Nucleic Acids Research, 1997, Vol. 25, No. 6 **1225–1232**

Isolation and characterization of the human nucleophosmin/B23 (NPM) gene: identification of the YY1 binding site at the 5′ enhancer region

Pui K. Chan*, Fung Y. Chan, Stephan W. Morris[1] and Zhong Xie

Department of Pharmacology, Baylor College of Medicine, Houston, TX 77030, USA and [1]Department of Experimental Oncology, St Jude Children's Research Hospital, Memphis, TN 38105, USA

Received November 13, 1996; Revised and Accepted January 23, 1997

DDBJ/EMBL/GenBank accession nos U89309–U89322

ABSTRACT

NPM (nucleophosmin/B23) is a major nucleolar protein which is 20 times more abundant in tumor or proliferating cells than in normal resting cells. Recently, it was found that NPM gene is located at the breakpoints of the t(2:5), t(3:5) and t(5:17) chromosome translocation. To understand the human NPM gene's structure and regulation, four genomic clones were isolated from the human chromosome 5 library and their DNA sequences analyzed. The human NPM gene has 12 exons of sizes ranging from 58 to 358 bp. The chromosome breakpoint for t(2:5) and t(5:17) translocation is within intron 4 and the breakpoint for t(3:5) translocation is within intron 6. The initiation site is located 96 bp upstream from the ATG site. A typical TATA box (at –25 nt) and a GC box (at –65 nt) were identified in the promoter region. We identified two gel-shift bands (A and B) with DNA fragment E (–741/–250 nt) by EMSA. A DNA footprint was observed at (–371/–344 nt) with the nuclear extract. A double stranded DNA with the footprint sequence (–371/–344 nt) competed the formation of gel-shift bands A and B in EMSA suggesting that proteins A and B bind to the footprint region. We confirmed that protein A is transcription factor YY1. These results suggest that YY1 may play a role in NPM gene expression. This is the first report on human NPM gene structure and sequence.

INTRODUCTION

Nucleophosmin (NPM, also called protein B23, numatrim, NO38) is a nucleolar phosphoprotein 20 times more abundant in cancer cells than in normal resting cells (1). Its putative function is the assembly and/or transport of ribosome. NPM is a mobile protein; its cellular localization is affected by growth conditions and influenced by certain cytotoxic drugs. During serum starvation (2) or treatments with anticancer drugs—daunomycin, actinomycin D, camptothecin or toyocamycin—it shifts from nucleoli to nucleoplasm (NPM-translocation) (3–7). The cellular location of NPM is also dependent on GTP and ATP levels. When the *de novo* synthesis of GTP is inhibited and cells are depleted of GTP, NPM-translocation occurs (8). On the other hand, when ATP levels are depleted, NPM-translocation is blocked and newly synthesized ribosomal RNA accumulate in nuclei and cannot be transported to the cytoplasm (9–10). NPM is associated with pre-ribosomal particles and other nuclear proteins as well. The proteins it associates with include transcription factor YY1 (11), nucleolar p120 protein (12) and HIV protein Rev/Rex (13–14).

NPM biosynthesis is related to cell proliferation and mitogenesis. We found that NPM mRNA is 50-fold higher in Novikoff hepatoma and 5-fold higher in hypertrophic rat liver than in normal rat liver (1). Feuerstein *et al*. (15–16) reported that when B cells, T cells and Swiss 3T3 cells were stimulated with various mitotic agents, NPM synthesis increased. On the other hand, down regulation of NPM was observed in Jurkat T-lymphoblasts during apoptosis (17). These studies indicate that NPM expression is associated with cell growth.

The rat NPM gene has been isolated and characterized (18). However, the structure and sequence of the functional human NPM gene is not yet known. Previously, we identified seven human processed pseudogenes of NPM (19). Genomic blot analysis indicated that there are at least 10 copies of NPM gene per haploid human genome. With the recent information that the chromosome breakpoints of t(2:5), t(5:17) and t(3:5) chromosome translocation associate with NPM gene (19–25), we searched for the functional gene in the human chromosome 5 library and isolated four clones. Here, we report the gene's physical structure and DNA sequence. The 5′ region and the DNA sequence of the intron 4 where t(5:2) chromosome break occurs were characterized. This is the first report on human NPM gene structure and sequence. We also found that the transcription factor YY1 (26–29) binds to the 5′ region of the NPM gene, suggesting that YY1 may play a role in NPM gene expression.

MATERIALS AND METHODS

Genomic sequences analysis

A human chromosome 5 genomic library was obtained from the American Type Culture Collection (cat. no. LA05NL03). The library was screened with cDNA probes (1) and the first intron probe (p16-3/1.2S) derived from the chromosome walk (20). Screening libraries, Southern blot analysis, DNA cloning, PCR and DNA sequencing by dideoxynucleotide termination reactions were performed according to the standard methods described in Sambrook *et al*. (30). Primers used in PCR and in cloning were derived from various regions of the cDNA.

*To whom correspondence should be addressed. Tel: +1 713 798 7902; Fax: +1 713 798 3145; Email: pchan@bcm.tmc.edu

