DNA BINDING AND GENE ACTIVATION PROPERTIES OF THE Nmp4
NUCLEAR MATRIX TRANSCRIPTION FACTORS

Kitti Torrungruang*1, Marta Alvarez*2, Rita Shah1, Jude E. Onyia3, Simon J. Rhodes4,
Joseph P. Bidwell1, 2§

1. Department of Periodontics, Indiana University School of Dentistry, Indianapolis, IN 46202
2. Department of Anatomy and Cell Biology, Indiana University School of Medicine, Indianapolis, IN 46202
3. Gene Regulation, Bone, and Inflammation Research Division, Lilly Research Labs, Indianapolis, IN 46285
4. Department of Biology, Indiana University-Purdue University Indianapolis, IN 46202

Running title: Nuclear matrix transcription factors

*Kitti Torrungruang and Marta Alvarez contributed equally to this study.

§Address correspondence to:
Joseph P. Bidwell
Department of Anatomy and Cell Biology
Medical Science Building 5035
Indiana University School of Medicine
635 Barnhill Drive
Indianapolis, IN 46202
E-mail: jbidwell@iupui.edu

Copyright 2002 by The American Society for Biochemistry and Molecular Biology, Inc.
SUMMARY

Splice variants of the Nmp4 gene include nuclear matrix transcription factors that regulate the type I collagen α1(I) polypeptide chain (COL1A1) promoter and several matrix metalloproteinase (MMP) genes. To date, these are the only Cys2His2 zinc finger proteins known to bind within the minor groove of homopolymeric (dA:dT) DNA. Nmp4 isoforms contain from five to eight Cys2His2 zinc fingers, an SH3-binding domain that overlaps with a putative AT-hook, and a poly-glutamine/alanine repeat [poly (QA)]. To determine the mechanistic significance of Cys2His2 zinc finger association with this unusual consensus DNA-binding element, we identified the Nmp4 DNA-binding and transcriptional activation domains. Zinc fingers #2, #3, and #6 mediated association with the homopolymeric (dA:dT) COL1A1/MMP DNA consensus element. The amino-terminus of the Nmp4 protein exhibited a strong transactivation capacity when fused to the GAL4 DNA-binding domain (DBD), but this activity was masked within the context of the full-length Nmp4-GAL4 DBD chimera. However, upon binding to the COL1A1/MMP homopolymeric (dA:dT) element, the native Nmp4 protein upregulated transcription and the poly (QA) domain acquired a significant role in transactivation. We propose that allosteric effects, induced upon zinc finger association with the homopolymeric (dA:dT) minor groove, confer context-specific functionality to this unusual family of Cys2His2 transcription factors.
INTRODUCTION:

The Cys$_2$His$_2$ zinc finger proteins comprise the largest group of eukaryotic DNA binding proteins (1, 2). This type of zinc finger contains two cysteine and two histidine residues that bind a single zinc atom stabilizing the finger, a loop consisting of a conserved $\beta\beta\alpha$ structure (2, 3). The amino acids on the surface of the $\alpha$-helix contact bases in the DNA major groove (2). This motif is repeated in tandem and allows for binding to DNA sequences of variable lengths (3, 4). The modular design of the zinc finger domain also affords many combinatorial possibilities for specific DNA recognition underlying the tremendous variety of consensus sequences recognized by zinc finger proteins (3, 4, 5).

Despite the vast DNA recognition potential of Cys$_2$His$_2$ zinc finger proteins, few are known to associate with AT-rich DNA. These DNA sequences can act as significant structural elements in transcription (6, 7, 8). The narrow minor groove of AT-rich DNA furnishes a structural recognition motif for many architectural transcription factors, proteins that bend DNA and influence transcription by altering the interactions between other trans-acting proteins (6). Matrix attachment regions are typically long stretches (200-800 bp) of AT-rich DNA that anchor chromatin loops to the nuclear matrix and provide chromatin boundaries and binding sites for nuclear matrix regulatory proteins (9). Homopolymeric (dA:dT) DNA can have functional significance apart from interaction with trans-acting proteins (7, 8). By comparison to B-form DNA, the homopolymeric (dA:dT) sequences have a shorter helical repeat (10.0 bp/turn as opposed to 10.5 bp/turn), a narrower minor groove (~9 Å as compared with ~15 Å), and additional bifurcated hydrogen bonds that provide added structural stability (7, 8). The rigidity of these elements resists conforming tightly to the face of the nucleosome, creating localized DNA distortions on either end of the element providing "access windows" for trans-acting factors (7, 10, 11, 12).

The Nmp4 proteins (Figure 1A) are the only known Cys$_2$His$_2$ zinc finger proteins which bind to the minor groove of homopolymeric (dA:dT) sites, such as those present in the promoters of many extracellular matrix genes, including the type I collagen $\alpha1(I)$ polypeptide chain (COL1A1) promoter and several matrix metalloproteinases (MMP) (13,
The MMP Nmp4 consensus element is embedded within both COL1A1 sites A and B in the reverse orientation (Figure 1B). Some Nmp4 binding sites in the MMP7 promoter contain a stretch of 7-9 A’s, similar to the stretch of 9 T’s in the COL1A1 consensus sequence.

