM) Polypeptide by Studies of the Amide Proton Exchange Rates and $T_2$ Relaxation Times.** To investigate whether any hydrogen bonds were formed involving amide protons of the N-terminal elongation peptide segment in Antp(YPWM) and whether the hydrogen-bonding network in the core of the protein was affected by the addition of the N-terminal elongation, quantitative measurements were made of the amide proton exchange rates in both the Antp(YPWM) polypeptide and the Antp(C39S) HD. Fig. 4 *A* and *B* shows plots of the backbone amide proton exchange rates $k_m$ vs. the amino acid sequence of Antp(C39S) and Antp(YPWM), respectively. [In the Antp(YPWM) polypeptide, the amide proton resonance of Ser-39 was not resolved, so that the exchange rate of this residue could not be measured.] Quite generally, more extensive and more accurate data were obtained for the Antp(C39S) HD than for the Antp(YPWM) polypeptide because Antp(C39S) was available in the uniformly $^{15}N$-labeled form and the two-dimensional [$^{15}N$,$^1H$]-COSY experiments performed are both faster and more sensitive than the two-dimensional [$^1H$,$^1H$]-NOESY experiments recorded with the unlabeled elongated polypeptide. Within the accuracy of the measurements, the amide proton exchange rates of corresponding residues are very similar for the two proteins. The patterns of slowly exchanging amide protons clearly reflect identical sequence locations of the helical secondary structures in both proteins. The gradual increase of the exchange rates toward the C-terminal end of helix IV is explained by the fact that helix IV points away from the protein and is, therefore, less well stabilized and protected from solvent contact (8). In the N-terminal segment containing residues −13 to +6 of the Antp(YPWM) polypeptide, no slowed amide proton exchange could be detected with the experiments used. This indicates that this N-terminal peptide segment is involved neither in the formation of stable local regular secondary structure nor in hydrogen bonding interactions with the structured core part of the protein.

Fig. 4*C* presents a plot of the amide proton $T_2$ relaxation times vs. the amino acid sequence of the Antp(YPWM) polypeptide. The $T_2$ values of residues 8–59 in the structured core are ≈20 ms, whereas the $T_2$ values of the amide protons in the N-terminal and C-terminal segments (residues −13 to 5 and 61 to 67) are ≈3 times longer. Since longer $T_2$ relaxation times reflect shorter effective rotational correlation times, we conclude that, compared to the structured core of the protein, significant additional segmental mobility on a nanosecond time scale prevails in the terminal polypeptide segments. The results obtained thus show in particular that a flexibly disordered state prevails for the entire N-terminal elongation of the polypeptide chain, including the conserved tetrapeptide segment YPWM.

## DISCUSSION

The close proximity of the YPWM motif to the DNA binding HD in a variety of homeotic proteins (Table 1) suggests the possibility that this tetrapeptide segment might be involved in some aspects of modulation of DNA binding. The present NMR investigations of the N-terminally elongated Antp(YPWM) polypeptide (Figs. 2–4) provided no direct indication of specific structural roles of this conserved tetrapeptide motif that could alter the DNA-binding properties of the Antp HD. In qualitative agreement with this conclusion, measurements of the apparent dissociation constant $K_d$ of the Antp(YPWM) polypeptide complex with the DNA binding site BS2-18 gave a value of 8.5 ± 2.2 nM, which is 5-fold higher than the value of 1.6 nM obtained with the Antp(C39S) HD (45). This indicates that the presence of the YPWM motif does not further stabilize the HD–DNA complex *in vitro*. Methylation and ethylation interference patterns with a further elongated YPWM-containing Antp HD polypeptide of 117 amino acid residues (45) showed no differences from those obtained with the 68-amino acid Antp(C39S) HD (A. Percival-Smith and W.J.G., unpublished observation). These experiments suggest that *in vitro* the YPWM motif neither increases the DNA-binding affinity nor changes the binding specificity of homeotic proteins. On the level of the molecular conformation, this can be rationalized by the present NMR studies, which provide no indication of a preferred nonrandom structure for this segment. Although it cannot be excluded that the 14-amino acid extension could be folded into a defined structure that requires sequences