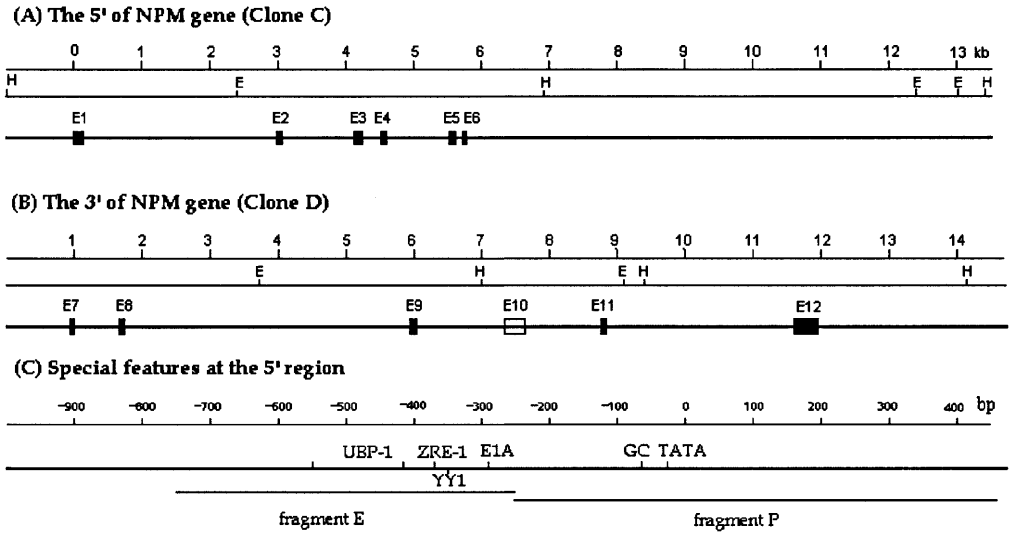

Figure 1. Physical structure of human NPM gene. E and H represent the locations of restriction enzyme site of *Eco*RI and *Hin*dIII, respectively. The relative size and position of exons 1–11 are represented by E1–E11. (**A**) The 5′ portion of NPM gene (clone C). (**B**) The 3′ portion of NPM gene (clone D). (**C**) Relative locations of special sequences in the 5′ region.

Determination of the initiation site

To identify the initiation site, the primer extension method was employed. Antisense primer P5 (1–17 from the ATG site of cDNA) was end-labeled with [γ-^{32}P]ATP by polynucleotide T4 kinase. RNAs were prepared from HeLa cells using the GIBCO/BRL Trizol reagent. Poly-A RNA were purified with oligotex-dT gel. Annealing of the primer to RNA was conducted at 43°C for 8 h. Primer extension was carried out by AMV reverse transcriptase at 42°C for 90 min. The primer extended fragment was analyzed in an 8% sequencing gel with both markers and a sequence ladder derived from the P5 primer.

Electrophoretic mobility shift assay (EMSA)

The procedure of EMSA was according to 'Current Protocol, vol II' (31) with slight modifications. DNA fragments, P (–249 to +465) and E (–741 to –250), were prepared by cleavages of the 5′ DNA with *Mlu*I (which cleaves at –249) and *Sac*I (which cleaves at –741 and +465) restriction enzymes. The reaction mixture (25 µl) containing the MCF-7 or HeLa cell nuclear (32) extract (1–3 µg protein), end-labeled DNA fragments (0.5–5 ng), binding buffer [12 mM HEPES, pH 7.9, 4 mM Tris–HCl, 60 mM KCl, 1 mM DTT, 1 mM EDTA, 0.05% NP-40, 5 mM $MgCl_2$, ($MgCl_2$ is required for YY1 binding), 12% glycerol], poly dI-dC (1.5 µg per reaction), BSA (300 µg/ml) ± competitors, were incubated at 25°C for 15 min. The mixture was loaded onto a 4% polyacrylamide gel for electrophoresis. Retardation of DNA fragments due to binding of protein factors was detected by autoradiography.

DNase I footprinting

DNA fragment E was end-labeled either by T4 kinase or Klenow reaction. The labeled DNA was incubated with nuclear extracts (3–15 µg) in a final volume of 50 µl containing 12 mM HEPES, pH 7.9, 4 mM Tris–HCl, 60 mM KCl, 1 mM DTT, 1 mM EDTA, 12% glycerol and 2 µg of poly(dI-dC) and BSA (100 µg/ml). After 15 min of incubation at 25°C, DNase I (0.5–1.5 unit) was added to the mixture and allowed to digest for 1–1.5 min at 0°C. The reaction was stopped and precipitated by adding ethanol/tRNA/ammonium acetate buffer mixture (31). The reaction product was analyzed by a 6% polyacrylamide–8 M urea gel. The position of the footprint was determined by an adjacent sequencing gel ladder (33).

Construction of 5′ deletion mutants and luciferase assay

Three DNA fragments from the 5′ region of the NPM gene: I (–741/+93), II (–405/+93) and III (–43/+93, which is the basal promoter with a TATA box) were prepared by PCR with appropriate primers. The 3′ end of these clones is located 3 bases upstream from the ATG site. These DNA fragments were subcloned into the luciferase reporter gene vector pXP1 (33) and were transfected into HeLa cells by electroporation (34) under the following conditions: 1.5 kV/cm, 1 ms pulse width, 2 trains with 15 pulses. Cells were cultured in Petri dishes (35 mm) for 48 h before harvest. Under these conditions, ~20% of cells were transfected. Luciferase activity (RLU, relative luciferase unit) was determined with a Luminometer using D-lucuferin as substrate. The luciferase activity for each sample was normalized with their co-transfected β-galactosidase activity (RLU/β-gal unit) which was determined by the ONPG colorimetric assay.

RESULTS

Gene structure of human NPM

About 2×10^5 plaques from the human chromosome 5 library were screened (see Materials and Methods). Four clones were identified and purified which have inserts of 13–20 kb. Restriction mapping and Southern blot analysis with probes derived from the 5′ and 3′ region of the NPM cDNA indicated that three clones (A, B and C) contained the 5′ portion of the NPM gene and one clone (D or clone 21-3) (20) contained the 3′ region. The inserts of these clones were fragmented with restriction enzymes and subcloned into Bluescript SK vectors. Subcloning of individual introns was

also achieved through PCR techniques using cDNA primers. The DNA sequences of these clones were determined.

Figure 1 shows the restriction enzyme sites, the relative size and the location of exons in clones C and D. More than 60% of the gene sequence has been determined (sequences have been submitted to the GenBank). Since some of the intron sequences (except introns 3, 4 and 5) were not fully determined, the relative position between exons was determined by PCR with cDNA primers and by restriction mapping. Clone C contains exons 1–6 and clone D contains exons 7–11. The combined sequence of the NPM gene in clones C and D span ~25 kb. Since the overlapping sequence has not been found between the 3′ of the C clone and the 5′ of the D clone, the estimated total length of the NPM gene could be >25 kb. There are 12 exons with sizes ranging from 58 (exon 7) to 358 bp (exon 12). The DNA sequences of the exons are identical to the cDNA sequence (1) except that the exon 11 is not identified in the major gene product. Table 1 shows the exon/intron junctions of the human NPM gene. The same exon/intron junctions are observed in the rat NPM gene (18). Figure 2A shows the DNA sequence of the 5′ region of the NPM gene and Figure 2B shows the sequence of the intron 4 where the chromosome breaks occur (discussed in the following).