In a previous study we compared the capacities of the Nmp4 isoforms 11H, 13H, 21H, and 28H, (accession numbers AF216804, AF216805, AF216806, AF216807, AB019281, see Figure 1A) to bind to the COL1A1/MMP regulatory sequence (13). The Nmp4 isoforms 11H, 13H, and 21H are full-length cDNAs encoding in-frame splice variants containing from 6-8 Cys2His2 zinc fingers. Nmp4/11H and Nmp4/13H have eight zinc fingers whereas Nmp4/21H has six fingers, missing #4 and #5. Nmp4/28H has only the first 5 zinc fingers, is lacking the poly(QA) domain and carboxy terminus common to the other isoforms, and has an insert just before the first finger (Figure 1A).

Using electrophoretic mobility shift analysis (EMSA), the Nmp4 proteins 11H, 13H, and 21H exhibited binding to the COL1A1/MMP consensus site and yielded binding profiles similar to that obtained with the nuclear matrix fraction from osteoblast-like cells (13). However, the Nmp4/28H clone did not bind to the consensus element (13). Nmp4/21H, therefore, is a native truncated isoform exhibiting COL1A1/MMP DNA binding.

Nmp4 proteins have an SH3-binding domain (14) that overlaps with a putative AT-hook motif (13), another potential DNA-binding domain (Figure 1A). The AT-hook is characteristic of the HMGA superfamily of architectural transcription factors and can mediate binding to the minor groove of AT-rich DNA (6, 16, 17). However, both the SH3-binding and AT-hook domains can mediate protein-protein interactions in the formation of higher-order protein complexes in the cytoplasm and nucleus, respectively (18, 19, 20). In the amino (N)-terminal region, Nmp4 proteins also possess a serine-threonine-rich motif (amino acids 51-173), similar to transactivation domains in ITF-1, ITF-2, and Pax6 (21, 22). Additionally, the Nmp4 isoforms include a poly (Q) motif encoded by a CAG trinucleotide repeat and a poly (QA) repeat (13, 14). Both of these motifs have been shown to influence the transcriptional activity of other key trans-acting proteins including Cbfa1/Osf2, CA150, and the human androgen receptors (23, 24, 25). Here, our objective was to identify the DNA-binding and transactivation domains of the Nmp4 proteins. We observed that only three specific zinc fingers (#2, #3, and #6 of
isoform 21H) were required for binding to the homopolymeric (dA:dT) \textit{COL1A1/MMP} Nmp4 consensus element. Transactivation experiments revealed a sensitivity of specific transcriptional regulatory domains to their attached DBD or to their DNA-binding state consistent with the apparent multiple functions of the proteins themselves.

\textbf{MATERIALS AND METHODS:}

\textit{Cell culture}

Human embryonic kidney 293T cells (ATCC, Rockville, MD) were maintained in DMEM (Gibco BRL) low glucose media supplemented with 100 IU/ml penicillin, 100\(\mu\)g/ml streptomycin, 25 \(\mu\)g/ml amphotericin, 2 mM L-glutamine (Gibco BRL), and 10\% fetal bovine serum (FBS; Sigma). The rat osteosarcoma cells, UMR 106-01, a generous gift from Dr. Nicola Partridge (St. Louis University, St. Louis, MO) were grown as previously described (15). All cells were maintained in humidified 95\% air/5\% CO\(_2\) at 37\(^\circ\)C.

\textit{Preparation of full-length and deletion expression constructs}

The cDNAs of the wild-type Nmp4 isoforms 11H, 13H, 21H, and 28H were subcloned into pcDNA-3\textsuperscript{TM} (Invitrogen, Carlsbad, CA) as previously described (13). The deletions from N-terminal or carboxyl (C)-terminal were generated by cloning the PCR products of truncated 11H or 21H clones into pcDNA-3\textsuperscript{TM} vectors. The internal deletions were prepared using ExSite\textsuperscript{TM} PCR-based site-directed mutagenesis kit (Stratagene, La Jolla, CA). The integrity of all plasmids was confirmed by DNA sequencing and restriction enzyme digestion.

\textit{In vitro transcription/translation of expression constructs, EMSA analysis, and dimethyl sulfate (DMS) interference fingerprinting.}

The Nmp4 pcDNA-3\textsuperscript{TM} constructs were expressed \textit{in vitro} using the TNT® Quick Coupled Transcription/Translation System following the manufacturers’ instructions.
The molecular weights of the protein products were confirmed by western analysis.

