Volume 17 Number 19 1989 Nucleic Acids Research

A *cis*-acting transcription element of the c-*myc* gene can assume an H-DNA conformation

Alan J.Kinniburgh

Department of Human Genetics, Roswell Park Memorial Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

Received June 26, 1989; Revised and Accepted August 23, 1989

ABSTRACT

I have used chemical probes and an oligonucleotide-association assay to determine the structure of a nuclease-sensitive, c-*myc* DNA region. I find that this DNA region can form a triplex-single stranded conformer *in vitro*—the H-DNA conformer. This DNA region has been shown previously to be a positive, *cis*-acting transcription element of the c-*myc* gene and to bind nuclear factors, including a base-paired ribonucleoprotein[10]. Therefore, H-DNA may be a functionally important *in vivo* topoisomer where the H-DNA and B-DNA conformers have different transcriptional activities.

INTRODUCTION

Both synthetic homopolymeric DNAs and naturally occurring heteropolymeric simple repeats with strong purine/pyrimidine strand asymmetry assume an H conformation when negatively supercoiled and/or at acidic pH[1-8]. The H conformer of DNA contains both triple-stranded and single-stranded regions. These structures are also sensitive to single-strand-specific nucleases when in superhelical plasmids. Such a DNA region is found at −125 bp in the 5′ flanking region of the human c-*myc* gene and has a large purine/pyrimidine strand symmetry[9,10]. We and others have shown that this DNA region is required for a high level of c-*myc* gene expression and binds several nuclear factors[10,11]. Recently, my co-workers and I have shown that one of these factors is a ribonucleoprotein which can base-pair to this nuclease-sensitive, *cis*-acting DNA element[10].

Since DNA denaturation is important in H-DNA formation *in vitro*[8], these nuclear factors may induce H-DNA formation *in vivo*. However, since no *perfectly matched* H-DNA structure can be drawn from the DNA sequence of this *cis*-acting element, it was important to determine if this DNA region could assume an H conformation. Therefore, I have examined this c-*myc* DNA region for its ability to form H-DNA in superhelical plasmids. Using chemical probes and an oligonucleotide-association assay, I find that the c-*myc* nuclease-sensitive DNA region shows reactivity expected of a triplex/single-stranded, H-DNA structure.

MATERIALS AND METHODS

Plasmid preparation, chemical modification, and gel electrophoresis

Plasmid DNA was purified from *E. coli* by the method of Holmes and Quigley[12]. Ten micrograms of a c-*myc* plasmid subclone, pc-*myc*282, were used in each sample. The human c-*myc* region of this subclone extends from a *Pvu* II site (−353 bp) to the *Sma* I site (−101 bp). The *Pvu* II/*Sma* I c-*myc* DNA was inserted at the *Hinc* II site of pUC 13. Each reaction

was carried out in 100μl of buffer. 50mM Na acetate at pH 4 or 5 and 100 mM Na cacadylate at pH 6 or 7 were utilized. Reactions were commenced by the addition of 0.5μl of DMS or 10μl of DEPC. The samples were vortexed and DMS samples were incubated 2′ at 22°C. Reactions were stopped using standard Maxam and Gilbert G-stop solution[13]. DEPC-treated samples were incubated for 20 min. Reactions were stopped by addition of 3 volumes of ethanol. OsO_4 modified samples were reacted in 100μl of buffers. Samples were made 10mM in OsO_4 and the reaction was started by the addition of pyridine (2μl). The samples were then incubated for 5 min. at 22°C. Reactions were stopped by the addition of ethanol (3 volumes). OsO_4 samples were processed as described above. The modified samples were digested with *Eco*RI. 3′ end-labeling was performed using ^{32}PαdATP (>3,000 Ci/mmol.) and the Klenow fragment of *E. coli* DNA polymerase I. 5′ end-labeling was performed as follows:

*Eco*RI digested DNA was treated with alkaline phosphatase, 5′ end-labeled with [^{32}P]-γATP (5,000 Ci/mmol.), and digested with *Hind* III (polylinker site near *Pvu* II site, −353 bp, of c-*myc*). The ~ 300 bp end-labeled c-*myc* DNA fragment was purified by agarose gel electrophoresis and electroelution. Purified DNA fragments were cleaved at modified bases by treatment with 1 M piperidine at 90°C for 30 min.[13]. After lyophilization, samples were resuspended in 99% formamide, 0.2% xylene cyanol, 0.2% bromphenol blue and electrophoresed through a 7 M urea−5% polyacrylamide gel (30×40cm) at 1,500V. Gels were dried and exposed to x-ray film.