Table 1. The exon/intron junctions of the human and the rat NPM gene

Exon no.	Exon size	5′ Splicing site	3′ Splicing site
Exon 1	154	TTC G/gtaactgc	cag/GT TGT
	154	TTC G/gtaactgg	cag/GT TGT
Exon 2	80	ACG/gtacttaa	cacag/GTC
	80	ACG/gtatgtaa	cacag/GTC
Exon 3	120	ACG/gtaagggca	ttctag/GTT
	120	AC<u>A</u>/gtaagtaca	ttttag/GTT
Exon 4	94	GTA G/gtatgt	ttag/CT GTG
	94	GTA G/gtagga	ttag/CT GT<u>A</u>
Exon 5	107	CAG/gtaga	aaag/AAA
	107	CAG/gtaaa	tcag/AAA
Exon 6	65	GAA GA/gtaagt	tgtag/T GAT
	62	GAA GA/gtaagt	tgtag/T GAT
Exon 7	58	AAA/gtgagt	tgcag/TCT
	58	AAA/gtacat	tgcag/TCT
Exon 8	87	AAA/gtaag	ttcag/GGA
	84	AA<u>G</u>/gtaag	ttcag/GG<u>T</u>
Exon 9	102	AAA/gtgagt	tgcag/GCG
	102	AAA/gtgagt	tgcag/GCG
Exon 10	332	..poly (A) site	aatag/GGT
	324	..poly (A) site	aatag/GGT
Exon 11	75	GAG/gtcaactg	tccag/GCT
	75	GAG/gtaactga	cctag/GCT
Exon 12	358	..poly (A) site	
	346	..poly (A) site	

The upper rows are the human sequences, the lower rows are the rat sequences; upper case letters represent the exon sequences; lower case letters represent the intron sequences. The difference between human and rat NPM coding sequences were underlined.

Location of the chromosome break point

It was recently discovered that the 5′ region of the NPM gene is fused with other genes in lymphoma cells having t(2:5), t(5:17) or t(3:5) chromosome translocation (20–25). Based on the RT–PCR transcript sequence of the fusion protein derived from cells with t(2:5) and t(5:17) chromosome translocation (20,24), the N-terminal 117 amino acids of NPM are fused to anaplastic lymphoma kinase (ALK) or retinoic acid receptor alpha (RARA). In t(3:5) chromosome translocation, the N-terminal 175 amino acids of NPM are fused to myepodysplasia leukemia factor 1 (MLF) (25). Compared to the cDNA sequence (1) and exon/intron positions of human NPM gene (Table 1 and Fig. 1), the 117 and 175 amino acids of NPM are located at the junctions of exon 4/exon 5 and exon 6/exon 7, respectively. Accordingly, the chromosome breakpoint for t(2:5) and t(5:17) translocation is within intron 4 and the breakpoint for t(3:5) translocation is within intron 6. The precise breakpoint location within these introns is not known. Intron 4 is 911 bp long (Fig. 2B). Sequence analysis indicated that this intron contains two *Alu* sequences (dotted underlines) which have 70–80% sequence homology to the *Alu* sequences observed in human ABL and BCR (break point cluster regions) genes (36). We also found that the 3′ end of this intron is particularly enriched in T (denoted with italic).

Identification of the initiation site and sequence analysis of the 5′ region of NPM gene

To identify the initiation site for NPM transcription, the primer extension method was employed. Figure 3 shows the autoradiograph of the primer extended DNA fragments. We found one major and three minor primer extended fragments with lengths of 111, 112, 113 (major) and 114 bp. Compared to the adjacent sequencing gel, the major fragment (113 bp long) extended to the cytosine (C) at 96 bp upstream from the ATG site. Accordingly, the initiation site (+1) is assigned to this position (Fig. 2A). A well defined TATA box (TATATAA) and a GC box (GGCG) were observed at –25 nt and –66 nt, respectively (Fig. 2A). Sequence analysis with a computer search program identified four potential *cis*-elements upstream from the promoter region. They are (i) E1A-F (AGGACGT) (37) at –293; (ii) UCR core (CGCCATTTT) (26) at –352; (iii) ZRE-1 (TTACACA) (38) at –377; and (iv) UBP-1 (CTCTCTGG) (39), located at –414 and –550.

Electrophoretic mobility shift assay (EMSA)

To identify proteins that bind to the 5′ region of the NPM gene, an electrophoretic mobility shift assay (EMSA) was employed. Two DNA fragments (P and E) were used (Fig. 1). Fragment P (–249 to +465) contains the TATA box, the GC box and part of the first intron. Fragment E (–741/–250) is located 5′ from fragment P. Nuclear protein extracts of HeLa or MCF-7 cells were incubated with the DNA fragments. Retardation of these fragments by proteins was identified (Fig. 4). One gel-shift band (arrow) was identified using fragment P (Fig. 4A), and two gel-shift bands (A and B) were identified using fragment E (Fig. 4B). Formation of these gel-shift bands was not affected by competition with non-specific competitors: calf thymus DNA (which was sheared to sizes ~500 bp), poly dI-dC/poly dI-dC and BSA, but was competed out by the unlabeled DNA fragments.