Protein-DNA interactions were characterized using electrophoretic mobility shift analysis (EMSA) as previously described (15). The 20 µl binding reactions included 75 mM KCl, 15% glycerol, 0.15mM EDTA, 500 ng of poly (dI:dC), 0.1 mM DTT, 19 mM Hepes (pH 7.5), 0.0075% NP-40, 2-4 µl of TNT lysate protein or 2 µg nuclear matrix protein, and 0.5 nM end-labeled rat COL1A1 5' regulatory fragment (site A -3489/-3434 nt and site B -1594/-1541 [15]) as the probe. These fragments also serve as Nmp4-MMP binding sites since the only difference is the orientation of the homopolymeric sequence within the promoters (13, 14, also see Figure 1B). The rat COL1A1 5' regulatory region (-3518/+115 nt), was a generous gift from A. Lichtler, B. Kream, and D. Rowe (The University of Connecticut Health Center, Framington, CT). We used Western blot analysis to relatively quantify the amount of each protein in the in vitro transcription translation lysate. The volume of the lysate used in the EMSA (2-4 µl) was based on the data from the Western analysis. Distamycin A hydrochloride and methyl green (Sigma Chemical Co. St. Louis, MO) were prepared as aqueous stock solutions (10 mM). These compounds were added to some of the EMSA binding reactions and incubated for 25-30 min at room temperature prior to electrophoresis which was performed on 8% polyacrylamide gels (80:1 acrylamide: N,N'methylbisacrylamide) in 1 x Tris/glycine EDTA buffer (4ºC).

For DMS interference fingerprinting analysis (26, 27), the rat COL1A1 5' regulatory region containing site A (-3518/-3406 nt) was singly end-labeled at either the 5' or 3' end. This DNA was then partially methylated by adding 2 µl of dimethyl sulfate (Sigma) for 1 min in 200 µl reaction buffer (60 mM NaCl, 10 mM Tris-HCl, [pH 8.0], 10 mM MgCl₂, and 1 mM EDTA). Methylation was stopped with the addition of 0.3 M sodium acetate (pH 7.0) and 200 mM 2-mercaptoethanol. The DNA was then precipitated with ethanol, washed, lyophilized, and resuspended in Tris-EDTA buffer. The partially methylated DNA was used as a probe in EMSA with the in vitro translated Nmp4/11H and Nmp4/21H proteins, as described above. The protein-DNA complexes and unbound probe were excised and eluted from the gel. The bound proteins were removed by extraction with phenol-chloroform and the DNA purified using the Qiagen
PCR purification kit (Qiagen, Inc, Valencia, CA) followed by ethanol precipitation. To cleave the modified A and G residues, the DNA pellet was re-suspended in 10 mM sodium phosphate (pH 6.8)/1 mM EDTA and incubated for 15 min. at 92°C, followed by the addition of 100 mM NaOH and incubated for another 30 min at 92°C. The DNA was then ethanol precipitated, re-suspended in Tris-EDTA buffer, equal counts were lyophilized, dissolved in formamide gel-loading dye, and resolved in a 8% sequencing gel. The EMSA and DMS interference gels were dried under vacuum at 80°C for 1 h and then at room temperature for 30 min. Kodak XAR film was exposed to the gel with an intensifying screen overnight at -80°C.

Promoter activation assays

GAL4 fusion constructs of full-length and truncated derivatives of Nmp4/21H were obtained by ligating the PCR-generated coding sequences downstream and in frame with a sequence coding for the GAL4-DNA binding domain (GAL4-DBD, amino acids 1-147) in the pBind vector (Checkmate Mammalian two-hybrid system, Promega). The integrity of all constructs was verified by DNA sequencing and restriction enzyme digestion, and their expression confirmed by western analysis.

For analysis of activation domains, pFRluc vector (Strategene) containing five copies of the GAL4 binding element and a basic promoter element (TATA box) upstream of a luciferase gene was used as a reporter. GAL4 fusion constructs and pFRluc reporter were used in 1:2 µg ratio for transient transfection in 293T cells. The human kidney 293T cells were seeded into 6-well plates at 1.0-1.5 x 10^5 cells/well. After 24 hr, the cells in each well were transiently transfected with DNA using the CalPhos system (Clontech, Palo Alto, CA). The transfected cells were harvested with the Reporter Lysis Buffer (Promega), 48 hr after transfection. Cells were washed two times with phosphate-buffered saline (PBS; 137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4,7H_2O and 1.4 mM KH_2PO_4, [pH 7.1]) and then exposed to lysis buffer for 15 min at room temperature. The lysates were processed for luciferase activity according to the manufacturers’ instructions (Promega) by adding 100 µl of Bright-Glo™ Reagent to 100 µl cell lysate and reading immediately on a Packard Bioscience Fusion Luminometer (Packard Instrument...
Company, Meriden, CT). Total protein concentrations were determined using the Coomassie Plus Protein Assay Reagent (Pierce, Rockford, IL). Results for the luciferase assay were reported as relative light units per µg of total protein.

The human MMP7 promoter, (-2,300/+30 nt, kindly provided by Dr. Lynn Matrisian, Vanderbilt University, Nashville, TN), was subcloned into the pGL-3 Basic (Promega) using BamHI and XhoI. The sequence fidelity of the MMP7/pGL-3 construct was verified by restriction digests. To assay MMP7 activation, the Nmp4/21H or its deletion constructs in the pcDNA-3™ expression vector and the MMP7 reporter construct were used in a 1:2 µg ratio for transient transfection in 293T cells. The luciferase assay was then performed as described above.

\textit{Nuclear matrix protein isolation}

Nuclear matrix proteins were extracted from UMR 106-01 cells using a sequential extraction protocol as previously described (28).

\textit{Statistical analysis}

A one-way analysis of variance (ANOVA) was used to determine the differences among groups with a predetermined p value of < 0.05 to achieve statistical significance. Tukey’s test for post hoc comparison was applied based on the ANOVA results.

