Different DNA contact schemes are used by two winged helix proteins to recognize a DNA binding sequence

Robert Bravieri, Tatiyana Shiyanova, Tsu-Hua Chen, Dave Overdier and Xiubei Liao*

Department of Biochemistry, College of Medicine, University of Illinois, 1819 West Polk Street, Chicago, IL 60612, USA

Received February 26, 1997; Revised and Accepted May 30, 1997

ABSTRACT

The hepatocyte nuclear factor 3 (HNF-3)/fork head (fkh) family contains a large number of transcription factors which recognize divergent DNA sequences via a winged-helix binding motif. In this report we present studies on the DNA binding properties of winged-helix HNF-3/fkh homologues 1 (HFH-1) and 2 (HFH-2) which recognize a shared DNA binding site with different affinities. To explore how HFH-1 and HFH-2 proteins recognize this DNA binding sequence, the binding affinities of these two HFH proteins toward a series of DNA sites containing a single strand trimer loop insertion at different positions were studied. This insertion induces a bend of ∼80° in the DNA binding site (prebending). HFH-1 and HFH-2 were shown to recognize DNA sites prebent at many nucleotide positions on both strands of the DNA sequence. Both HFH-1 and HFH-2 were more sensitive to mismatch insertions on the sense strand of the DNA binding site, especially within the AAAATAAC sequence. Our data suggest that the recognition helix (helix 3) recognizes the AAAATAAC sequence and that the helix 3/DNA interaction results in bending of the DNA which narrows the major groove in the AAAATAAC sequence. Furthermore, the binding affinities of HFH-1 and HFH-2 toward DNA binding sites with base-pair reversion in the AAAATAAC sequence was also investigated. Different patterns of response from HFH-1 and HFH-2 to both prebent and base-pair reverted binding sites were observed. Our results demonstrate that even two highly conserved members of the winged-helix family may contact the same DNA sequence differently.

INTRODUCTION

Transcription plays a central role in cell growth, development and differentiation by regulating gene expression at well timed growth stages in specific tissues (reviewed in 1–3). Regulation of gene expression is controlled primarily at the level of transcription initiation which is coordinated by an array of transcription factors. Although several control mechanisms operate at the level of transcriptional initiation, one of the most critical steps is the recognition of a specific DNA sequence by transcription factors leading to enhanced and/or repressed transcription of target genes. This recognition process requires non-specific and base-specific interactions between amino acid residues of a transcription factor and nucleotide residues of its cognate DNA site. Therefore, an understanding of the basic interactions between transcription factors and their target sequences is crucial for the understanding of transcriptional regulation. It has been shown that many transcription factors can be grouped into families based on their conserved DNA recognition motifs (reviewed in 4,5), such as zinc finger (6), homeodomain (7,8), basic leucine zipper (9) and basic helix–loop–helix motif (5,10). The DNA binding properties and the structural basis of these motifs in recognizing specific DNA sites are well studied, and many different DNA contact schemes are observed. These differences may play roles in gene specific DNA recognition at various stages of cellular differentiation and organogenesis.

The winged helix–turn–helix motif (11) was initially identified in the mammalian hepatocyte-enriched transcription factor HNF-3 (12) and the Drosophila fork head homeotic protein (fkh; 13). However, since then many transcription factors containing the winged-helix motif have been cloned from organisms ranging from yeast to human (14–24). Numerous studies indicate that the HNF-3/fkh family members play important roles in tissue-specific gene and developmental regulation. The transcription factors HNF-3α, -3β and -3γ were shown to regulate the restricted expression of several genes important for liver function (25–27,28). HNF-3β was further shown to be essential for node and notochord formation in the mouse embryo (29,30).

All members of the HNF-3/fkh family contain amino acid homology in the winged-helix DNA binding domain consisting of ∼100 amino acids, which bind to DNA as monomers. The crystal structure of the HNF-3γ–DNA complex showed that the DNA binding domain adopts a modified helix–turn–helix motif on a DNA binding site (11). In this motif, the N-terminal region of the HNF-3γ folds into three α-helices resembling the helix–turn–helix motif, and the third helix (helix 3) is presented into the major groove of DNA and provides the critical base-specific interactions. The C-terminus of HNF-3γ contains
two wings in a coiled structure which contribute one additional minor groove base-specific contact as well as several backbone contacts. However, the site used in the cocrystal with HNF-3γ does not resemble the PCR selected strong binding sites and shows a reduced affinity with the HNF-3 protein in gel mobility shift assays (31). Therefore, the mechanism by which winged-helix proteins recognize the selected strong binding sites cannot be deduced clearly.