A DNA sequence of the 5' region.

```
                                                                    aagctttatgtgacacaggag-1201
ctctcagaaaggaacagttaaattataacatgtccatattatgcgttagtatgtgtccacaactaacccataatatggacatgttataatttaactgttc-1101
attcctctcttgatagacactgaaatcgtctctataatttttcctgtaagaaataatgccttggggtgggggtgtcttcctttcgaggctatcatttgtac-1001
ttcctcacttcttaaatttgtttgatatgttgaggcttaaaaaaaaaagattacgggtcttcagaacgccccaatgcctgcggggtgctggggtgcctcg -901
gagtttgtcaaggtgtagcttaaaacgttaattcccatgctcgctcccaaatattcgaattaaggcttctgtagttcttacaagtcacccgctttctttc -801
aggaggaatgttaaagattaagagcggcaagaagtcagagtcggcccaccctccgagctctcttagggcgatgtccttgctaatttggagactgattcag -701
tccccttttggccccaagttacgtaaagataaggactttggagatgttttctcaggaaggacagagctgaaaaacaaagttccgtaagtctggccggta -601
actacatctttccttcctaacaaagaataagccgcaattcactctctctggcacctgaactttggggtaacgattaactgcgatccgaagccttcaggca -501 UBP-1
agtaaaaaattcctgaagtgattacgcctgtttggaggcttgcagggcactaggggatggggaaaggtgaatcgaggtgctctctggctcattcgcagcc -401 UBP-1
ggctaacccgccatatcttacacagctaagggctgccgacgccattttgcagggtgggctgcgcagactcttggcgggaggccggcgcgcttgagcggga -301 ZRE,UCR
gaggacgtggaagggttggaggtggggagtgcgcgttacgactggaaagcacgcgtgcgcacaggcgacagcagcggaggggtggggccagtgtacgagt -201  E1A-F
gcgcgtgctcggtgggagcccgcggagtacctggaaggaggtgggggcgaggtagaaaggagtgggggttgaaaagccttgcgcaggacggctacggtacg -101
ggggtgggagggcttcggagcacgcgcgcgga ggcg ggacttgggaagcgctcgcgagatcttcagggtcta tatataa gcgcgggagcctgcgtccttt   -1 GC,TATA
CCCTGGTGTGATTCCGTCCTGCGCGGTTGTTCTCTGGAGCAGCGTTCTTTTATCTCCGTCCGCCTTCTCTCCTACCTAAGTGCGTGCCGCCACCCGATGG  100  E1
AAGATTCGATGGACATGGACATGAGCCCCCTGAGGCCCCAGAACTATCTTTTCGgtaactgctgggggagctggagcgaggccgagcggggcctggtgg  200
cggtgagggtgggggtgaggggcgggaatccggctgcaaccgggtctggtggagaccgccagaccgacgggaggcctgagcgggtgggaagagctgcctg  300
agccctggacctggagggattggagccgcgggggagccggtggcgtgaaggggaggagacgagcggcctgaagcgtctggggcgtgagcggtccgagc  400
ccggggcctgaggtaaagtggaaccggccgcctggaagttcggcctttgccggggatacgtggcg----------------------------------  500
```

B DNA sequence of the intron 4.

```
CACTTAGTAGgtatgttatttttatatattatactacttagtttgtcctctttagtgcagttgcttggttcccagtttggacttaaagcatgggtatagt      E4
actactgtctttttaataggttccaatgtgagtctagaaattggagaggacaaataaatttttgggggcgggggggagaggaaatctttgctgtcacccag
gctagagtacagtggcacgatcttggctcactggaacctctccgggattcaagcgatcctcctgtctcagcctccccagtagctggggccacagacgtgc
accaccaagtccagttgcgtttcatagttatagtagagaagggtttcgtgatgttggccaggccgatcttgaactcctggccttaacgtgatctgcacg
ccctggcctctcaaagtgctagtattacacgtgtcagccactgtgcctggcctaaaaattatttttaatgaaagacagtctcattataacggctggagtg
cagtgatgtgatcatagcttgctatatcctcgaactgctactgggttcacctcagcctctggaatagctagaactacaggcacactccacgcctggctaa
ttttttttgtatatgtgcagatggggtctcagtatgttgcccagattggactcttggcctcaagtggtccgccttggcctccccaaagtgagattacagg
catgagccaccctccccaggcttcttgcatttaaaacctggcagtgaacattaggcctcaaaatacttttgttaaaaagttccttttcccatgtgctctt
tttttttttttttttaaatagaatagaagtctcagtttttagagtatttactatcagtgtttctttttttttctgacttcttgctgcttgagttttataat
gtctaataaattgtattttagCTGTGGAGGAAGATGCAGAGTCAGAAGATGAAGAGGAGGAGGATGTGAAACTCTTAAGTATATCTGGAAAGCGGTCTGC      E5
```

Figure 2. (**A**) DNA sequence of the 5′ region. (**B**) DNA sequence of intron 4. The TATA box, the GC box, the initiation site and the ATG site are blotted. Other features were underlined and described in the text. Exon sequences (E1, etc.) are typed in upper case letters and double underlined. Repetitive and *Alu* sequences were underlined with dots.

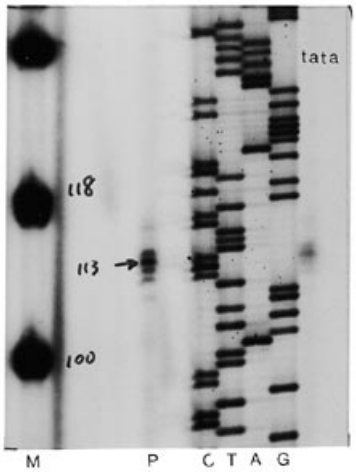

Figure 3. Determination of the initiation site. The primer extended DNA fragment (lane P) was analyzed on a 8% sequencing gel. Lane M has DNA size-markers; lanes G, A, T and C represent the nucleotide sequence of the 5′ region determined by dideoxynucleotide chain-termination method using the same primer. Compared to the adjacent sequencing gel and the markers, the major fragment (113 bp long) corresponds to an extension to 96 bp from the ATG site.

Identification of protein binding site by footprinting analysis

To determine the DNA region that bound to protein(s), DNase I footprinting technique was employed. Fragment E was end labeled with [γ-^{32}P]ATP and incubated with the nuclear extract of MCF-7 cells. The DNA region protected by proteins was determined by DNase I digestion and footprinting analysis. As shown in Figure 5, a footprint was identified from –371 to –344 with the sequence of GGGCTGCCGACGCCATTTTGCAGGGTGG (NPM-*cis*-1). To study whether the protein(s) protecting this region is related to the gel-shift band(s) observed in EMSA (in Fig. 4B), a double-stranded DNA oligo with the footprint sequence (NPM-*cis*-1) was made and used as a competitor in EMSA. As shown in Figure 6, this synthetic oligo effectively competed against the protein binding of fragment E in EMSA (lane 2–8). Both the major and minor gel-shift bands (proteins A and B) were effectively competed off by the oligo (12.5 ng). A modified oligo [NPM-*cis*-A, of which three Gs in the nonconsensus region of NPM-*cis*-1 are removed and inserted into the consensus region of the NPM-*cis*-1, (see Table 2)] was not effective in competition (lane 10–16). This result indicates that a specific DNA sequence is required for the binding of proteins A and B.