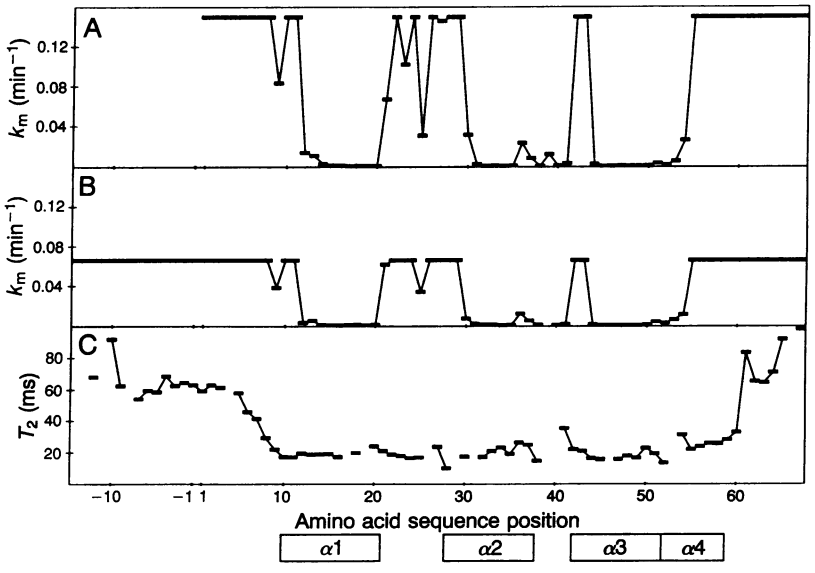


FIG. 4. (*A*) Backbone amide proton exchange rates ($k_m$) of the Antp(C39S) HD [3.3 mM sample of uniformly $^{15}N$-enriched Antp(C39S) HD, pH 4.3, 20°C; $^1H$ frequency, 500 MHz; [$^{15}N$,$^1H$]-COSY]. Exchange rates faster than 0.15 min$^{-1}$ could not be measured. (*B*) Backbone amide proton exchange rates ($k_m$) of the Antp(YPWM) polypeptide {9 mM sample of the Antp(YPWM) polypeptide, pH 4.3, 20°C; $^1H$ frequency, 600 MHz; [$^1H$,$^1H$]-NOESY}. Exchange rates faster than 0.07 min$^{-1}$ could not be measured. (*C*) $T_2$ relaxation times of the amide protons of the Antp(YPWM) polypeptide [9 mM sample of the Antp(YPWM) polypeptide, pH 4.3, 20°C; $^1H$ frequency, 600 MHz; experiment of Fig. 1]. At the bottom the locations of the four helices in the HD structure (8) are indicated.

N-terminal to it, the fact that the corresponding region connecting the N-terminal domain and the HD of the yeast MAT$\alpha$2 protein is highly susceptible to proteases (46) supports the view that this conserved motif is also flexible in the intact Antp protein. In the MAT$\alpha$2 protein, the flexibility in the linker between the N-terminal domain and the HD is crucial for the flexibility of the MAT$\alpha$2 dimer and plays an important role in target site selection (47).

When the modular nature of eukaryotic transcription regulatory sequences (for a review, see ref. 48) is considered, it seems likely that the specificity of homeotic gene products with respect to target site selection *in vivo* is also influenced by interactions with protein cofactors. This opens the possibility that the YPWM tetrapeptide motif is involved in protein–protein interactions (49). Its high degree of conservation might then reflect that different homeotic proteins interact with common cofactors. Interestingly, the equivalent part of the yeast MAT$\alpha$2 protein is the major protein–protein contact site between the MCM1 and MAT$\alpha$2 proteins (A. K. Vershon and A. D. Johnson, personal communication). *In vivo* assays of the functional properties of homeotic genes carrying mutations in the YPWM motif (see refs. 50–52) as well as biochemical studies will be required to further investigate possible roles of this "unstructured" polypeptide segment in DNA recognition, either by direct protein–DNA interactions or by mediation through protein–protein interactions.