The amino acid sequence of the recognition helix is highly conserved among the winged-helix family members, while each winged-helix protein possesses distinct binding specificity. In some cases, the same DNA site binds two family members. Whether the DNA binding activities of two members can be controlled individually by DNA modification is not clear. Therefore, understanding whether the same DNA contact scheme is used for two highly homologous family members to bind one DNA sequence would provide important information on this question.

In order to understand these unique properties of the winged-helix proteins, we studied the DNA binding properties of two winged-helix family members: HFH-1 and HFH-2 (18) which is also known as GENESIS (24). Both proteins possess >50% amino acid sequence identity in the winged-helix domain with HNF-3γ. Although the base-specific DNA contact residues identified in the cocrystal of the HNF-3γ-DNA complex are conserved in HFH-2 and HFH-1, the three proteins bind to different sets of DNA sequences. Starting from a PCR-selected 26 bp DNA binding site (31), which is recognized by both HFH-2 and HFH-1, we further determined that a 13 bp sequence of DNA is the core sequence for HFH-1 and HFH-2 binding. Additionally, we also designed a series of oligonucleotides, where each of them carries a single strand trimer loop insertion which induces a bend of ~80° in the DNA binding site (32,33). The interactions between HFH-2 or HFH-1 with the prebent DNA sites and the sites with a base substitution were studied by gel mobility shift assays. These studies suggest that the two winged-helix proteins bind to the same sequences with slightly different DNA contact schemes.

MATERIALS AND METHODS

Expression and purification of the DNA binding domain of HFH-2 and HFH-1

The gene encoding the DNA binding domain of HFH-2 (31) was generated by PCR amplification from the rat HFH-2 genomic clone (18). The endonuclease recognition sites NdeI and EcoRI were engineered for the cloning of the DNA binding domain of HFH-2 in a correct reading frame. The PCR product was cloned into pET21b (Novagen) vector using NdeI and EcoRI restriction endonuclease sites, and the coding sequence was confirmed by DNA sequencing (34). The expressed protein (101 amino acids, 12 kDa) contains the functional DNA binding domain of HFH-2 and an extra Met at the N-terminus.

Production of the HFH-2 DNA binding domain was achieved by induction of T7 polymerase in Escherichia coli strain HMS-174 (35) by adding 1 mM IPTG to the media at an OD600 of ~0.6. The cells were then grown overnight at 37°C and collected by centrifugation. The majority of overproduced HFH-2 is located in inclusion bodies, as judged by SDS–PAGE. The cell pellet was resuspended in 3 vol of lysis buffer (50 mM Tris, pH 7.5, 50 mM NaCl and 1 mg/ml lysozyme) and incubated at room temperature for 30 min. After lysozyme treatment, Triton X-100 was added to the cell suspension to a final concentration of 0.5% v/v. Following brief sonication, insoluble inclusion bodies were collected by centrifugation at 28 000 g for 20 min. The pellet was resuspended in solubilization buffer (50 mM Tris–HCl, pH 7.5, 50 mM NaCl, 6 M Guanidine–HCl, 5 mM DTT) and solubilized by brief sonication, followed by vigorous stirring for 30 min. The denatured HFH-2 in 6 M guanidine was passed through a desalting column (Bio-Rad Laboratories, CA) to reduce the guanidine concentration. Desalted HFH-2 was then loaded onto a heparin sepharose column (Pharmacia Biotech Inc. Piscataway, NJ 08854). After removing the residual guanidine, renatured HFH-2 was eluted from the resin by a salt gradient of 0.2–0.8 M NaCl. Fractions from the elution were analyzed by silver-stained SDS–PAGE. The fractions containing the DNA binding domain of HFH-2 were pooled and concentrated by ultrafiltration centricon (Amicon). The pooled fractions ran as a major band on SDS–PAGE with purity >90% as judged by silver-stained SDS–PAGE. The renatured HFH-2 is folded in the winged-helix motif, as determined by NMR NOE constraints, and demonstrates the expected DNA binding affinity toward several PCR selected HFH-2 DNA binding sites (31).