YY1 binds to NPM-*cis*-1 and fragment E

The footprint (NPM-*cis*-1) has the sequence CCATTTTG which is a potential binding site for the zinc finger DNA binding protein, YY1 (δ factor, UCRBP, NF-E1) (26–29). Table 2 shows the comparison of this sequence to other sequences that associate with protein YY1. To confirm whether YY1 is involved in the binding of DNA and causes the gel-shift, recombinant His-YY1 (a gift from Dr E. Seto) was employed in EMSA. Purified His-YY1 (a single band in SDS gel with >95% pure) was incubated with the end-labeled oligo NPM-*cis*-1. As shown in Figure 7, YY1 caused the gel-shift of NPM-*cis*-1 (lane 2). Compared to the gel-shift band positions induced by the nuclear protein extract which produced a major and a very weak minor band (lane 3), the band produced by YY1 corresponded to the major band. The slightly slower mobility of the YY1 shifted band could be due to a small difference in the charge and M.W. of the recombinant His-YY1. However, this difference in mobility was

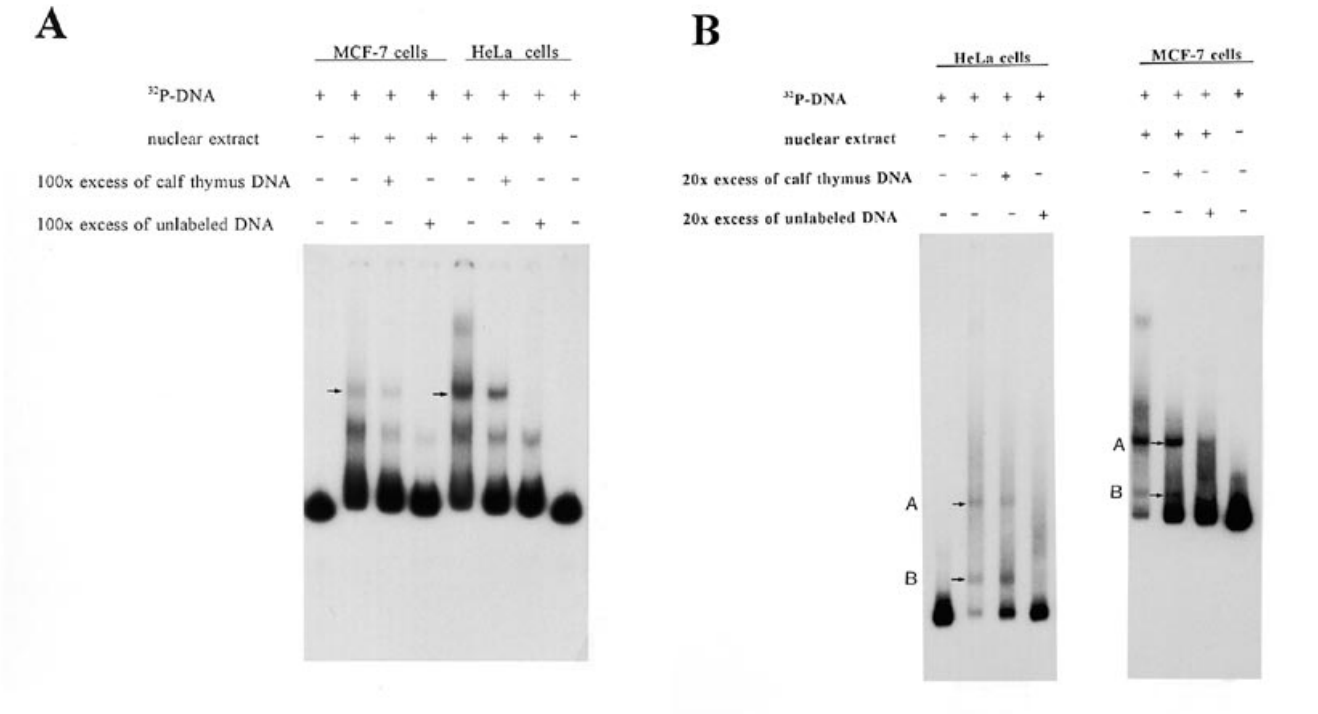

Figure 4. Electrophoretic mobility shift assay (EMSA). Nuclear extracts of HeLa or MCF-7 cells were incubated with (**A**) DNA fragment P (–249/+465) or (**B**) DNA fragment E (–741/–250). Gel-shift bands were denoted with arrows. These shifted bands were not affected by competition with calf thymus DNA (20–100-fold excess), poly dI-dC/poly dI-dC and BSA, but were competed out by the unlabeled DNA fragment.

not observed with the longer DNA fragment E. Addition of YY1 antibody to the reaction mixture containing nuclear protein extract (lane 5) or purified YY1 (lane 10) produced a weak supershift band. Only a small fraction of band was supershifted. This could be due to a partial dissociation of the immuno-complex under high ionic strength PAGE conditions (31). The supershift is specific because normal rabbit serum or purified rabbit IgG (lanes 6,7) had no such effect. These results indicate that YY1 binds to the footprint sequence NPN-*cis*-1.

Table 2. DNA sequences that bind to protein YY1

NPM (*cis*-1)	GGGCTGCCGACGCCATTTTGCAGGGTGGG
MuLV UCR (UCRBP)(25)	CTGCAGTAACGCCATTTTGCAAGGCAT
Adenovirus P5(YY1)(26)	AGGGTCTCCATTTTGAAGCGGG
L30,L32 mouse(delta)(27)	CTCGCTCCCCGGCCATCTTGGCGGCTGG
Igk3′enhancer(NF-E1)(28)	CCTACCCCACCTCCATCTTGTTTGATA
NPM-*cis*-1	GGGCTGCCGACGCCATTTTGCAGGGTGG
NPM-*cis*-A[a]	GGCTGCCGACGCgCATgTTgTGCAGTGG

[a]Double-stranded oligo NPM-*cis*-A was modified from NPM-*cis*-1 by rearranging three Gs in the nonconsensus region (underlined) to the consensus region (denoted by lower case letters).