All the three structure determinations of HD–DNA complexes reported so far (9–11) used short HD polypeptides, with at most one residue added at the N terminus of the HD sequence. Since the N-terminal arm of the HDs in these structures contacts bases in the minor groove of the DNA and has also been shown to significantly affect the complex stability (53), it was of considerable interest to ascertain that these N-terminal protein–DNA interactions are not a specific property of the polypeptides selected so far for the structure determinations. The results on the structure and the binding properties of the elongated Antp(YPWM) polypeptide now strongly support that the data collected with the shorter HD polypeptides are a valid representation of the properties of HDs incorporated into longer polypeptide chains, including most probably the intact expression products of the homeobox-containing genes.

We thank C. Bartels for the use of a modified form of the EASY program, P. Güntert for the program NOLIFI used to perform the nonlinear fits, G. Halder for sequencing the pAop2CS-YPWM plasmid insert, P. Baumgartner for help with the protein purification, and R. Marani for the careful processing of the manuscript. Financial support was obtained from the Schweizerischer National Fonds (Project 31.32033.91).

1. McGinnis, W., Levine, M., Hafen, E., Kuroiwa, A. & Gehring, W. J. (1984) *Nature (London)* **308**, 428–433.
2. McGinnis, W., Garber, R. L., Wirz, J., Kuroiwa, A. & Gehring, W. J. (1984) *Cell* **37**, 403–408.
3. Scott, M. P. & Weiner, A. J. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 4115–4119.
4. Gehring, W. J. (1987) *Science* **236**, 1245–1252.
5. Scott, M. P., Tamkun, J. W. & Hartzell, G. W. (1989) *Biochim. Biophys. Acta* **989**, 25–48.
6. Affolter, M., Schier, A. & Gehring, W. J. (1990) *Curr. Opin. Cell Biol.* **2**, 485–495.
7. Hayashi, S. & Scott, M. P. (1990) *Cell* **63**, 883–894.
8. Qian, Y. Q., Billeter, M., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1989) *Cell* **59**, 573–580.
9. Otting, G., Qian, Y. Q., Billeter, M., Müller, M., Affolter, M., Gehring, W. J. & Wüthrich, K. (1990) *EMBO J.* **9**, 3085–3092.
10. Kissinger, C. R., Liu, B. S., Martin-Blanco, E., Kornberg, T. B. & Pabo, C. O. (1990) *Cell* **63**, 579–590.
11. Wolberger, C., Vershon, A. K., Liu, B. S., Johnson, A. D. & Pabo, C. O. (1991) *Cell* **67**, 517–528.
12. Schneuwly, S., Kuroiwa, A., Baumgartner, P. & Gehring, W. J. (1986) *EMBO J.* **5**, 733–739.
13. Cribbs, D. L., Pultz, M., Johnson, D., Mazzulla, M. & Kaufman, T. C. (1992) *EMBO J.* **11**, 1437–1449.
14. Regulski, M., McGinnis, N., Chadwick, B. & McGinnis, W. (1987) *EMBO J.* **6**, 767–777.
15. Le Motte, P. K., Kuroiwa, A., Fessler, L. I. & Gehring, W. J. (1989) *EMBO J.* **8**, 219–227.
16. Weinzierl, R., Axton, M., Ghysen, A. & Akam, M. (1987) *Genes Dev.* **1**, 386–397.
17. Kornfeld, K., Saint, R. B., Beachy, P. A., Harte, P. J., Paetti, D. A. & Hogness, D. S. (1989) *Genes Dev.* **3**, 243–258.
18. Karch, F., Bender, W. & Weiffenbach, B. (1990) *Genes Dev.* **4**, 1573–1587.
19. Mlodzik, M., Fjose, A. & Gehring, W. J. (1988) *EMBO J.* **7**, 2569–2578.