\textbf{RESULTS:}

\textit{Zinc fingers #2, #3, and #6 mediate binding to the COL1A1/MMP consensus sequence.}

The truncated Nmp4/28H isoform has an insert before the first zinc finger (Figure 1A). Removal of this insert failed to confer DNA binding to this isoform (data not shown). We therefore used Nmp4/21H to characterize the minimum Nmp4-COL1A1/MMP DNA binding domain because it is the shortest naturally occurring isoform that binds to these regulatory sites (13, 14). Wild-type Nmp4/21H and derivatives were subcloned into pcDNA-3™, expressed \textit{in vitro}, and characterized by EMSA using the COL1A1 site A as the probe (13, 14).
The full-length Nmp4/21H isoform is comprised of 518 amino acids (aa), which include zinc fingers #1-#3 and #6-#8 (construct 1, Figure 2A). Deletion of the C-terminal region [construct 2], the poly [QA] domain [construct 3], or the AT-hook/SH3-binding domain [construct 7], did not abrogate DNA-binding to the \textit{COL1A1/MMP} consensus sequence (Figure 2A/2B). Removal of larger regions of the Nmp4/21H protein including the N-terminus [construct 5] and the C-terminus plus the poly (QA) repeat [construct 4] had no impact on DNA-binding as determined by EMSA (Figure 2A/2B).

Deletion of the Cys$_2$His$_2$ zinc fingers from Nmp4/21H [construct 8] abrogated DNA-binding (Figure 2A/2B). Conversely, this domain alone [construct 9] exhibited binding to the consensus sequence (Figure 2A/2B). To determine whether any specific zinc fingers conferred DNA-binding to the \textit{COL1A1/MMP} binding site we deleted different combinations of three, two, or single zinc fingers. We found that zinc fingers #1, #7, and #8 are not required, while the combination of zinc fingers #2, #3, and #6 were necessary for mediating Nmp4/21H binding to the \textit{COL1A1/MMP} consensus elements. Identical results were obtained using \textit{COL1A1} site B (data not shown). Nmp4/28H is missing zinc finger #6 which would explain its inability to bind to the \textit{COL1A1/MMP} consensus element (13).

To determine whether the Cys$_2$His$_2$ zinc fingers alone mediated binding to the minor groove of the AT-rich Nmp4 consensus sequence, we used EMSA in combination with distamycin and methyl green. Distamycin is used as a competitor for protein binding to the minor groove of AT-rich DNA (29) and methyl green competes for the occupation of the major groove of DNA (30). The binding of the full-length construct of Nmp4/21H to site A of the \textit{COL1A1} promoter was sensitive to the presence of distamycin (Figure 3A) as previously demonstrated (13, 15). Deletion of the AT-hook/SH3-binding domain [construct 7] did not alter the effect of distamycin on Nmp4/21H-\textit{COL1A1} binding (Figure 3A). The zinc finger domain alone [construct 9] exhibited the same sensitivity to distamycin as the full-length protein (Figure 3A). Interestingly, the full-length Nmp4/21H construct exhibited a weak sensitivity to the same concentrations of methyl green (Figure 3B), yet this sensitivity was attenuated upon removal of the putative AT-hook/SH3-binding domain. Finally, DNA-binding by the zinc finger domain alone showed no sensitivity to the presence of methyl green (Figure 3B).
In a previous study, we used methidium-propyl EDTA (MPE) footprinting to demarcate the perimeter of the nuclear matrix extract binding activity along site A of the \textit{COL1A1} 5' regulatory region (15). Here, we used DMS interference fingerprinting to further delimit the Nmp4 protein-DNA contacts and to compare Nmp4/11H and Nmp4/21H association with this consensus element (Figure 4). The Nmp4/11H isoform contains eight zinc fingers whereas the Nmp4/21H has six. Both the Nmp4/11H and Nmp4/21H fingerprints extended across nucleotides -3463/-3458 (Figure 4), consistent with the MPE footprint of the nuclear matrix extract obtained in our earlier study (15). The sequence of the opposing strand in the area of the Nmp4 binding site contains only Cs and T's and therefore was not suitable for methylation interference analysis (data not shown).

\textit{The N-terminus of Nmp4/21H exhibits strong transactivation capacity when tethered to the GAL4 DBD but this is masked within the context of the full-length protein.}

We used Nmp4-GAL4-DBD chimeras, consisting of either the full-length Nmp4/21H protein or its truncated derivatives, fused to the GAL4-DBD, to identify domains exhibiting autonomous transactivation capacity. This was accomplished by co-transfecting 293T cells with the chimeras and a reporter construct driven by a basic promoter linked to five copies of the GAL 4 consensus sequence (see Materials and Methods). The N-terminus [1-187aa] of Nmp4/21H exhibited strong transactivation capacity and increased promoter activity 20-30 fold over baseline promoter activity (Figure 5). The poly[QA] [396-453aa] motif lacked transactivation activity, but in combination with the C-terminus [396-518aa] showed a weak transactivation capacity increasing promoter activity ~2-fold over baseline (Figure 5).