The binding of purified YY1 to DNA fragment E was also studied with EMSA. We found that YY1 shifted fragment E to the major band position (protein A) like the nuclear protein did (data not shown). We also found tht addition of YY1 to the nuclear extract enhanced the intensity of the major band (data not shown). These results indicated that the major band was a result of binding to YY1.

Expression of the 5′ deletion mutant of NPM gene

A luciferase reporter gene assay was employed to study the 5′ DNA domain(s) for NPM expression. The 5′ deletion mutants of NPM gene were subcloned into the luciferase reporter gene vector pXP1 (34). These constructs were transfected into HeLa cells and the expression of luciferase activity was determined (Materials and Methods). As shown in Figure 8, the promoter/enhancer activity with the DNA sequence up to –405 (fragment II) including the YY1 binding site (–361/–353) is $93.19 \pm 19.45 \times 10^6$ RLU/β-gal unit. This activity is 157 times higher than the basal promoter (fragment III) activity ($0.59 \pm 0.12 \times 10^6$ RLU/β-gal unit). It was also found that the construct with fragment I (which extended from –405 to –741) has a lower activity ($48.01 \pm 9.8 \times 10^6$ RLU/β-gal unit), indicating a possible negative regulator element in the region between –741 and –405.

DISCUSSION

Sequence comparison between human and rat NPM genes

The exon/intron junctions for the human NPM gene are identical to those of the rat NPM gene (Table 1). The number and the size of exons between human and rat NPM are the same, although the human gene spans >25 kb while the rat gene spans ~11 kb. Many repetitive sequences were found in the introns of the human gene. While the sequence homology between the coding portions of human and rat NPM is ~94%, we observed only 75% sequence homology in the 5′ noncoding region, the promoter and the enhancer regions. Chang and Olson (40) identified two NPM

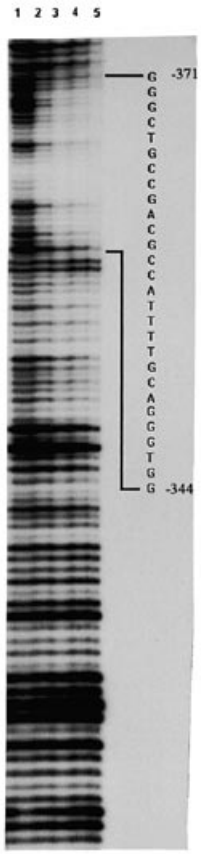

Figure 5. DNase I footprint analysis of binding sites. DNA fragment E (–741/–250) was end labeled with ^{32}P and incubated with the nuclear extract of MCF-7 cells. The DNA protected region was identified by DNase I digestion. A footprint was identified within the DNA sequence of –371 and –344. Samples in lane 1 contain no nuclear extract; lanes 2, 3, 4 and 5 contain 8, 16, 24 and 32 μg protein, respectively.

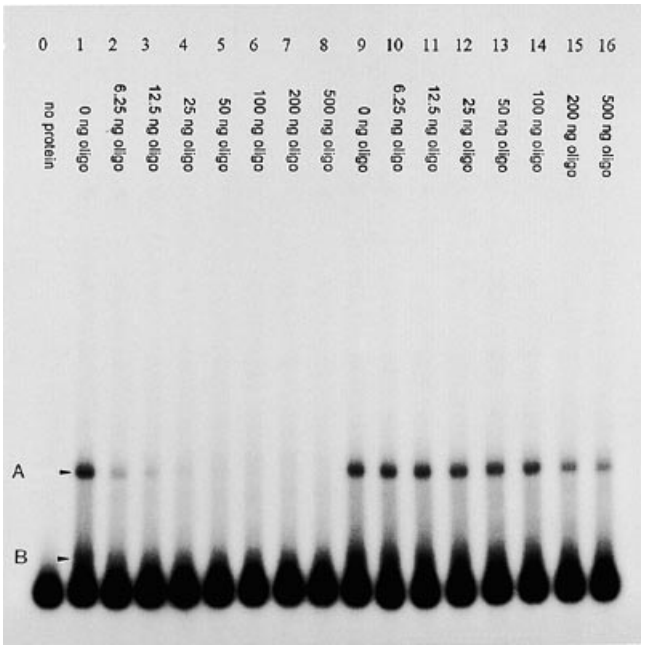

Figure 6. Competition of NPM-*cis*-1. DNA fragment E was incubated with nuclear extract of MCF-7 cells with the competing oligo NPM-*cis*-1 (lane 2–8) or the modified oligo NPM-*cis*-A (lane 10–16). Both the major and the minor gel shifted bands (arrows) were competed out by 12.5 ng of NPM-*cis*-1 but not by NPM-*cis*-A.

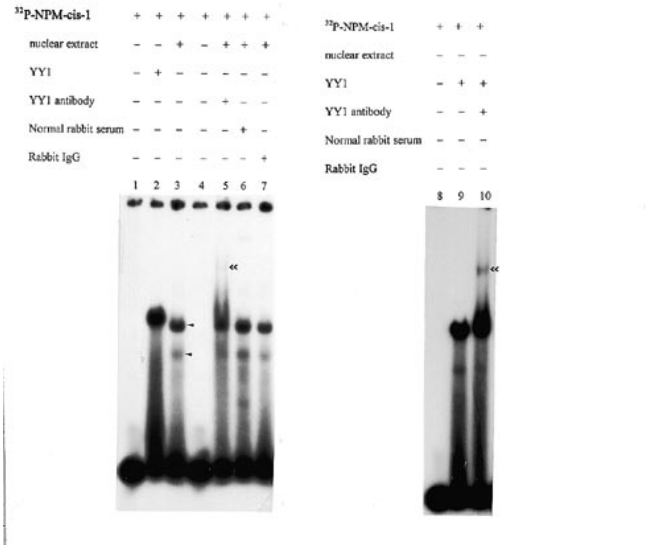

Figure 7. YY1 and nuclear protein extract bind to NPM-*cis*-1. Purified YY1 or nuclear protein extract was incubated with the synthetic oligo NPM-*cis*-1 and their binding was analyzed by EMSA. Lanes 1, 4 and 8, [^{32}P]DNA (NPM-*cis*-1) alone; lane 2 and 9, [^{32}P]DNA plus YY1 (0.5 μg); lane 3, [^{32}P]DNA plus nuclear proteins; lane 5, [^{32}P]DNA plus nuclear proteins and YY1 antibody; lane 6, [^{32}P]DNA plus nuclear proteins and normal rabbit serum; lane 7, [^{32}P]DNA plus nuclear proteins and rabbit IgG; lane 10, [^{32}P]DNA plus YY1 (0.5 μg) and YY1 antibody.