The DNA binding domain of HFH-1 was cloned into the pET21b vector and fused to 6xHis at C-terminus (18). The plasmid was expressed in E.coli strain BL21 (DE3), and protein extracts were harvested according to the manufacturer’s recommendations. The HFH-1 protein was purified from inclusion bodies by using a standard protocol from Qiagen (CA). The expressed HFH-1 protein contains 106 amino acids (13 kDa) including an extra Met at N-terminus and 6xHis at C-terminus and is folded uniformly.

<table>
<thead>
<tr>
<th>NAME</th>
<th>SEQUENCE</th>
<th>BINDING ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>HFH2</td>
<td>GAFYTTTTTATTGTACGCCGCT-5'</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>A</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>A17</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>A19</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>A20</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>A26</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B13</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B15</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B17</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B18</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
<tr>
<td>B26</td>
<td>5'----GAFYTTTTTATTGTACGCCGCT</td>
<td>3'----GAFYTTTTTATTGTACGCCGCT</td>
</tr>
</tbody>
</table>

Double strand DNA products are shown after the primer extension and mung bean nuclease treatment. Sense strands only are shown for A17–A26 and B13–B26. B13 is the product of the primer B annealing to the antisense strand of HFH-2#12. TTP and dATP were added to extend A17 and B15, respectively; dTTP and dATP were added to generate both A19 and B17; dTTP, dATP and dGTP were added in A20; while dTTP, dATP and dCTP were added in B18; all four dNTPs were used in A26 and B26. The right column shows the estimated ratios of complexes to free DNA: –, no binding; +, low complex to free DNA ratio; ++, medium ratio; ++++, high ratio.
judged by NMR methods and demonstrates the expected DNA binding property toward PCR selected HFH-1 DNA binding sites (31).

**Primer extension experiments**

The DNA binding site HFH-2#12 was originally selected by a PCR mediated DNA site selection with recombinant HFH-2 protein (31). A primer extension experiment was used to further define the minimum DNA binding site from HFH-2#12. Two 13 base primers were synthesized for the HFH-2#12 DNA site (Table 1A and B). Four DNA elongation reactions were set up with one strand of HFH-2#12 and the corresponding primer for T7 DNA polymerase (US Biochemical) at room temperature for 20 min. In each reaction, either one, two, three or all four types of nucleotides were added to a final concentration of 20 nM each depending on the sequences to be extended. After the DNA synthesis, T7 DNA polymerase was deactivated at 70°C for 10 min. Unreplicated single strand DNA was removed by adding 150 U of mung bean nuclease (Pharmacia) to the reaction at pH 4.5 at room temperature for 60 min and the mung bean nuclease reaction was stopped by adding EDTA to a final concentration of 20 mM. Depending on the nucleotides provided in the primer extension reactions, this method generates different lengths of DNA binding sites (Table 1). The activities of these sites toward HFH-1 and HFH-2 were evaluated via gel shift assays.

**Gel shift assay**

Gel shift assays (36) were performed as described previously (31). The binding reaction was performed in 20 µl of 1× gel shift buffer (20 mM HEPES pH 7.5, 40 mM KCl, 2 mM MgCl₂, 1 mM DTT, 4% Ficoll) containing the proper amount of the DNA binding domain of HFH-2 or HFH-1, 1 ng of labeled DNA, 100 ng poly(dI–dC) and 200 ng bovine serum albumin incubated at room temperature for 20 min before loading onto the gel. A 9% nondenaturing gel was run at 4°C to separate the complexes from free DNA.

**RESULTS**

**Determination of a common DNA binding sequence for HFH-1 and HFH-2**

Previous studies defined the DNA binding consensus sequences for both HFH-2 and HFH-1 proteins (31). A 26mer DNA binding site (HFH-2#12) can be recognized by both HFH-2 and HFH-1.