Interestingly, the strong transactivation activity of the N-terminus was completely masked within the context of the full-length Nmp4/21H-GAL4-DBD chimera (Figure 5). In fact, the full-length Nmp4-GAL4-DBD chimera [1-518aa] and the zinc finger-GAL4-DBD chimera [231-395aa] repressed promoter activity 5- and 9-fold below baseline, respectively (Figure 5). The portion of the Nmp4 protein [188-230aa], that includes the AT-hook/SH3-binding domain, significantly attenuated the strong transactivation
capacity of the N-terminus [1-187aa] but did not completely inhibit it (2.1-fold over baseline), suggesting that the zinc finger domain contributed to this masking effect (Figure 5). Promoter-reporter constructs lacking the GAL4 DNA binding sites did not respond to the various Nmp4/21H-GAL4-DBD chimeric constructs (data not shown) indicating the absence of background effects caused by the zinc finger-containing constructs targeting other sites.

The native Nmp4/21H protein acts as an activator within the context of the MMP7 promoter.

To determine whether the Nmp4/21H regulatory domains behave as autonomous modules, or instead, are sensitive to the identity of their attached DBD or to their DNA-binding state, we removed the GAL4-DBD. Kidney 293T cells were co-transfected with a human MMP7 promoter-reporter construct, containing the homopolymeric (dA:dT) Nmp4 consensus sequence and an expression vector containing either the full-length Nmp4/21H isoform or one of its truncated derivatives. Full-length Nmp4/21H isoform [1-518aa] upregulated the basal activity of the MMP7/luciferase promoter-reporter construct approximately 2-fold over the empty pcDNA-3™ expression vector (Figure 6). This was consistent with previous observations on the effect of overexpression of Nmp4 proteins on the promoter activity of MMP1, MMP3, and MMP7 (14). Next, we compared the transactivation capacity of Nmp4/21H with its truncated derivatives. The Nmp4/21H zinc finger domain construct [202-398aa], lacking the N-terminus, AT-hook/SH3-binding, poly[QA], and C-terminus domains, upregulated the activity of the MMP7 reporter-promoter construct approximately 1.3-fold over the empty expression vector but was significantly less efficacious than the full-length Nmp4/21H (p=0.024, Figure 6). The truncated Nmp4/21H derivative missing only the poly[QA] domain [Δ401-453aa] exhibited a similar attenuation of transactivation capacity as the zinc finger domain-only construct (p=0.020, Figure 6). This demonstrates that the the poly[QA] domain contributes to transactivation in the context of its native protein and DNA binding site but was silenced when tethered to the GAL4-DBD. The truncated derivatives, missing either the AT-hook/SH3-binding domain [Δ188-199aa] or the N-terminus [186-518aa], both
upregulated transcription of the MMP7 promoter 1.6-fold over the empty expression vector and did not significantly differ from the activity of the full-length Nmp4/21H protein (Figure 6). Finally, removal of the C-terminus [1-453aa] had no significant effect on the transactivation capacity of Nmp4/21H (Figure 6).

DISCUSSION:

The Nmp4 zinc finger domain serves multiple functions including DNA-binding and, as shown in our previous study, as a nuclear localization signal (NLS) and nuclear matrix targeting signal (NMTS) (31). Nmp4 proteins have 5-8 zinc fingers (13). At least five zinc fingers (#1-5 or #4-8) are required for exclusive nuclear localization and the number and arrangement of the zinc fingers encode specific nuclear matrix "zip codes" for these proteins (31). The Nmp4 zinc fingers mediate association with the nuclear matrix in the absence of DNA (31), consistent with the capacity of this domain to mediate protein-protein interactions (32). The capacity of the Cys2Hys2 zinc finger domain to mediate a variety of functions in one protein is not unusual (2, 33).

In the present study, EMSA indicated that only zinc fingers #2, #3, and #6 were necessary for mediating association of the native isoform Nmp4/21H with the COLIA1/MMP regulatory site. The DMS interference analysis indicated that the Nmp4/21H and 11H binding pattern over the homopolymeric (dA:dT) DNA consensus element were similar. This suggests that zinc fingers #4 and #5 of the 11H isoform participate in DNA-binding and are not moved out of the way. This is possible due to the amino acid sequence similarity between zinc fingers #2 and #4 (90% identity) and between zinc fingers #3 and #5 (81% identity). Additionally, since the COLIA1/MMP consensus sequence consists of 16 T residues out of 19 nucleotides, nearly all, if not all, binding sites are equivalent, eliminating any steric hindrance introduced by the organization of the binding element. Although zinc fingers #2, #3, and #6 are the minimum DNA-binding domain of the native truncated derivative Nmp4/21H, point mutagenesis and finger "swapping" may further clarify the interactions with this unusual consensus sequence. The present data are consistent with observations of other Cys2Hys2 zinc finger proteins in which only two to four tandemly arranged zinc fingers are required
for DNA recognition (2). Additional fingers may be necessary for high affinity and high specificity of binding or for mediating protein-protein interactions (2, 32, 33). The Nmp4 protein zinc fingers #1, #7, and #8 contribute to both the NLS and NMTS (31).