cDNA clones (B23.1 and B23.2) from rat tissues. The long form (B23.1) contains exons 1–9 and 11–12, and the short form (B23.2) contains exons 1–10. The short form of NPM (B23.2) is a result of alternative splicing of the NPM gene (18). A sequence corresponding to the rat's exon 10 (expressed in the short form) was also identified in the human NPM gene (with 80% sequence homology). The expression of exon 10 in human tissue is not known and remains to be investigated. Intron 4 of the human NPM gene where chromosome break occurs (discussed below) is longer (911 bp) than the corresponding rat sequence (558 bp), but there is no sequence homology between them.

Sequence analysis of intron 4, chromosome breakpoint

NPM gene rearrangement is observed in certain types of non-Hodgkin's lymphoma (20–25). The 5′ portion of the NPM gene (in chromosome 5) is fused to other genes in t(2:5), t(5:17) and t(3:5) chromosome translocations. These findings indicate that within the NPM gene, there are regions vulnerable to chromosome breaks. Our results indicate that the chromosome break occurs [for t(2:5) and t(5:17) translocation] within intron 4 of the NPM gene. Intron 4 is relatively smaller (911 bp long) than those in Philadelphia chromosome translocation where the breakpoints are located in a 20 kb region inside a 70 kb intron (41). The highly consistent breakpoint of NPM observed in t(2:5) and t(5:17) suggests that there may be a 'fragile site' in intron 4 vulnerable to cleavage instead of rejoining to the correct chromosome. Intron 4 contains two *Alu* sequences and T-stretches

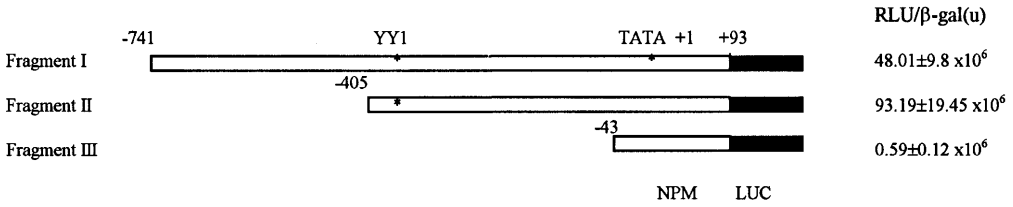

Figure 8. Determination of NPM expression with luciferase assay. DNA fragments (I, II, III) from the 5′ region of NPM gene were subcloned into luciferase reporter gene vector pXP1. These constructs were transfected into HeLa cells and their expression activities were determined. The luciferase activity was normalized with β-galactosidase reporter gene (co-transfected) activity and the data represents the average ± SD of five experiments.

(9–17 Ts) at the 3′ end (Fig. 2B). The *Alu* sequences have 70–80% sequence homology to the *Alu* sequences observed in the human ABL and BCR genes (36). The T-stretches at the 3′ end may facilitate triplet formation (42), altering the DNA structure to favor breakage.

It is hypothesized that abnormal expression of genes due to gene rearrangement may affect normal cell growth and differentiation. The NPM promoter may cause overproduction of hybrid proteins which have tyrosine kinase (ALK) or retinoic acid receptor activities. These hybrid proteins may play a role in tumorigenesis. It was reported that over-expression of the NPM–ALK hybrid protein caused a malignant transformation of NIH 3T3 cells (43).

Analysis of the 5′ region of the NPM gene

As a first step toward understanding NPM gene regulation and expression, we investigated protein bindings in the 5′ region of the NPM gene with EMSA and footprinting analysis. With DNA fragments E (–741/–250) and P (–249/+465), we identified three DNA binding proteins (Fig. 4). Footprint analysis and competition EMSA (Figs 5 and 6) showed that the binding site for proteins A and B is between –371/–344. We identified that protein A is YY1. The binding of YY1 to this sequence is specific because a modified oligo (NPM-*cis*-A) with changes in the consensus sequence could not effectively compete for its binding to YY1 (Fig. 6). YY1 is a zinc finger transcription factor that binds to promoters and enhancers of many viral or cellular genes (27–29). Depending on which proteins it associates with, YY1 can either be a positive or negative factor (27–29,44). It is also reported that binding of YY1 bends the DNA near the promoter/enhancer region which facilitates the interaction of protein factors on either side of its binding site (45).

There are three potential factor binding sites around the YY1 binding site (Figs 1 and 2). They are (i) E1A-F binding site (AGGACGT) (37), located at –294; (ii) ZRE-1 site (TTACACA) (38), located at –377; and (iii) two copies of UBP-1 binding site (CTCTCTGG), located at –414 and –550 (39). E1A-F, ZRE-1 and UBP-1 are cellular proteins which bind to enhancer elements and activate viral gene expression. One could speculate, based on this observation, that the factors responsible for viral gene activation may also affect NPM gene expression. Nonetheless, the bending of DNA (45) in the enhancer/promoter region of NPM gene by YY1 could help to facilitate interactions among factors and eventually control NPM gene expression. Our finding that YY1 binds to the 5′ region of the NPM gene is particularly significant, since YY1 is also found to associate with NPM protein (11). These findings suggest that NPM may be involved a feedback control mechanism.

ACKNOWLEDGEMENTS

We thank Dr Edward Seto for the gift of YY1 protein and antibody, and Dr Ram Reddy for his critical reviewing of this manuscript. This work was supported by a NIH grant (CA 42476) from the National Cancer Institute.