---

**Figure 1.** (A–D) Binding activities of partial HFH2#12 sites toward HFH-2 and HFH-1. The partial HFH-2#12 sites were generated by primer extensions in the presence of limiting deoxynucleotides followed by mung bean nuclease treatment to remove unreplicated single stranded DNA. In each reaction, 1 ng of the template DNA and 1 ng of the corresponding primer were used for the limited DNA synthesis. The products of the primer extension reactions are summarized in Table 1. The full length HFH-2#12 was used as a control of the reaction condition. Fifty ng (0.2 µM) of the DNA binding domain of HFH-2 (A and B) or HFH-1 (C and D) was added in each reaction (20 µl) of the gel shift assay. Gel shift assays of HFH-2 (E) and HFH-1 (F) to synthetic partial DNA binding sites. Each binding reaction contains 1 ng of the labeled DNA binding site and a specific amount of the HHF protein. The average molecular weights and concentrations for these partial sites are from 8.4 kDa (B13) to 16.8 kDa (A26 and B26).
In order to define the minimum common DNA sequence for the two proteins, different lengths of partial HFH2#12 sites were obtained by primer extensions from A or B (Table 1) in the presence of limiting deoxynucleotides followed by digestion of single stranded DNA with mung bean nuclease. These sites were used to examine the DNA binding affinities of HFH-2 and HFH-1 proteins via gel shift assays (Fig. 1). As shown in Table 1 and Figure 1A and B, the HFH-2 DNA binding domain demonstrated high affinity binding with the products of the primer extension A20, A26, B17, B18 and B26 (number indicates length of double stranded DNA); while the HFH-2 DNA binding domain showed a reduced affinity to A19 and B15 sites, and did not form a detectable complex with the A17 site. The HFH-1 DNA binding domain bound to A20, A26, B15, B17, B18 and B26 sites with high affinity, and to the A19 site with a much reduced affinity. Also, HFH-1 did not form a detectable complex with either A17 or B13 sites. In our primer extension experiment, the amount of double stranded DNA products could not be controlled accurately due to either incomplete or over digestion of mung bean nuclease. Therefore, in order to further evaluate the binding affinity of these sites toward HFH-1 and HFH-2, a titration experiment was performed on four synthetic DNA binding sites (A19, A17, B15 and B13) which show low affinities for HFH-1 and HFH-2 (Fig. 1F and G). Based on these gel-shift data, we deduced that the overlapped DNA sequence required for HFH-1 and HFH-2 binding is a 15mer between B17 and A20 (C1T2T3A4A5A6T7A8A9A10A11A12A13). Since the extension reactions start DNA synthesis at A12 on the sense strand, this method cannot generate the DNA binding site terminating at T14 when both dTTP and dATP are added. Accordingly, two 13mers, one 14mer and one 15mer as well as several 11 and 12mers were synthesized and tested for HFH-2 and HFH-1 binding activities by gel shift assays. Our data show that the 13 bp sequence (C1T2T3A4A5A6T7A8A9A10A11A12A13) is required for high affinity HFH-2 binding (Fig. 2A) and that the DNA sequences shorter than 13 bp do not bind to either HFH-2 or HFH-1 (data not shown). Interestingly, this 13 bp site does not support high affinity HFH-1 binding in our gel shift assays (Fig. 2B), and the 15 bp sequence (C1T2T3A4A5A6A7A8A9A10A11A12A13T14A15) is needed for HFH-1 recognition (Fig. 2B). This result indicates that HFH-1 recognizes a slightly longer sequence in the HFH-2#12 site.