20. Dalton, D., Chadwick, R. & McGinnis, W. (1989) *Genes Dev.* **3**, 1940–1956.
21. Mlodzik, M. & Gehring, W. J. (1987) *Cell* **48**, 465–478.
22. McGinnis, W. & Krumlauf, R. (1992) *Cell* **68**, 283–302.
23. Mavilio, F., Simeone, A., Giampaolo, A., Faiclla, A., Zappavigna, V., Acampora, D., Poiana, G., Russo, G., Peschle, C. & Boncinelli, E. (1986) *Nature (London)* **324**, 664–668.
24. Müller, M., Affolter, M., Leupin, W., Otting, G., Wüthrich, K. & Gehring, W. J. (1988) *EMBO J.* **7**, 4299–4304.
25. Güntert, P., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1991) *J. Mol. Biol.* **217**, 531–540.
26. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
27. Studier, F., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1991) *Methods Enzymol.* **185**, 60–89.
28. Anil-Kumar, Ernst, R. R. & Wüthrich, K. (1980) *Biochem. Biophys. Res. Commun.* **95**, 1–6.
29. Griesinger, C., Otting, G., Wüthrich, K. & Ernst, R. R. (1988) *J. Am. Chem. Soc.* **110**, 7870–7872.
30. Brüschweiler, R., Griesinger, C., Sørensen, O. W. & Ernst, R. R. (1988) *J. Magn. Reson.* **78**, 178–185.
31. Billeter, M., Qian, Y. Q., Otting, G., Müller, M., Gehring, W. J. & Wüthrich, K. (1990) *J. Mol. Biol.* **214**, 183–197.
32. Otting, G., Orbons, L. M. & Wüthrich, K. (1990) *J. Magn. Reson.* **89**, 423–430.
33. Bodenhausen, G. & Ruben, D. (1980) *Chem. Phys. Lett.* **69**, 185–189.
34. Otting, G. & Wüthrich, K. (1988) *J. Magn. Reson.* **76**, 569–574.
35. Arseniev, A. S., Sobol, A. G. & Bystrov, V. F. (1986) *J. Magn. Reson.* **70**, 427–435.
36. Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **63**, 207–213.
37. Hoult, D. I. & Richards, R. E. (1975) *Proc. R. Soc. London A* **344**, 311–320.
38. Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **113**, 967–974.
39. Mao, J., Mareci, T. H., Scott, K. N. & Andrew, E. R. (1986) *J. Magn. Reson.* **70**, 310–318.
40. Eccles, C., Güntert, P., Billeter, M. & Wüthrich, K. (1991) *J. Biomol. NMR* **1**, 111–130.
41. Billeter, M., Braun, W. & Wüthrich, K. (1982) *J. Mol. Biol.* **155**, 321–346.
42. Wagner, G. & Wüthrich, K. (1982) *J. Mol. Biol.* **155**, 347–366.
43. Wider, G., Lee, K. H. & Wüthrich, K. (1982) *J. Mol. Biol.* **155**, 367–388.
44. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids* (Wiley, New York).
45. Affolter, M., Percival-Smith, A., Müller, M., Leupin, W. & Gehring, W. J. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4093–4097.
46. Sauer, R. T., Smith, D. L. & Johnson, A. D. (1988) *Genes Dev.* **2**, 807–816.
47. Smith, D. L. & Johnson, A. D. (1992) *Cell* **68**, 133–142.
48. Mitchell, P. & Tjian, R. (1989) *Science* **245**, 371–378.
49. Schughart, K., Utset, M. F., Awgulewitsch, A. & Ruddle, F. H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5582–5586.
50. Gibson, G., Schier, A., LeMotte, P. & Gehring, W. J. (1990) *Cell* **62**, 1087–1103.
51. Mann, R. S. & Hogness, D. S. (1990) *Cell* **60**, 597–610.
52. Kuziora, M. A. & McGinnis, W. (1989) *Cell* **59**, 563–571.
53. Percival-Smith, A., Müller, M., Affolter, M. & Gehring, W. J. (1990) *EMBO J.* **9**, 3967–3974.