To our knowledge, Nmp4 is the only known Cys2His2 zinc finger protein that recognizes homopolymeric (dA:dT) DNA. However, a small number of zinc finger proteins are known to bind non-homopolymeric AT-rich DNA (34-38) and these share some characteristics with Nmp4. The *Drosophila* chorion transcription factor CF2-II also uses a subset of three Cys2His2 zinc fingers to associate with the minor groove of an AT-rich consensus sequence (36, 39, 40). Interestingly, CF2-II zinc finger #4, required for association with the AT-rich minor groove, exhibits a 62% homology to Nmp4 zinc finger #3. MIG1 has two Cys2His2 zinc fingers that bind to a GCGGGG motif in addition to an AT-rich region 5' to this consensus sequence (41). The MIG1 protein recognizes the unique structure of the AT-rich region, the consensus context (42), instead of the specific sequence, and mediates bending within the AT element (41). Similarly, Nmp4 induced DNA bending upon association with the homopolymeric (dA:dT) consensus element (15).

The Nmp4 zinc fingers may recognize the local structural contour of the *COL1A1/MMP* regulatory site, i.e., the rigid narrow minor groove of the homopolymeric (dA:dT) DNA instead of the sequence unambiguously presented in the major groove. The binding of the zinc finger domain alone to the *COL1A1/MMP* consensus element was sensitive to distamycin, but not to methyl green, indicative of an association with the minor groove (29, 30). Interestingly, the AT-hook/SH3-binding domain conferred weak methyl green sensitivity to the full-length Nmp4/21H protein. The AT-hook domain, as its name implies, typically binds to the narrow minor groove of AT-rich DNA (6, 16, 42, 43), although a weak association of this domain with the DNA major groove has been observed in HMGA proteins (44). The recognition of DNA conformation rather than the nucleotide sequences is well known in several minor groove-binding transcription factors including TATA-box-binding protein, integration host factor, SRY, LEF-1, and HMGA (45). Upon binding, these proteins induce conformational changes in the DNA, thereby facilitating the assembly of higher order transcription enhancer complexes (45).
The present data demonstrate that the zinc finger-mediated binding to the homopolymeric (dA:dT) consensus element has profound effects on the transactivation capacity of the Nmp4 protein domains. Tethering the Nmp4 proteins to the GAL4 DBD, thus neutralizing the effects of zinc finger association with the homopolymeric (dA:dT) consensus sequence, revealed that the N-terminus containing serine/threonine-rich domain had a strong transactivation capacity that was masked by the full-length protein. In fact, the full-length Nmp4/21H-GAL4-DBD chimera repressed baseline transcription of the heterologous promoter. However, removal of the GAL4 DBD from the Nmp4/21H chimeras resulted in the full-length Nmp4/21H upregulating \(MMP7\) promoter activity. Additionally the poly[QA] domain, silent in the Nmp4/21H-GAL4-DBD chimeras, contributed significantly to this activation. The N-terminus and AT-hook/SH3-binding domains made no significant contributions to transactivation. Little to no contribution from the Nmp4 C-terminus was required for upregulating \(MMP7\) promoter activity.

Together these data reveal that the Nmp4 proteins are comprised of modular domains whose isolated activities are dependent on the attached DBD. The homopolymeric (dA:dT) consensus site appears to contain information that is interpreted by the bound Nmp4 proteins. Nmp4 stimulates all tested MMP genes to similar degrees (14, and this study). Under physiological conditions, Nmp4 may act synergistically with other transcription factors resulting in higher stimulation of a natural promoter.

Allosteric control of transcription, i.e., the sensitivity of some transcriptional regulatory domains to their attached DBD or to their DNA-binding state is emerging as a significant governing mechanism in gene expression (46-48). Like Nmp4, the nuclear matrix transcription factors YY1 and Osf2/Cbfa1 both contain separable activation domains that function differently when tethered to a heterologous promoter by the GAL4 DBD as opposed to binding to their native consensus elements (25, 49). The three activation domains of Osf2/Cbfa1, AD1, AD2, and AD3, all contributed to transcriptional activation in the context of the native DBD. However, when attached to the GAL4 DBD in a heterologous system, only AD3 functioned independently while AD1, AD2, and the full-length protein, all were silenced (25). Additional examples of these context-specific transcription factors include USF2 (50), serum-response factor (SRF) (51), and ATF2 (52). These kinds of allosteric effects can have significant physiological consequences.
Recent evidence demonstrates that the allosteric effect on Pit-1 binding to the growth hormone promoter, in combination with other DNA binding factors, mediates Pit-1 activation of growth hormone expression in somatotropes, but repression in lactotropes (46). A similar phenomenon could contribute to the observed Nmp4-mediated repression of collagen in some osteoblast-like cells and activation in others (13, 53).

The present data, as well as earlier studies, support our hypothesis that Nmp4 proteins recognize a structural consensus context within the homopolymeric (dA:dT) sites, and that in turn, these AT-rich elements contain information that is interpreted by these proteins. The allosteric effects of the AT-rich DNA on the Nmp4 proteins may vary with slight differences in DNA sequence, and therefore structure, as well as with each specific Nmp4 isoform, of which there are many. This allows the Nmp4-DNA interactions to generate the pattern of regulation that is gene- and cell-type specific.