Oligonucleotide-Association assay

Two μg of p*myc* 282 (−353 to −101 of c-*myc* in pUC 13) was mixed with 50 ng of ^{32}P-labeled oligonucleotides in 100 *mM* Na acetate pH 4.0. The samples were incubated at 45°C for 20 min. and then electrophoresed through a 1.5% agarose gel buffer with 10 *mM* Na acetate pH 4.0 for 20 hrs. at 50V.

RESULTS

A hypothetical H-DNA conformer of a c-myc nuclease-sensitive region

Several eukaryotic genes contain homopurine-homopyrimidine stretches which are sensitive to single-strand-specific nucleases, such as S_1-nuclease, when in supercoiled DNA[14,15]. A generally accepted model for these DNA regions is a triplex-single stranded structure termed H-DNA[1−8]. Although the c-*myc* NSE sequence is S_1-nuclease sensitive in superhelical plasmids, it is not a simple sequence homopurine-homopyrimidine DNA region as many previously described H-DNA structures. However, recent work from Frank-Kamenetskii and Wells' laboratory[2,5] has shown complex, imperfect homopurine-homopyrimidine stretches can form H-DNA in superhelical plasmids. A hypothetical H-DNA structure can be drawn for the c-*myc* nuclease sensitive element (NSE) with one Hoogsteen base-pair missing and one A-T *not* Hoogsteen paired (Fig. 1). To confirm or refute this structure, I assayed the c-*myc* NSE with chemical probes and an oligonucleotide-association assay to obtain evidence for an H-DNA structure *in vitro*.

Chemical probing for H-DNA structures

Negative supercoiling and acidic pH favor the H-DNA conformer[2]. The protonation of cytosine residues facilitates Hoogsteen base-pairing, whereas high negative supercoiling can alleviate this requirement in certain instances[6]. I have utilized chemical probes which react with single-strand DNA in a base-specific and structure-specific manner. Diethylpyrocarbonate (DEPC) is hyper-reactive with adenines in single-stranded DNA and osmium tetroxide (OsO_4) is hyper-reactive with unpaired thymine bases[16]. In addition, dimethyl sulfate (DMS) reacts with the N-7 of purines, G > A. Therefore, DMS will

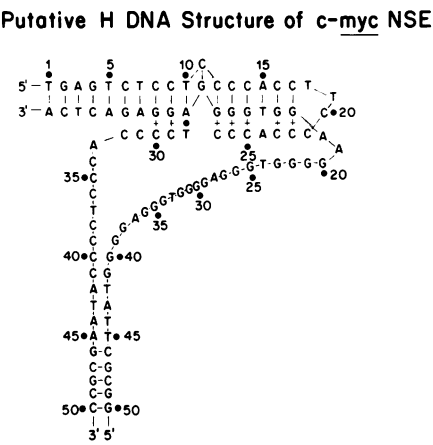

Figure 1. Putative H-DNA structure of the c-*myc* nuclease-sensitive element (NSE). An H-DNA structure was drawn using the model proposed by Frank-Kamenetskii and co-workers[1-4]. The data presented in this report was used as a guide (see text for details). Watson–Crick base-pairs are shown by a dash, AT Hoogsteen base-pairs by a dot and GC Hoogsteen base-pairs by a 'plus' sign.

be *hypo*-reactive when guanines are Hoogsteen base-paired[16].