Analysis of the binding affinities of HFH-2 and HFH-1 to prevent DNA sites

Although HFH-1 and HFH-2 are highly homologous to HNF-3γ, the high affinity DNA sequence for HFH-1 and HFH-2 is different from the DNA site (G1A2C3T4A5A6G7T8C9A10A11C12C13) used in the HNF-3γ cocrystal. Therefore, the base residues which are contacted by DNA recognition helix (helix 3) of HFH-1 or HFH-2 cannot be deduced from the structure of the HNF-3γ-DNA complex. Furthermore, HFH-1 and HFH-2 proteins bind to the HFH-2#12 site with different affinities, and so whether different contact schemes are used in HFH-1 and HFH-2 to recognize this DNA binding site is not clear. To explore these two issues, a family of DNA binding sites with an insertion mismatch loop were synthesized (Table 2). Each site contains a tri-nucleotide single strand insertion on either the sense or the antisense strand before it binds to a protein (32,33), and distorts local DNA conformation in a protein–DNA complex. As a result, the loop insertion disrupts expected critical phosphate and/or base specific interactions at the inserted positions, thus leading to lower or even a total loss of binding affinity. Furthermore, we also expect that a mismatch insertion introduces an ∼80–90° bend into the DNA site. The base pairs which are indicated by (i).
HFH-1 and HFH-2 were evaluated by gel shift assays (Fig. 4A–F and Table 2). Although both HFH-2 and HFH-1 bind to the HFH-2#12 site, the two proteins demonstrate different binding affinity patterns toward the prebent sites. The insertions between A4 and A5 (A4^A5), A5 and A6 (A5^A6), and A6 and A7 (A6^A7) of the sense strand inhibit its binding activities toward both HFH-1 and HFH-2. The insertion between T10 and G11 (T10^G11) of the antisense strand completely abolishes binding activity toward HFH-2, but not to HFH-1 (Fig. 4A–F). Contrary to this, insertions between T3 and A4 (T3^A4) and C11 and A12 (C11^A12) reduce its binding activity more toward HFH-1 than HFH-2 (Fig. 4A–F).

In general, single strand loop insertions in the double stranded AAAATAAC sequence, especially in the sense strand of AAAATAAC, have a more severe effect on the affinity of the binding site than insertions at other positions. The insertion in the double stranded AAAATAAC sequence that has the minimum effect on HFH-1 and HFH-2 binding affinity is between A7 and T8 (A7^T8) on the sense strand. If a mismatch insertion loop is placed within the AAAATAAC sequence, HFH-2 protein–DNA complexes migrate at almost the same speed through the gel as the wild-type HFH-2–DNA complexes (Fig. 4A and B), even though the protein free DNA sites with an insertion loop migrate slower than the protein free wild-type DNA site. However, if the insertions are placed outside of the double stranded AAAATAAC sequences, six out of eight HFH-2–DNA complexes show slower mobility than the wild-type HFH-2–DNA complex (Fig. 4A and B). This pattern of gel shift mobility indicates that the DNA binding sites with a mismatch insertion in both strands of the AAAATAAC sequence adopt approximately the same conformation in the HFH-2 complex as that of the wild-type DNA sequence, while the DNA sites with an insertion outside the AAAATAAC sequence adopt different conformations as that of the wild-type DNA site in the complex. In contrast, all the HFH-1–DNA complexes in the gel shift assay show almost the same migration distance (Fig. 4C and D). This indicates that, unlike in HFH-2 complexes, prebent DNA sequences in HFH-1 DNA complexes all adopt a similar conformation. This is another one of the binding mode differences between HFH-1 and HFH-2. It seems that the structure of HFH-2 is more plastic than HFH-1 and can fit many higher order DNA conformations in protein–DNA complexes. This observation would also seem to support the previous result that HFH-2 has a broader DNA sequence specificity than many other HFH family members (31).

Analysis of the DNA binding activities of HFH-1 and HFH-2 to sites with a base-pair reversion in the core sequence

By introducing the tri-nucleotide mismatch loop into the HFH-2 #12 binding site we identified several positions which are comparably more sensitive to HFH-2 and/or HFH-1 binding than other positions. The tri-nucleotide insertion distorts the DNA structure, especially 2 bp immediately flanking an insertion loop. Thus, an insertion loop may lead to the disruption of both non-specific and base-specific protein–DNA interactions at nucleotides flanking the insert. In order to determine possible base-specific interactions, six DNA binding sites were synthesized. Each of these DNA sites contains a base-pair reversion immediately adjacent to a position (A4^A5, A6^A7 and T10^G11) at which the tri-nucleotide insertion severely disrupts the interaction between the proteins and DNA. The binding affinity of HFH-2 and HFH-1 to these sites was studied by gel shift assays (Fig. 5A and D). Our data show that five out of six base-pair reversions (A4T, A6T, A7T, A10T and C11G) reduce the binding activity for both HFH-1 and HFH-2. Interestingly, although the insertion between T10 and G11 on the antisense strand of the binding site abolishes its binding activity toward HFH-2 but not HFH-1, the base pair reversion C11G abolished its binding activity toward HFH-1 but not HFH-2 (Fig. 5B and D). This observation further supports our hypothesis that HFH-1 and HFH-2 have different DNA contact schemes for the site HFH-2#12.