ACKNOWLEDGEMENTS:
This work was supported by NIH grant DK53796-01A1 (JPB) and NSF grant 9729669 (SJR). We thank Drs. Kyle Sloop and Andre van Wijnen for technical advice.

REFERENCES:


U S A. 95, 2938-2943.

Nucleic Acids Res. 24, 2567-2574.

U S A. 93, 2159-2164.


**FIGURE LEGENDS:**

**Figure 1:** The Nmp4 proteins are Cys$_2$His$_2$ zinc finger proteins that bind to homopolymeric (dA:dT) DNA. (A). Schematic representations the Nmp4 isoforms. Abbreviations: $I_1$ and $I_2$ inserts 1 and 2, AT/SH3B, Class II SH3 binding domain overlapping with the AT-hook domain (vertical striped box); numerals over open boxes indicate the zinc finger number; poly[QA], poly glutamine-alanine domain (horizontal striped box). Accession numbers AF216804, AF216805, AF216806, AF216807. (B). Nmp4 binding sites are imperfect homopolymeric (dA:dT) sequences in the $COL1A1$ (13) and $MMP$ (14) promoters. The Nmp4 binding sites in the $MMP$ promoters are typically in the reverse orientation as those characterized in the $COL1A1$ promoter. Nmp4 binding sites are indicated by the vertical arrows.

**Figure 2:** EMSA of full-length and truncated derivatives of Nmp4/21H demonstrate that zinc fingers #2, #3, and #6 are necessary for DNA-binding. (A). Schematic representation of full-length Nmp4/21H and the truncated derivatives used in EMSA analysis with $COL1A1$ site A. Construct identity number is indicated in parentheses. The identity of the deleted ($\Delta$) amino acids in brackets [ ]. Abbreviations: AT/SH3B, Class II SH3 binding domain overlapping with the AT-hook domain (vertical striped box); numerals over open boxes indicate the zinc finger number; poly[QA], poly glutamine-alanine domain (horizontal striped box). The removal of poly [QA] domain in all constructs also includes poly [A] and poly [Q] in their deletion. (B). EMSA analysis of selected Nmp4/21H constructs. Numbers above wells represent construct identity number indicated in (A); FP, free probe; V, pcDNA3, parent expression vector.

**Figure 3:** EMSA, using (A) distamycin or (B) methyl green as competitors, demonstrates that the zinc finger domain associates with the minor groove of $COL1A1$ site A. (A). The full-length Nmp4/21H [21H] and the truncated derivatives containing only the zinc fingers (ZF) or missing only the AT-hook/SH3-binding domain (NO AT), are competed off the poly(dT) DNA sequence with distamycin, a competitor for the minor groove (0µM, 15µM, 20µM, and 50µM). (B). The full-length Nmp4/21H construct [21H] exhibited some sensitivity to the same concentrations of methyl green. This sensitivity
was attenuated with removal of the AT-Hook/SH3-binding domain [NO AT]. DNA-binding of the zinc finger domain alone [ZF] was not altered by methyl green.

Abbreviations: AT/SH3B, Class II SH3 binding domain overlapping with the AT-hook domain (vertical striped box); numerals over open boxes indicate the zinc finger number; poly[QA], poly glutamine-alanine domain (horizontal striped box); the identity of the amino acids (aa) or deleted (∆) amino acids in brackets [ ].

**Figure 4:** DMS interference fingerprinting analysis reveals that Nmp4/11H (A) and Nmp4/21H (B) have similar binding patterns along site A (-3518/-3406 nt) of the COLIA1 gene. Asterisks indicate adenine nucleotides at which methylation interfered with Nmp4 isoform binding. Abbreviations: GA, Maxam-Gilbert sequencing ladder; F, methylated input DNA; U, unbound methylated DNA; B, methylated DNA bound to Nmp4 isoforms (11H or 21H).

**Figure 5:** The N-terminus of Nmp4/21H exhibits strong transactivation capacity that is masked by the full-length protein in a heterologous system. To identify domains exhibiting autonomous transactivation capacity, kidney 293T cells were transfected with Nmp4-GAL4 chimeras and a reporter construct driven by a basic promoter linked to five copies of the GAL 4 consensus sequence (5X-GAL4-LUC, see Materials and Methods for details). The N-terminus [1-187aa] exhibited strong transactivation capacity, whereas the C-terminus + the poly QA region [396-518aa] showed much weaker transactivation capacity. The strong transactivation activity of the N-terminus was completely masked within the context of the full-length Nmp4/21H protein. The portion of Nmp4 protein [188-230aa] that includes AT-hook/SH3-binding domain significantly attenuated the strong transactivation capacity of the N-terminus [1-187aa]. The full-length protein [1-518aa] as well as the zinc finger domain alone [231-395aa] significantly repressed the basal transcriptional activity of the GAL4-promoter/reporter construct. Data show mean ± SE and is representative of three independent experiments. Abbreviations: DBD, GAL4 DNA-binding domain; AT/SH3B, Class II SH3 binding domain overlapping with the AT-hook domain (vertical striped box); numerals over open boxes indicate the zinc finger number; poly[QA], poly glutamine-alanine domain (horizontal striped box), the
poly [QA] domain in all constructs also includes poly [A] and poly [Q]; TATA, TATA box; LUC, luciferase.