I first examined the purine-rich strand by DMS and DEPC modification of a superhelical plasmid containing the c-*myc* NSE. DMS, at acidic pHs, shows a pronounced hypo-reactivity (compare the reactivity in NSE region with the reactivity in adjacent regions). There is a several fold decrease in reactivity throughout the entire NSE at pH 4 and pH 5 (Fig. 2). The general hypo-reactivity of the NSE may indicate that the NSE can assume alternate conformers. I shall discuss this possibility below. However, there are 8 G residues, bases 8,9,11–14,16 and 17 of the purine-rich strand of the NSE, which are particularly unreactive (Fig. 2; see Fig. 4 for summary). The results were quantitated by densitometric scanning for the pH 4,5 and 6 reactions (Fig. 2). The hypo-reactivity of these 8 G residues is readily observable at pH 5.0 where they are suppressed 40–70% as compared to pH 4.0 (Fig. 2). This asymmetry of DMS reactivity is consistent with protection of the N-7 position of these 8 guanines by Hoogsteen base-pairing (Fig. 1). Such asymmetric hyporeactivity is a hallmark of H-DNA structure, where these 8 guanines are both Watson–Crick and Hoogsteen base-paired (they are part of the H-DNA triplex). Curiously, G11 should be DMS reactive since it is not Hoogsteen base-paired. However, the prediction of figure 1 is a 'bulge' at this point and, therefore, C11 may not be accessible to DMS. Alternatively, a C residue further 3′ on the pyrimidine-rich strand may loop back and Hoogsteen base pair with G11. There are three other Gs, bases 25,26 and 36 of the purine-rich strand, which are slightly under reactive (Fig. 2 and 4).

DEPC modification of the purine-rich strand demonstrates four strongly reactive adenines, bases 18,19,28 and 37, one moderately reactive adenine, base 43, and a moderately reactive guanine, base 41, all at acidic pH (Figs. 2 and 4). The hyper-reactivity of these bases is not observed at pH 6.0 (Fig. 2). The doublet adenines, bases 18 and 19, define the border between triplex and single-stranded DNA since they abut the cluster of hypo-reactive guanines (Figs. 2 and 4). On the other hand, the moderately reactive adenine (base 43) may indicate that this region is 'breathing'. This interpretation is consistent with this base adjoining the junction between single-stranded DNA and B-DNA (Fig. 1). The DEPC

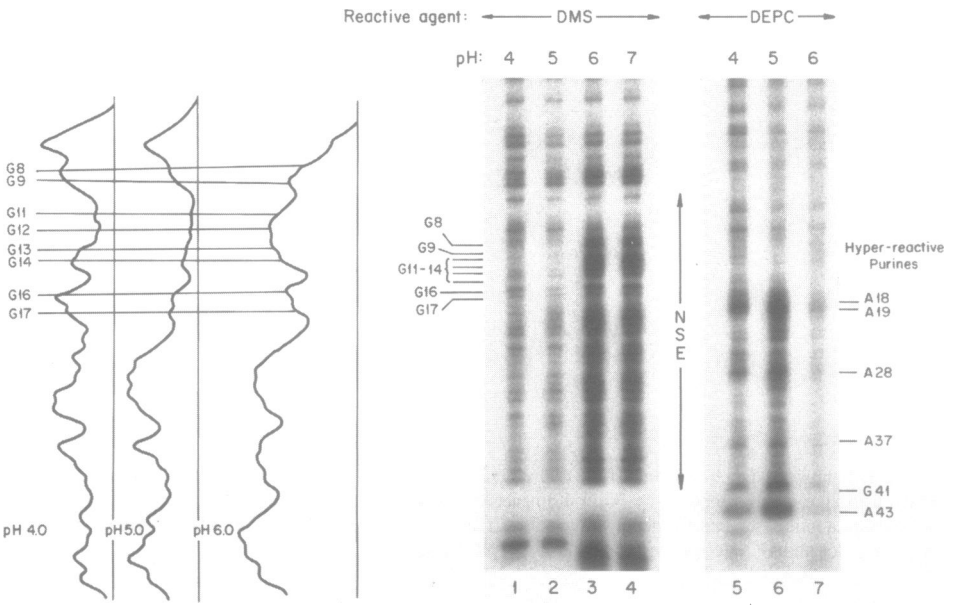

Figure 2. Dimethylsulfate (DMS) and Diethylprocarbonate (DEPC) modification of a superhelical c-*myc* plasmid subclone: analysis of the purine-rich strand. Ten micrograms of a c-*myc* plasmid subclone, pc-*myc*282, were used in each sample. The human c-*myc* region of this subclone extends from a *Pvu* II site (−353 bp) to the *Sma* I site (−101 bp). Each reaction was carried out as described in Materials and Methods. The plasmid DNA was cleaved with *Eco*RI, phosphatased, and 5′ end-labeled as described in Materials and Methods. Lane 1, DMS, pH 4; lane 2, DMS, pH 5; lane 3, DMS, pH 6; lane 4, DMS, pH 7; lane 5, DEPC, pH 4; lane 6, DEPC, pH 5; lane 7, DEPC, pH 6. The left hand side of this figure shows densitometric scans of the DMS reactions. pH 7.0 reaction is not shown but was similar to the pH 6.0 reaction.

and DMS experiments have been repeated using different c-*myc* subclones with identical results (data not shown).