DISCUSSION

The crystal structure of the HNF-3γ–DNA complex (Fig. 6) shows that the DNA binding domain of HNF-3γ, when binding
to DNA, is folded into a winged helix–turn–helix motif in which
the third helix (helix 3) in the N-terminal part of HNF-3γ provides
critical base-specific interactions to the DNA binding site, while
the second wing (wing 2) at the C-terminus of HNF-3γ makes
many backbone contacts (11). The winged-helix family proteins
are highly conserved and the base specific residues observed in
HNF-3γ–DNA complex are almost invariable in many other
family members, such as HFH-1 and HFH-2. However, both
HFH-1 and HFH-2 show different DNA binding specificity to
HNF-3γ. The site HFH-2#12 which is recognized by both HFH-1
and HFH-2 is rather different from the site used in the cocrysal
of HNF-3γ. Thus, the DNA contact schemes used by HFH-1 and
HFH-2 cannot be deduced from the crystal structure.

Our data show that the DNA binding activities of HFH-2 and
HFH-1 are more severely inhibited by the loop insertions into the
double stranded AAAATAAC sequence, especially on the sense
strand, than those in flanking sequences. Therefore, this result is
consistent with the observation in the crystal structure of the
HNF-3γ–DNA complex that DNA contacts made by helix 3 and
helix 1 of HNF-3γ are primarily in the middle of one strand of the
DNA site (11). It seems that the double stranded AAAATAAC
core sequence is contacted by helix 3 of HFH-1 and HFH-2. It is
likely therefore that the insertion-sensitive positions are the result
of disruption of critical interactions between the DNA site and the
protein and/or a steric effect which blocks helix 3 insertion into
the major groove of the DNA site. Furthermore, HFH-2–DNA
complexes with a loop insertion within both strands of the
AAAATAAC sequence show the same migration rate as the
wild-type HFH-2–DNA complex, while several HFH-2–DNA
complexes with a loop insertion outside the AAAATAAC core
sequence show slower migration rates than the wild-type
HFH-2–DNA complex. This result indicates that prebent
HFH-2#12 sites with a loop insertion in either strand of the
AAAATAAC sequence adopt approximately the same conforma-
tion in HFH-2 complexes as that of the wild-type DNA site, while
HFH-2#12 sites prebent outside the AAAATAAC sequence can
adopt alternative conformations. These data further support that
helix 3 directly contacts the double stranded AAAATAAC core
sequence. Since an α-helix is a relatively well defined structure,
the DNA recognition surface provided by an α-helix which is
inserted into the major groove of DNA likely contacts DNA in
one conformation for maximum affinity. Loop insertions in this
region therefore severely disrupt these DNA contacts resulting in
reduced affinity. On the contrary, wing 2 and/or wing 1 of HFH-2
and HFH-1 are highly disordered in solution (Marsden and Liao,
unpublished), and these wings may have the ability to adopt
alternative conformations in the protein–DNA complex. Therefore,

Figure 4. Binding of HFH-2 and HFH-1 to the prebent DNA sites. One ng DNA
(17.6 kDa on average) was used in each reaction of 20 µl of the final volume.
In each gel shift assay shown in (A–D), 100 ng of HFH-2 or HFH-1 was used
to form the complex. In each gel shift assay shown in (E) and (F), the amount
of the HFH-1 or HFH-2 is indicated as in the figure. (A) Gel shift assay of
HFH-2 with the DNA sites containing the insertion loops in the sense strand of
the HFH-2#12 sequence. (B) Gel shift assay of HFH-2 with the DNA sites
containing the insertion loops in the antisense strand of the HFH-2#12
sequence. (C) Gel shift assay of HFH-1 with the DNA sites containing the
insertion loops in the sense strand of the HFH-2#12 sequence. (D) Gel shift
assay of HFH-1 with the DNA sites containing the insertion loops in the
antisense strand of the HFH-2#12 sequence. Titration study of binding activities
of T3^A4, A7^T8, T10^G11 and C11^A12 sites to HFH-2 (E) and HFH-1 (F).
Figure 5. Gel shift assays of binding affinities of HFH-2 (A) and HFH-1 (B) to DNA binding sites of 20 bp with a base substitution in AAAA TAAC sequence (12.8 kDa on average). The final volume for each reaction is 20 µl. Titration studies of HFH-2 (C) and HFH-1 (D) to A10T and C11G. In each reaction 1 ng of the labeled DNA binding site (10.3 kDa) and a specified amount of the HFH protein were used to form protein–DNA complex.