REFERENCES

1 Chan, W.Y., Liu, Q.R., Borjigin, J., Busch, H., Rennert, O.R., Tease. L.A., and Chan, P.K. (1989) *Biochemistry,* **28,** 1033–1039.
2 Chan, P.K., Aldrich, M., and Busch, H. (1985) *Exp. Cell Res.,* **161**, 101–110.
3 Yung, B. Y.-M., Busch, R. K., Busch, H., Mauger, A. B., and Chan, P. K. (1985) *Biochem. Pharmacol.* **34**, 4059–4063.
4 Chan, P.K. (1992) *Exp. Cell Res.*, **203**, 174–181.
5 Chan, P.K., Aldrich, M.B., and Yung, B. Y-M. (1987) *Cancer Res.* **47,** 3798–3801.
6 Sweet, P., Chan, P.K. and Slater, L.M. (1989) *Cancer Res.* **49,** 677–680.
7 Yung, B. Y.-M., Busch, H. and Chan, P. K. (1986) *Cancer Res.* **46,** 922–925.
8 Finch, R.A., Revankar, G.R. and Chan, P.K. (1993) *J. Biol. Chem.* **268,** 5823–5827.
9 Wu, M.H., Lan, C.Y. and Yung, B. Y-M, (1995) *Biochem. J.*, **305,** 987–992.
10 Finch, R.A. and Chan. P.K. (1996) *Biochem. Biophys. Res. Comm.*, **222,** 553–558.
11 Inouye, C.J. and Seto, E. (1994) *J. Biol. Chem.* **269,** 6506–6510.
12 Valdez, B.C., Perlaky, L., Henning, D., Saijo, Y., Chan, P.K. and Busch, H. (1994) *J. Biol. Chem.* **269,** 23776–23783.
13 Fankhauser, C., Izaurralde, E., Adache, Y., Wingfield, P., Laemmli U.K. (1991) *Mol. Cell. Biol.* **11,** 2567–2575.
14 Adachi, Y., Copeland, T.D., Hatanaka, M. and Oroszlan, S. (1994) *J. Biol. Chem.* **268,** 13930–13934.
15 Feuerstein, N., and Mond, J.J. (1987) *J. Biol. Chem.* **262,** 11389–11397.
16 Feuerstein, N., Chan, P.K., and Mond, J.J. (1988) *J. Biol. Chem.* **263,** 10608–10612.
17 Patterson, S.D., Grossman, J.S., D'Andres, P. and Latter, G.I. (1995) *J. Biol. Chem.* **270,** 9429–9436.
18 Chang, J-H. and Olson, M.O.J. (1990) *J. Biol. Chem.* **265**, 18227–18233.
19 Liu, Q.R. and Chan, P.K. (1993) *DNA Cell Biol.* **12,** 149–156.
20 Morris, S.W., Kirstein, M.N., Valentine, M.B., Dtiimer, K.G., Shapiro, D.N., Saltman, D.L. and Look, A.T. (1994) *Science* **263,** 1281–1284.
21 Sandlund, J.T., Downing, J.R. and Crist, W.M. (1996) *N. Engl. J. Med.*, **334,** 1238–1248.
22 Bullrich, F. Morris, S.W. Hummel, S., Pileri, H., Stein, H. and Croce, C.M. (1994) *Cancer Res.* **54,** 2873–2877.
23 Lamant L. Meggetto F. al Saati T. Brugieres L. de Paillerets BB. Dastugue N. Bernheim A. Rubie H. Terrier-Lacombe MJ. Robert A. Rigal F. Schlaifer D. Shiuta M. Mori S. Delsol G. (1996) *Blood* **87,** 284–291.
24 Redner, R.L., Ruch, E.A., Faas, S., Rudert, W.A. and Corey, S.J. (1966) *Blood*, **87,** 882–886.
25 Yoneda-Kato, N., Look, A.T., Kirstein, M.N., Valentine, M.B., Raimondi, S.C., Cohen, K.J. Carroll, A.J. and Morris, S.W. (1996) *Oncogene*, **12,** 255–275.
26 Flanagan, J.R., Becker, K.G., Ennist, D.L., Gleason, S.L., Driggers, P.H., Levi, B-Z., Appella, E. and Ozato, K. (1992) *Mol. Cell. Biol.*, **12,** 38–44.

27 Shi. Y., Seto, E., Chang, L-S., and Shenk, T. (1991) *Cell*, **67,** 377–388.
28 Hariharan, N., Kelley, D.E. and Perry, R.P. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9799–9803.
29 Park, K. and Atchison, M. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 9804–9808.
30 Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989) *Molecular Cloning.* 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
31 Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1994) *Current Protocols in Molecular Biology.* Vol I–III. Harvard Medical School. John Wiley & Sons, Inc.
32 Dignam, J.D., Martin, P.L., Shastry, B.S. and Roeder, R.G. (1983) *Methods Enzymol.* **101,** 582–598.
33 Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
34 Nordeen, S.K. (1988) *BioTechniques* **6**, 454–457.
35 Chang, D.C. (1992) *Guide to Electroporation and Electrofusion.* Academic Press, pp. 429–455.
36 Heisterkamp, N., Stam, K., Groffen, J., DeKlein, A., and Grosveld, G. (1985) *Nature*, **315,** 758–761.
37 Yoshida, K., Narita, M. and Fujinaga, K. (1989) *Nucleic Acid Res.*, **17,** 10015–10034.
38 Lieberman, P.M., Hardwick, J.M., Sample, J., Hayward, G.S. and Hayward, S.D. (1990) *J. Virol.*, **64,** 1143–1155.
39 Wu, F.K., Garcia, J.A., Harrich, D. and Gaynor, R.B. (1988) *EMBO J.*, **7,** 2117–2129.
40 Chang, J.H. and Olson, M.O.J. (1988) *J. Biol. Chem.* **264,** 11732–11737.
41 Denny, C.T., Shah, N.P., Ogden, S., Willman, C., McConnell, T., Crist, W., Carroll, A. and Witte, O.N. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4254–4258.
42 Wells, R.D., Collier, D.A., Hanvey, J.C., Shimizu, M. and Wohlrab, F. (1988) *FASEB J.* **2,** 2939–2949.
43 Fujimoto, J., Shiota, M., Iwahara, T., Seki, N., Satoh, H., Mori, S. and Yamamoto, T. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 4181–4186.
44 Yang, W-M., Inouye, C.J. and Seto, E. (1995) *J. Biol. Chem.* **270,** 15187–15193.
45 Natesan, S. and Gilman, M.Z. (1993) *Genes Dev.* **7,** 2497–250.