I next reacted the c-*myc* plasmid subclone with OsO_4 to examine unpaired thymines in the pyrimidine-rich strand. At acidic pH (4 and 5) three thymines on the pyrimidine-rich strand show hyper-reactivity with OsO_4, bases 18,19 and 43 of the pyrimide-rich strand (Fig. 3B and Fig. 4). The reactivity is asymmetric, with a doublet of Ts near the 5′ end of the NSE and a single T near the 3′ end of the NSE, base 43 (Fig. 3B). These data are consistent with the H-DNA structure shown in figure 1. The doublet Ts on the left side (nos. 18 and 19) are reactive because they are in a single-stranded loop between the Watson−Crick base-pairs and the Hoogsteen base-pairs of the triplex (Figs. 1, and 4). The hyper-reactive T near the 3′ end (right side, base 43) is unpaired as the model predicts (Fig. 1, and 4). What is not consistent with the model of figure 1 is the lack of reactivity in the penultimate T, base 37, of the pyrimidine-rich strand (Fig 4, marked X). This base should be single-stranded and hyper-reactive since the A-37 residue on the lower strand is hyper-reactive and therefore unpaired (Figs. 1, 2 and 4). In defense of the H-DNA model, I observe modest reactivity of this thymine at acidic pH in certain experiments.

I have examined the pyrimidine-rich strand with DEPC to score any unpaired adenines which punctuate this strand. There are five hyper-reactive adenines, bases 24,33,42,44 and 45 of the pyrimidine-rich strand (Figs. 3A, and 4). The two 3′ most adenines help