Figure 6. The crystal structure of HNF-3γ-DNA complex reported by Clark et al. (11). HNF-3γ adopts winged-helix motif on the DNA binding site. The amino (N) and carboxyl (C) termini, α helices (H1, H2, H3), sheets (S1, S2, S3), and wings (W1 and W2) are labeled. The recognition helix (H3) provides base-specific contacts to the DNA site in the major groove. W2 contributes one additional base-specific interaction in minor groove and many non-specific interactions in the complex.

many non-specific DNA contacts made by the two wings show less sensitivity to a loop insertion.

By using a circular permutation assay, Pierrou et al. (23) demonstrated that the binding of winged-helix proteins to their cognate DNA sites causes an ∼80–90° bending in the DNA. However, the permutation method cannot indicate whether the bending is sharp or gradual over an entire site. The HNF-3γ–DNA cocystal structure indicates a 13° bend which narrows the major groove to fit helix 3. On the basis of our study, we rationalize that if a loop insertion causes the DNA site to bend at the same position and in the same direction as the wild-type DNA binding site does in a HFH–DNA complex, then the activity of this present site will not be severely reduced. In our study, results show that the A7^T8 site which contains a loop insertion between A7 and T8 in the sense strand of the core sequence has the highest binding affinity for HFH-2 and HFH-1 (Fig. 4). This indicates a possible sharp bending of the DNA site between A7 and T8 of the sense strand which is the center of the double stranded AAAATAAC sequence. The 8 bp AAAATAAC sequence is 4/5 of a turn in B type DNA with the 5'-AAAA-3' sequence on the sense strand and 3'-ATGTG' sequence on the antisense strand defining roughly a major groove on one side of DNA. Thus, the AT in the middle of this sequence is on the opposite side of the DNA double helix, and the loop insertion between the AT bases narrows the major groove in AAAATAAC which hosts the recognition helix of HFH-2 and HFH-1.

In the structure of the HNF-3γ-DNA complex, the HNF-3γ DNA binding domain recognizes T_7XXG_T_8/X_10 in a base-specific fashion, with no base-specific interactions being observed for the bases marked as X. The contacts to G7, T8 and A10 are made by helix 3, while the contact to T4 is made by the flexible wing 2. Since helix 3 of HFH-1 or HFH-2 makes contacts with the AAAATAAC sequence, the base-specific contacts made by HFH-1 or HFH-2 to the AAAATAAC sequence are expected to be rather different from HNF-3γ. For example, base pair inversions at underlined bases in the A_7AAA_T8AC sequence reduces its binding activities toward HFH-2 and HFH-1. This result cannot be predicted on the basis of the crystal structure of the HNF-3γ–DNA complex. However, three HNF-3γ, HFH-1 and HFH-2 also show a similarity in DNA binding. Our results show that both HFH-1 and HFH-2 are more sensitive to the loop insertions into the sense strand of the HFH-2#12 site than those into the antisense strand. These results indicate HFH-1 and HFH-2 make more DNA contacts to the sense strand of this site which is consistent with the crystal structure of the HNF-3γ–DNA complex in which the most of the DNA contacts are located on one strand of the DNA site.