**Figure 6:** Removal of the GAL4-DBD from the Nmp4/21H chimeras and allowing the protein to interact with the homopolymeric (dA:dT) response element reveals that this protein acts as an activator within the context of the *MMP7* promoter. Full-length Nmp4/21H [1-518aa] induces a two-fold upregulation of the *MMP7* promoter-luciferase reporter construct in 293T kidney epithelia cells as compared to the pcDNA3 parent expression vector alone. The Nmp4 truncated derivatives exhibit an attenuated transactivation capacity of the *MMP7* promoter-luciferase reporter construct in 293T kidney epithelia cells as compared to the full-length Nmp4/21H protein. The zinc finger domain construct [202-398aa] upregulated MMP7 promoter activity only by 1.3-fold, equivalent to only removing the poly[QA] domain [∆401-453aa]. The transactivation activity of both of these truncated derivatives were significantly less than the full-length protein. Removal of the N-terminus [186-518aa] or the AT-hook/SH3-binding domain [∆188-199aa] did not significantly attenuate the transactivation capacity. Data represent mean ± SE of five experiments. Abbreviations: ∆, deleted; AT/SH3B, AT-hook/SH3-binding domain; numerals indicate the zinc finger (ZF) number; poly[QA], poly glutamine-alanine domain. The removal of poly [QA] domain in all constructs also includes poly [A] and poly [Q] in their deletion. The asterisk and diamond indicate significant difference (p<0.05) between comparison groups using ANOVA and Tukey’s post hoc test, see text.
A.

\[ I_1 \quad \text{AT/SHB} \quad I_2 \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad \text{poly[QA]} \]

- \text{Nmp4/11H}  
- \text{Nmp4/13H}  
- \text{Nmp4/21H}  
- \text{Nmp4/28H}

B.

\text{COL1A1 PROMOTER}

\text{Site A [-3469/-3450 nt]}  
\text{TTCTTTTTTTTTCTTTTCT}

\text{Site B [-1574/-1555 nt]}  
\text{TTATTTTTTTTTCTTTGCCT}

\text{MMP PROMOTERS: (G/C)AAAAAA(A)}

\text{MMP7}  
\text{MMP3}  
\text{MMP1}

-1500 \quad -1000 \quad -1
## A.

<table>
<thead>
<tr>
<th>Nmp4/21H construct</th>
<th>AT/SH3B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>poly[QA]</th>
<th>DNA-binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) [1-518]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(2) [Δ454-518]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(3) [Δ401-453]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(4) [Δ399-518]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(5) [Δ1-185]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(6) [Δ1-201]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(7) [Δ188-199]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(8) [Δ231-395]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(9) [Δ1-201:Δ399-518]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(10) [Δ308-395]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(11) [Δ231-314]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(12) [Δ231-286]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(13) [Δ259-314]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(14) [Δ231-258:Δ287-314]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(15) [Δ231-258]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(16) [Δ259-286]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(17) [Δ287-314]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(18) [Δ231-258:Δ315-344]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-</td>
</tr>
<tr>
<td>(19) [Δ231-258:Δ345-376]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(20) [Δ231-258:Δ375-395]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
<tr>
<td>(21) [Δ231-258:Δ338-395]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>+</td>
</tr>
</tbody>
</table>
Control  

<table>
<thead>
<tr>
<th>Region</th>
<th>GAL4-DBD-</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1-518aa]</td>
<td></td>
</tr>
<tr>
<td>[1-187aa]</td>
<td></td>
</tr>
<tr>
<td>[1-230aa]</td>
<td></td>
</tr>
</tbody>
</table>

[231-395aa]  

[396-453aa]  

[396-518aa]  

PROMOTER  

<table>
<thead>
<tr>
<th>Region</th>
<th>GAL4-DBD-</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1-518aa]</td>
<td></td>
</tr>
<tr>
<td>[1-187aa]</td>
<td></td>
</tr>
<tr>
<td>[1-230aa]</td>
<td></td>
</tr>
<tr>
<td>[231-395aa]</td>
<td></td>
</tr>
<tr>
<td>[396-453aa]</td>
<td></td>
</tr>
<tr>
<td>[396-518aa]</td>
<td></td>
</tr>
</tbody>
</table>
AT/SH3B [1-518aa] [1-454aa] MMP7 [-1400 nt] PROMOTER LUC

[1-518aa] [1-453aa] 
[202-398aa] [188-199aa] [186-518aa] [401-453aa]

FOLD DIFFERENCE FROM ACTIVITY OF CONTROL

1X 1.2X 1.4X 1.6X 1.8X 2.0X 2.2X

∗ ∗ ∗ ∗♦

by guest on September 1, 2017 http://www.jbc.org/ Downloaded from
DNA binding and gene activation properties of the Nmp4 nuclear matrix transcription factor
Kitti Torrungruang, Marta Alvarez, Rita Shah, Jude E. Onyia, Simon J. Rhodes and Joseph P. Bidwell

J. Biol. Chem. published online February 26, 2002

Access the most updated version of this article at doi: 10.1074/jbc.M107496200

Alerts:
- When this article is cited
- When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/early/2002/02/26/jbc.M107496200.citation.full.html#ref-list-1