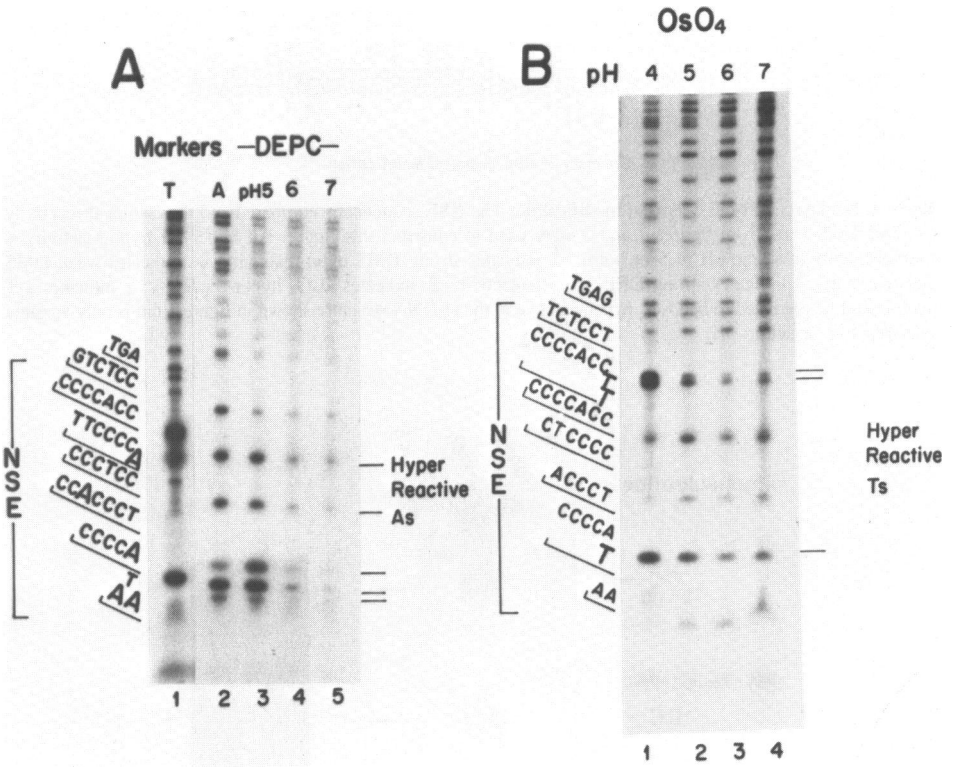

Figure 3. Diethylpyrocarbonate (DEPC) and osmium tetroxide ($0s0_4$) modification of a superhelical c-*myc* plasmid subclone: analysis of the pyrimidine-rich strand. The c-*myc* plasmid subclone used is described in figure 2. DEPC and modification was as described in Materials and Methods. The plasmid DNA was cleaved with *Eco*RI and 3′ end-labeled as described in Materials and Methods.

A. DEPC modification of c-*myc* NSE. Lane 1, marker T reaction performed with linear DNA and OsO_4; lane 2, marker A reaction performed with linear DNA and DEPC; lane 3, DEPC, pH 5; lane 4, DEPC, pH 6; lane 5, DEPC, pH 7.

B. OsO_4 modification of c-*myc* NSE. Lane 1, OsO_4, pH 4; lane 2, OsO_4, pH 5; lane 3, OsO_4, pH 6; lane 4, OsO_4, pH 7.

to further define the extent of the single-stranded regions within the H-DNA structure (Figs. 1 and 4). The three remaining hyper-reactive adenines, bases 24,33 and 42, are consistent with the triplex-single-stranded regions proposed (Figs. 1 and 4). Adenine, base 24, cannot Hoogsteen base-pair and is therefore not predicted to be within the triplex region (Fig. 1). The adenines base 33 and 42 are predicted to be in the single-stranded region of the pyrimidine-rich strand and are therefore hyper-reactive (Figs. 1 and 4). I have attempted to further probe the cytosines of the pyrimidine-rich strand with methoxylamine but find it gives variable results and only modest hyper-reactivity (data not shown). A summary of the chemical modification data is shown (Fig. 4).

An alternate assay of H-DNA

In further support of the H-DNA structure, an additional experiment was performed. Lyamichev and others[3] have shown that H-DNA-forming sequences can stably associate

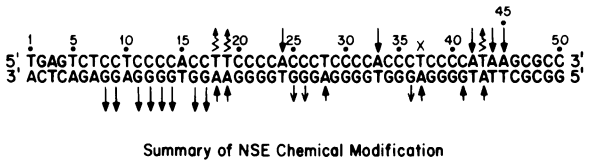

Figure 4. Summary of NSE chemical modification. The NSE sequence is numbered and the chemical reactivity is summarized. Data from figures 2 and 3 were used to construct this figure. All hyper- or hypo-reactivity is observable only at acidic pH (4,5 or both). ↓ indicates strong DMS hypo-reactivity; ↓ indicates weak DMS hypo-reactivity. ↑ indicates strong DEPC hyper-reactivity; ↑ indicates OsO_4 hyper-reactivity; x indicates a T which would be predicted to be OsO_4 hyper-reactive in the H-DNA structure shown in figure 1 but is only variably hyper-reactive at acidic pH.