Even though HFH-1 and HFH-2 both recognize the HFH-2#12 site, they show different responses to the loop insertions and base pair inversions on this site. The insertion between T10 and G11
of the antisense strand abolishes its binding activity for HFH-2 completely, while only reducing the binding activity toward HFH-1. A base pair reversion C11G abolishes binding activity for HFH-1, but not HFH-2; and the loop insertion between T3 and A4 and between C11 and A12 of the sense strand reduces the binding activity more to HFH-1 than to HFH-2. Furthermore, HFH-2 can recognize a 13 bp sequence from the HFH-2#12 site, while HFH-1 needs 15 bp to support binding. These results indicate that the observed different binding affinities to the site HFH-2#12 are likely the result of slightly different DNA contact schemes used by HFH-1 and HFH-2 to this site. Since the DNA contact residues observed in HNF-3γ are conserved in HFH-1 and HFH-2, the differences in the binding affinities of the two proteins to loop insertions and base-pair inversions may be explained by a model proposed by R. Costa and colleagues (31) in which a 20 amino acid sequence outside the recognition helix (helix 3) is proposed to regulate the recognition schemes in different winged-helix family members. This 20 amino acid sequence is the most divergent sequence in otherwise conserved winged-helix proteins and is different in HFH-1 and HFH-2. Interestingly, our NMR studies show that HFH-2 contains a short helix in the middle of the 20 amino acid sequence between helix 2 and helix 3 and this short helix is conserved in HFH-2 complexed with the HFH-2#12 DNA site (Jin, Marsden and Liao, unpublished). Due to the insertion of this helix (helix 4), helix 3 in HFH-2 is slightly shorter and is moved toward the C-terminus of the protein and the overall folding of HFH-2 in the DNA complex is expected to be different from HNF-3γ in the DNA complex. These folding differences may change the position of helix 3, as well as the relative sites of DNA contact residues located on different secondary elements and lead to different DNA contact schemes in HFH family members. Since helix 3 is highly conserved and recognizes the AAAATAAC sequence in both HFH-1 and HFH-2, the DNA contacts made by helix 3 of the two proteins are likely similar, but not identical. The position of helix 3 relative to other secondary elements which make DNA contacts, such as wing 1, wing 2 and helix 1 will fine-tune the final DNA contact schemes used by the two HFH proteins. These DNA contact differences are the result of folding differences regulated by the sequence between helix 2 and helix 3. Since the core sequence AAAATAAC in our studies is different from the sequence contacted by helix 3 of HNF-3γ, it is possible that the DNA contact scheme used by helix 3 of HNF-3γ is rather different from the scheme used by helix 3 of HFH-2 or HFH-1. This hypothesis needs further structural studies of the HFH family members binding to different DNA binding sequences.

The fundamental question then is, how do HFH-1 and HFH-2 proteins recognize DNA sites which are prebent into many different conformations? Previous data have demonstrated that DNA with single strand loop insertions show a population in rapid transition between states ranging from straight to highly bent (33). The fraction of the straight DNA may be the target for the HFH protein binding in our experiments. However, since the two wings are highly dynamic in solution, we cannot rule out the possibility that the prebent DNA can be directly recognized by the HFH proteins. If this is the case, the indication is that a less structured DNA binding domain may have an advantage since it can interact with DNA in different conformations.

Previous results show that many tissues contain more than one expressed winged-helix family member (31,37), and that although each member has its unique DNA binding specificity, some sequences can be recognized by more than one HFH protein. Thus, balancing the activities of two HFH family members expressed in the same tissue is important for transcription regulation. Our studies indicate that although the DNA binding site HFH-2#12 is recognized by both conserved HFH-2 and HFH-1, the site is contacted by the two proteins with slightly different binding schemes. These differences in protein–DNA recognition schemes might be used by cells to control the activity of one HFH family member without interfering with the activity of the other family members. Furthermore, we show that HFH-2 can tolerate different higher orders of DNA conformations outside the core binding sequence in the HFH-2–DNA complexes, while the HFH-1 protein cannot. Thus, it is possible that the higher order of conformations in DNA binding sites play a direct role in transcription factor recognition. Many studies have demonstrated that bent DNA sequences can be a result of protein binding (38–41). Since transcription initiation usually requires the participation of an array of transcription factors, the ability of HFH-2 to recognize and to tolerate a higher order DNA structure, possibly caused by another transcription factor, may be required for transcription regulation.

The method outlined in this paper can be used to study protein–DNA interactions of other DNA binding proteins. It can be used to determine the DNA recognition sequence for each domain of a multi-domain DNA binding protein, such as zinc-finger proteins, and to study whether a strand of DNA is preferably contacted by a DNA binding protein. This method provides easily interpretable data to study protein–DNA interactions.

ACKNOWLEDGEMENTS

We thank Dr Robert Costa for the gift of HFH genes, and for fruitful discussion and communication of results prior to publication. We also thank Emily Myrick and Ian Marsden for reviewing the manuscript. This work was funded by grant GM 52034 from the National Institutes of Health.

REFERENCES