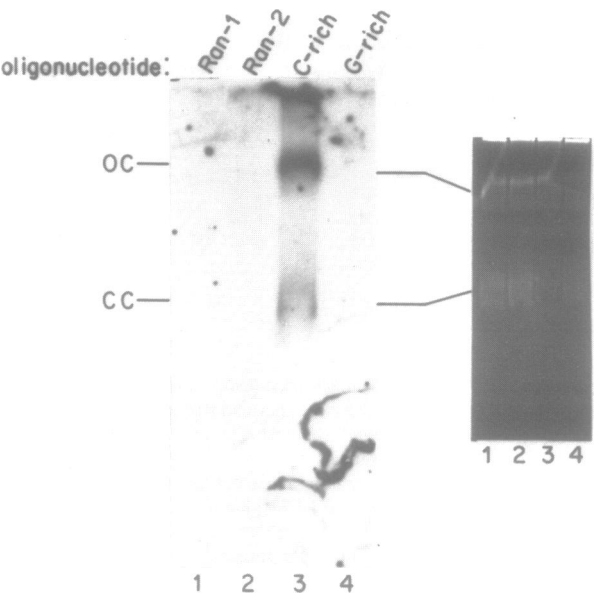

Figure 5. Co-electrophoresis of a nuclease-sensitive element, c-*myc* subclone and a ^{32}P-labeled oligonucleotide homologous to the c-rich strand. Two μg of p*myc* 282 (−353 to −101 of c-*myc* in pUC 13) was mixed with 50 ng of ^{32}P-labeled oligonucleotides in 100 *mM* Na acetate pH 4.0. The samples were incubated at 45°C for 20 min. and then electrophoresed through a 1.5% agarose gel buffer with 10 *mM* Na acetate pH 4.0 for 20 hrs. at 50V. The oligonucleotides were:

lane 1, a random sequence oligonucleotide of sequence, CAGCTAGCGTCGTGGCTGCGACTGACCTCACGGA.

lane 2, a random sequence oligonucleotide of sequence, TCCGTGAGGTCAGCTCGCAGCCACGACGCTAGCTG.

lane 3, an oligonucleotide homologous to the NSE pyrimidine-rich (C-rich) strand, GGATCCTTCCCCACCCTCCCCACCCTCCCAGATCT.

lane 4, an oligonucleotide homologous to the NSE purine-rich (G-rich) strand, AGATCTGGGAGGGTGGGGAGGGTGGGGAAGGATCC. The panel on the right is the ethidium bromide-stained gel. CC and OC refer to the position of migration of closed circular and open circular plasmid DNA.

with oligonucleotides that are homologous to their pyrimidine-rich strand but not those homologous to their purine-rich strand. This association occurs whether the H-DNA forming sequence is present in plasmids which are supercoiled, closed circular or linear. These associations are believed to result from Hoogsteen base-pairing of the pyrimidine-rich strand to the B-DNA form of the homologous region[3] and *not* by denaturation of the plasmid DNA with subsequent Watson–Crick base-pairing of the oligonucleotide. Since the purine-rich oligonucleotide cannot Hoogsteen base-pair to B-DNA, it should not be capable of a stable association with the plasmid bearing nuclease sensitive element. I have performed such an experiment utilizing plasmids which contain the natural c-*myc* NSE sequence. When either of two ^{32}P-labeled, random sequence oligonucleotides or ^{32}P-labeled, G-rich (purine-rich) NSE oligonucleotide is pre-incubated with NSE-containing plasmid, no association is observed after electrophoresis (Fig. 5, lanes 1,2 and 4). However, when the ^{32}P-labeled, C-rich (pyrimidine-rich) oligonucleotide is pre-incubated with the NSE-containing plasmid, a complex is observed (Fig. 5, lane 3). These results were obtained at pH 4.0, whereas no NSE plasmid oligonucleotide complexes were observed at pH 7.0 (data not shown).

DISCUSSION

These data, together with the S_1-nuclease mapping data presented elsewhere[10], support the idea that the c-*myc* NSE sequence can assume an H-DNA conformation similar to the structure proposed (Fig.1). One feature which is still uncertain, however, is the exact point at which the 3′ B-DNA structure is resumed (assumed to be position 46 GC base-pair shown in figure 1). The single-stranded character could extend, at most, another six base-pairs, since the T at this position is not hyper-reactive (Fig. 3B and data not shown).

Although the H-DNA structure may be more stable than other structures at acidic pH, it is most likely in equilibria with the B-DNA form and perhaps the H-y3 H-DNA form[6–8,16]. This latter structure would involve a triplex structure at the 3′ end of the pyrimidine-rich strand, rather than at the 5′ end as shown (Fig. 1). The two H-DNA conformers could account for the generalized reduction in DMS reactivity over the entire NSE region (Fig. 2) with the 5′ end or H-y5 triplex form (as shown in figure 1) being the predominant form (as Htun and Dahlberg claim for perfect H-DNA structures)[8].

Other possible models to account for these results are less satisfying than the proposed H-DNA structure. For example, I observe that the left side of the sequence is either unreactive or hypo-reactive, while the right side of the sequence is hyper-reactive with various chemical probes (indicating unpaired bases). Slippage loop models and alternating Watson–Crick/Hoogsteen duplexes cannot account for the asymmetry of these chemical modifications (see refs. 6–8 for a more extensive discussion of this subject). It is therefore likely that the c-*myc* NSE sequence can assume an H-DNA conformation in supercoiled DNA at low pH. This model accounts well for the S_1-nuclease sensitivity of the NSE[9,10] as well as its reactivity with chemical probes.

The data presented here has several important aspects. This is the first demonstration of an H-DNA structure within a defined, *cis*-acting transcriptional element. My laboratory has shown previously that deletion of this NSE significantly reduces the expression of the c-*myc* gene[10]. Deletions which drastically reduce the transcriptional activity of the c-*myc* gene also eliminate cutting by S_1-nuclease[10]. Therefore, it seems possible that the transcriptional activity of the NSE is dependent on a non-B DNA structure, presumably an H-DNA structure.

Several transcription factors bind to this NSE region[10,11]. One factor is a ribonucleoprotein wherein the RNA component base-pairs to the NSE DNA[10]. I speculate that this hybridization event may help to displace the NSE pyrimidine-rich strand and may be the first step in H-DNA formation *in vivo*. Since the c-*myc* NSE is a positive transcription element[10,11], H-DNA formation in this model would increase the transcriptional activity of the c-*myc* gene. Perhaps only the H-DNA conformer is actively transcribed whereas the B-DNA conformer is transcriptionally inactive.

In vivo the requirement for protonation of cytosines (low pH effect) might be overcome by the ribonucleoprotein and protein factors which specifically bind this DNA region. The ribonucleoprotein factor appears to base-pair with a region of the purine-rich strand which, in the H-DNA conformer, is single-stranded (Terri L. Davis and AJK, unpublished results). Should this interaction occur *in vivo*, it should lower the activation energy of the B-DNA to H-DNA transition. Locally high levels of DNA supercoiling might also participate in overcoming the activation energy required for the B-DNA to H-DNA transition. Further *in vivo* and *in vitro* experiments will be necessary to support these speculations.

ACKNOWLEDGEMENTS

I wish to acknowledge the helpful comments of Drs. David Kowalski, Robert Umek and Maarten Linskens. I also thank Donna Oleszek for technical assistance and Nancy Frame for secretarial assistance. This work was supported by a grant from the National Institutes of Health (CA 43661).

REFERENCES

1. Lyamichev,V.I., Mirkin,S.M. and Frank-Kamenetskii,M.D. (1985) J. Biomol. Struct. Dyn. **3**,327–338.
2. Mirkin,S.M., Lyamichev,V.I., Drushlyak,K.N., Dobrynin,V.N., Filippov,S.A. and Frank-Kamenetskii,M.D. (1987) Nature **330**,495–497.
3. Lyamichev,V.I., Mirkin,S.M., Frank-Kamenetskii,M.D. and Cantor,C.R. (1988) Nucleic Acids Res. **16**,2165–2178.
4. Voloshin,O.N., Mirkin,S.M., Lyamichev,V.I., Belotserkovskii,B.P. and Frank-Kamenetskii,M.D. (1988) Nature **333**,475–476.
5. Hanvey,J., Shimizu,M. and Wells,R.D. (1988) Proc. Natl. Acad. Sci. USA **85**,6292–6296.
6. Htun,H. and Dahlberg,J.E. (1988) Science **241**,1791–1795.
7. Johnston,B.H. (1988) Science **241**,1800–1804.
8. Htun,H. and Dahlberg,J.E. (1989) Science **243**,1571–1575.
9. Boles,T.C. and Hogan,M.E. (1987) Biochem. **26**,367–376.
10. Davis,T.L., Firulli,A.B. and Kinniburgh,A.J. (1989) submitted.
11. Hay,N., Bishop,J.M. and Levens,D. (1987) Genes and Dev. **1**,659–671.
12. Holmes,D.S. and Quigley,M. (1981) Anal. Biochem. **114**, 193–201.
13. Maxam,A.M. and Gilbert,W. (1980) Meth. Enzymol. **65**, 499–560.
14. Schon,E., Evans,T., Welsh,J. and Efstratiadis,A. (1983) CELL **35**,837–848.
15. Larsen,A. and Weintraub,H. (1982) CELL **29**,609–622.
16. Wells,R.D., Collier,D.A., Hanvey,J.C., Shimizu,M. and Wohlrab,F. (1988) FASEB J. **2**,2939–2949.,